Improved methods and actionable tools for enhancing HTA

WP10 - Guidance to support consistent HTA appraisal for orphan medicinal products (OMPs)

Deliverable 10.1

HTA Appraisal Framework Suitable for Rare Disease Treatments

K Facey[1], A Whittal[2], M Drummond[2,3], S Upadhyaya[4], T Junghans[1], E Nicod[2]

[1]University of Edinburgh – Scotland, [2]Bocconi University – Italy, [3]University of York – England, [4]National Institute for Health and Care Excellence – England

12 May 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779312.
The results presented reflect the author's views and not those of the European Commission.

Contents

Executive Summary **4**
List of Abbreviations **6**
I. Background **7**
IMPACT HTA Project 7
The challenge of appraising rare disease treatments 7
Terminology 8
Overview of WP10 research 8
Table 1: Ethnographic Observations of Scottish Medicines Consortium (SMC) 10
Table 2: Ethnographic Observations of National Institute for Health and Care Excellence, England 11
Table 3: Ethnographic Observations of CADTH, Canada 12
II. The focus of the appraisal framework **13**
Non-cancer rare diseases 13
Appraisal of RDTs, not pricing 14
An economic setting 14
III. Methods to develop this appraisal framework **14**
IV. The appraisal framework **16**
The ethos of consistent flexibility for any HTA system 16
A. Evidence Submission and Critical Assessment processes address all dimensions of value and identify uncertainties **18**
1. The entire HTA process should be shaped around clearly defined decision-making domains and any decision modifiers 19
2. All relevant evidence is obtained for each domain of decision-making and all decision modifiers 21
3. Critical assessment of clinical evidence explicitly considers what evidence could have been generated in the rare condition 25
4. Critical assessment of economic models takes account of the paucity of knowledge in rare diseases and judges whether the model is sufficient for decision-making 28
B. Structured Appraisal Deliberation considers all dimensions of value **31**
5. Appraisal committees are bespoke for RDTs, or general appraisal committees include several rare disease specialists 32
6. The deliberative appraisal discussion is driven by the domains of decision-making and use of modifiers is clearly understood 32
7. Uncertainties are characterized in terms of form, extent and implications for decision-making 34
8. Outcomes-Based Managed Entry Agreements may be used to resolve decision-relevant uncertainties, if collection of sufficient data is feasible 35
C. Underpinning: Iterative expert clinical and patient input throughout the process **36**
V. Discussion **42**

Appendix 1. HTA Methods Guidance on Handling Appraisal of Treatments for RDs .. 43

1.1 Methods for RDTs (OMP definition) 43

1.2 Methods for RDTs (Smaller populations than OMPs – "Ultra-Rare") 46

1.3 Methods that may be relevant for RDTs 49

Appendix 2 Country Guidance on Deliberation and Data Collection Specific to Rare Diseases 51

2.1 Deliberation Guidance for RDTs 51

2.2 Deliberation Guidance for Single and Short-Term Therapies (e.g., cell and gene therapies) that may be used in RDs..... 52

2.3 Deliberation and Further Data Collection for some RDTs 53

References 54

Please cite this publication as:

Facey K, Whittal A, Drummond M, Upadhyaya S, Junghans T, Nicod E. IMPACT HTA WP10 HTA Appraisal Framework Suitable for Rare Disease Treatments. 12 May 2021. [Online]. Available from: Impact HTA | Health Technology Assessment | Work Package 10 (impact-hta.eu) (or Zenodo)

For queries about this work contact: k.facey@btinternet.com or elena.nicod@dolon.com .

Executive Summary

Background

Orphan Medicinal Products or Rare Disease Treatments (RDTs) are challenging to appraise as the limited numbers for study often lead to small and/or uncontrolled trials of short duration. In addition to the limited clinical evidence base, there is often a paucity of knowledge about the disease and how patients progress. Furthermore, RDTs may have a high price to seek a return on investment from a small treatable population and strong demand given high unmet needs. These issues lead to major uncertainties in HTA appraisal of clinical and/or cost effectiveness. Some HTA bodies address this by using special processes to appraise RDTs; others add special features, or decision modifiers, to standard processes; others give no leniency to RDTs. The aim of this research has been to create an appraisal process beyond conventional economic evaluation that is suitable for RDTs within the context of fair resource allocation and equitable service provision for all people in a health system.

Methods

This research has explored how different forms of evidence and inputs are curated, input, assessed and appraised to inform appraisal recommendations about use of an RDT. It has included contact with HTA experts in countries across Europe, Canada, Australia and New Zealand to document processes and compare case studies. Innovative ethnographic research has been undertaken of appraisal committees to explore in-depth the nuances of appraisal and its deliberative processes. It has also undertaken focused work on two particularly challenging issues for RDTs – namely use of Patient-Reported Outcomes and implementation of Outcomes-Based Managed Entry Agreements. This work has involved literature reviews and several multi-stakeholder activities to explore issues, share processes and develop tools.

The Appraisal Framework

We originally aimed to develop an appraisal framework for RDTs, but as a result of consultation with HTA leaders we have developed recommendations for appraisal that are suitable for RDTs. These recommendations could be used to develop a bespoke appraisal process for RDTs within an organisation or across organisations or to enhance existing appraisal processes to ensure that they are fair for RDTs.

The essential things required to create such a fair appraisal process for RDTs are:

- leniency in critical assessment of evidence relating to RDTs to recognise the inherent limitations
- flexibility in process to support determination of the value of each RDT
- consistency in application of leniency and flexibility.

Hence, we make recommendations for an appraisal process that ensures every RDT is appropriately critically assessed and in which gaps in clinical evidence are complemented by other forms of evidence and inputs. The recommendations aim to ensure that leniency and flexibility are applied consistently in the appraisal of RDTs.

The appraisal framework recommendations are:

Evidence Submission and Critical Assessment processes address all dimensions of value and identify uncertainties	Appraisal Deliberation considers all dimensions of value
1. The entire HTA process is shaped around clearly defined decision-making domains and any decision modifiers	5. Appraisal committees are bespoke for rare disease treatments, or general appraisal committees include several rare disease specialists
2. All relevant evidence is obtained for each domain of decision-making and all decision modifiers	6. The deliberative appraisal discussion is driven by the domains of decision-making, and use of modifiers is clearly understood
3. Critical assessment of clinical evidence explicitly considers what evidence could have been generated in the rare condition	7. Uncertainties are characterized in terms of form, extent and implications for decision-making
4. Critical assessment of economic models takes account of paucity of knowledge in rare diseases and judges whether the model is sufficient for decision-making	8. Outcomes-Based Managed Entry Agreements may be used to resolve decision-relevant uncertainties, if collection of sufficient data is feasible
Clinical and patient experts are involved iteratively throughout the appraisal process to explain context of condition, existing care pathway and help resolve uncertainties related to determination of treatment value	
Delivering fair appraisal of Rare Disease Treatments through consistent flexibility	

At the heart of the IMPACT HTA Appraisal Framework is the requirement to define a decision-making framework beyond clinical and cost effectiveness (e.g., including nature of condition, organisational issues, ethical issues), along with modifiers that can deliver levels of flexibility in appraisal. This decision-making framework and modifiers then need to drive all parts of the HTA process, so that all those involved (stakeholders and appraisal committee members) can better contribute evidence and inputs. Detailed guidance is presented for each recommendation from our three-year research programme that will be of interest to all HTA stakeholders.

This will lead to a stronger evidence base for decision-making and best use of expert knowledge to resolve uncertainties and inform value judgements about an RDT. Different deficiencies in evidence, care pathways and access to expertise in a health jurisdiction, means this is not a rigid framework; deliberation and hard choices will still need to be made. However, the aim is that this appraisal framework will ensure a thorough but fair appraisal process for RDTs that takes account of the paucity of clinical evidence and knowledge.

Conclusion

This appraisal framework with enactment of the detailed guidance will support consistency of flexibility in appraising RDTs to ensure fairness, within a framework of accountability for reasonableness (Daniels and Sabin 2008). A fair process is one that ensures inclusion of all relevant evidence and knowledge, consistently throughout the process, to allow for the best decision possible given the unique circumstances of the disease.

List of Abbreviations

ATMP	Advanced Therapeutic Medicinal Product
CDEC	Canadian Drug Expert Committee
CF	Cystic Fibrosis
D	Deliverable
DSU	Decision Support Unit
EMA	European Medicines Agency
EU	European Union
EUnetHTA	European Network for HTA
HST	Highly Specialised Technologies
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
I.C.E.R	Institute for Clinical and Economic Review
HTA	Health technology Assessment
ICER	Incremental Cost Effectiveness Ratio
MAH	Marketing Authorisation Holder
MS	Milestone (report)
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OBMEA	Outcomes-Based Managed Entry Agreement(s)
OMP	Orphan Medicinal Product
PACE	Patient and Clinician Engagement
PRO	Patient-Reported Outcome
PROM	Patient-Reported Outcome Measure
PSA	Probabilistic Sensitivity Analysis
QALY	Quality-Adjusted Life Year
QoL	Quality of Life
RCT	Randomized Controlled Trial
RDT	Rare Disease Treatment
SMA	Spinal Muscular Atrophy
SMC	Scottish Medicines Consortium
TA	Technology Appraisal
WP	Work Package
WS	Workstream
WTP	Willingness To Pay

I. Background

IMPACT HTA Project

The IMPACT HTA project has been funded as part of the EC H2020 call for research on improved methods in economic evaluation and tools and guidance to support health technology assessment (HTA) and health system performance. It proposes new and improved methods, tools and guidance for decision-makers across ten research areas, which, overall, contribute to the understanding of costs and health outcomes variations within and across countries and to costs and health outcomes data integration from different sources.

The challenge of appraising rare disease treatments

Work Package 10 (WP10) has focused on an area where use of economic evaluation within HTA is most contentious, that is to inform decisions about access and reimbursement of orphan medicinal products (OMPs).

The European Medicines Agency (EMA) can grant an OMP (sometimes called “orphan”) designation to a medicine that treats a rare disease that occurs in less than 1/2,000 people, is life-threatening or chronically debilitating and where there is no satisfactory treatment. This attracts additional regulatory support for clinical development of the OMP and regulatory review by a specialist committee. Regulatory authorisation may also be expedited based on interim analyses, with a requirement for post licensing evidence generation. However, the HTA body must make a recommendation about use of the treatment based on the evidence available at the time of authorisation.

Internationally the definition of a rare disease differs and may depend on a prevalence figure or total number of patients in the country or it may not be defined. Additionally, in some countries, there may be an additional classification to delineate treatments for the most rare (ultra-rare) diseases. Some HTA bodies make this delineation for diseases with a prevalence of <1/50,000 people and refer to “ultra-orphans”, ultra-OMPs. In fact, recent research has identified 6,172 distinct rare diseases (Nguengang Wakap et al. 2020) and of the 5,304 with a recorded prevalence, 85% have a prevalence of <1/1,000,000. This is much lower than the ultra-rare definition used by most organisations and shows that many rare diseases will have very limited numbers for study. Furthermore, approximately 70% of these diseases are single cases and 10% reports of families.

Despite the large number of rare diseases, only 193 OMPs (some for the same condition) have been authorised over the past 20 years (up to 7 March 2021)[1]. So, the unmet need for rare diseases it still large.

In HTA, OMPs are challenging to appraise as the limited numbers available for study often lead to uncontrolled or small confirmatory trials of short duration. Furthermore, there is often limited clinical knowledge about many rare diseases. This lead to uncertainties about treatment benefit and challenging value judgements in economic modelling, within a context of strong demand given significant unmet need and high prices (Nicod et al. 2019). Some HTA bodies have stated that conventional estimates of cost/Quality Adjusted Life Year (cost/QALY) may not capture all elements of value of an OMP and that wider considerations are needed from a multi-stakeholder perspective, but this view is not shared by all HTA bodies.

[1] https://ec.europa.eu/health/human-use/orphan-medicines_en

At the outset of this project, it appeared likely that as a result of increased genetic testing and development of stratified/personalised therapies more OMPs may come to market and thus be subject to HTA. This has been the case and during this research the advent of Advanced Therapeutic Medicinal Products (ATMPs), such as cell and gene therapies, for rare diseases has expanded the appraisal challenge. How should one-off treatments that have very high upfront costs be evaluated when their benefit is estimated to be in the much longer term? As a result of these therapeutic developments, we have also seen countries that have exempted OMPs from appraisal, now requesting some form of appraisal. For example, in Lithuania an added benefit assessment is now mandatory for OMPs based on comparative trial non-surrogate endpoint outcomes. Furthermore, as reported in D10.3, several countries are implementing new processes for additional data collection post HTA appraisal for OMPs or ultra-OMPs (Germany, Scotland, Austria).

Given all these issues, appraisal of OMPs within a wider setting of fair resource allocation and equitable service provision for all people in a health system is challenging. Countries handle these challenges in different ways, but none have the extensive differences in process as implemented by the regulators. Some HTA bodies apply special appraisal and reimbursement mechanisms that can be used with a range of innovative or challenging products, others have bespoke features that allow flexibility in the appraisal processes used for OMPs (Nicod et al. 2020). All countries agree on the challenge of appraising these treatments and are seeking guidance to understand good practices and support fair decision-making. Hence, WP10 has developed guidance on novel approaches to appraise OMPs that will support robust, accountable, consistent decision-making across Europe.

Terminology

During our research we found it necessary to switch our terminology from OMP to “rare disease treatment” (RDT) to be more widely encompassing. In the EU, RDTs include OMPs, but may also include treatments for rare diseases for which the Marketing Authorisation Holder (MAH) has not sought an OMP designation and treatments that may be considered in reimbursement as “procedures” rather than medicines (such as cell therapy). Furthermore, we have seen international interest in our work, where the term OMP does not apply. Hence, we use the term RDT going forward unless we make reference to an organisation’s process that specifically uses the term OMP or “orphan”.

Overview of WP10 research

WP10 was established to develop a toolkit providing guidance for HTA bodies on novel approaches to appraising RDTs beyond conventional economic evaluation. It planned to explore how different forms of evidence and inputs obtained from a range of data sources and stakeholder perspectives could be integrated with economic modelling to inform robust, accountable and patient-centered decisions for RDTs.

The objectives of WP10 were threefold:

1. To develop guidance on novel approaches to appraising RDTs beyond conventional economic evaluation - *including tools to support use of Patient-Reported Outcomes (PROs) and Outcomes-Based Managed Entry Agreements (OBMEA).*
2. To identify how evidence and knowledge obtained from a range of sources (including economic evaluation) can be integrated into an HTA appraisal to inform robust and accountable decisions.
3. To bring together existing initiatives and different stakeholder perspectives to advance the understanding on the appropriate ways forward for OMP appraisals.

Objectives 2 and 3 outline the manner in which objective 1, the key deliverable of this work package, has been developed.

Four workstreams, were originally established to explore how evidence and knowledge from different sources and stakeholders can be integrated to inform robust, accountable, deliberative decisions about the value of treatments for rare diseases.
WS1: Documentation of existing HTA appraisal/reimbursement processes for rare disease treatments;
WS2: Ethnographic observation to optimize evidence and knowledge input to HTA appraisal processes for rare disease treatments;
WS3: Use of patient-reported outcome measures (PROMs) in appraisal of rare disease treatments;
WS4: Implementation of Outcome Based Managed Entry Agreements for rare disease treatments.

During the project an additional workstream (WS5) was added as we realised extra activities were needed to develop the appraisal framework drawing on emerging findings from all workstreams, other sources (such as international methods and process guides) and engaging with stakeholders to obtain their feedback.

WP10 used a range of qualitative, quantitative and action-oriented research methods with an interdisciplinary research team including expertise in statistics, social anthropology, health services research, HTA, health psychology, health service commissioning and health economics.

In WS1, we worked with country experts to document appraisal processes in jurisdictions across Europe, Canada, Australia and New Zealand. Each country's approach has been published in a country "vignette" on our website, showing standard appraisal processes, adaptations for RDTs and any specific aspects relating to PROs or OBMEA.[2] We then reviewed all responses to categorize special "features" or "decision modifiers" relating to RDT appraisal to understand how the uncertainties associated with RDTs are addressed and whether there is determination of value beyond cost effectiveness (Nicod et al. 2020). Then we explored how these features are reflected in practice through case study analyses comparing the appraisal processes for two RDTs in countries with and without special processes.

The WS1 country "vignettes" have been updated during the project and are now widely available in the rare disease community through our partnership working with the Orphanet, who link to the vignettes on their database. The learnings from WS1 laid the foundation for all other workstreams.

WS2 used desktop analysis and ethnographic observation to explore the gap between stated processes and principles for appraisal of RDTs and the actuality of the complex deliberative processes. Observations were undertaken at the National Institute for Health and Care Excellence (NICE, England), the Scottish Medicines Consortium (SMC) and CADTH (Canada). In close collaboration with the HTA leaders in these organisations, challenging RDT cases were identified, going through different processes and meetings were observed at various points in the appraisal process. These are outlined in Tables 1-3. These committee observations were augmented by 25 interviews with 30 individuals from all stakeholder groups, which helped explore the interstitial space between what happened and the perceptions of those involved.

[2] Impact HTA | Health Technology Assessment | Country Vignettes (impact-hta.eu)

Table 1: Ethnographic Observations of Scottish Medicines Consortium (SMC)

Treatment indication and regulatory authorisation	**Process** ***Modifiers***	**Patient & Clinician Engagement (PACE)**	**SMC Appraisal**	**Recommendation** (5 weeks after appraisal)
Tisagenlecleucel (Kymriah) One-off cell therapy EU PRIME pats up to 25 years old with B-cell acute lymphoblastic leukaemia in relapse post-transplant or 2nd/later relapse	Ultra-orphan framework *End-of-life*	11 Nov 2018	8 Jan 2019	Accepted as per indication
Patisiran (Onpattro) IV infusion every 3 weeks UK EAMS – 2/8/18-27/8/19 hATTR amyloidosis in adults with stage 1 or 2 polyneuropathy	Ultra-orphan framework *Unmet need*[1], *Substantial improvement in QoL*	9 Apr 2019	7 May 2019	Accepted as per indication
Lumacaftor-ivacaftor (Orkambi)[2] Tablet or granules twice daily cystic fibrosis ≥6 years old and 2-5 years homozygous for 508del mutation	Orphan-equivalent *Unmet need*[1]	11 Jun 2019	2 Jul 2019	Not recommended[3]
Tezacaftor-ivacaftor (Symkevi) & Ivacaftor Combination tablet in morning, monotherapy in evening Cystic fibrosis ≥12 years old homozygous 508 del mutation or with a range of other stated mutations	Orphan-equivalent *Unmet need*[1], *Substantial improvement in QoL*			Not recommended[3]
voretigene neparvovec (Luxturna) One-off gene therapy Vision loss due to inherited retinal disorder caused by biallelic RPE65 mutations and with sufficient viable retinal cells	Ultra-orphan pathway *None stated as no recommendation*	Not applicable, but attended New Drugs Committee 24 Sep 2019	5 Nov 2019	Initial assessment of clinical and cost effectiveness before mandatory data collection
Onasemnogene abeparvovec (Zolgensma) One-off gene therapy EU PRIME 5q SMA with bi-allelic mutation in SMN1 gene and clinical diagnosis of spinal muscular atrophy (SMA) Type 1, or up to 3 copies of SMN2	Orphan-equivalent *Substantial improvement in life expectancy* *Substantial improvement in QoL*	13 Jan 2021	2 Feb 2021	Accepted with restriction specifying pre-symptomatic patients "expected to develop SMA Type 1"

[1] *"absence of other treatments of proven benefit";* [2] previously rejected by SMC in 2016;
[3] Revised pricing agreement 4/9/20 to include - ivacaftor-tezacaftor-elexacaftor (Kaftrio) at time of licensing "which means that many patients with rare mutations which fall outside of scope of EMAs current licensing considerations will also be able to benefit from Kaftrio" CF Trust explains this is for patients with rare mutations covered in US FDA license."

Table 2: Ethnographic Observations of National Institute for Health and Care Excellence, England

Treatment indication and regulatory authorisation	Process	Appraisal Observation	Process after Observation	Recommendation
Voretigene neparvovec 2020) (Luxturna) One-off gene therapy Vision loss due to inherited retinal dystrophy caused by confirmed biallelic RPE65 mutations with sufficient viable retinal cells	Highly Specialised Technology (HST)	25 July 2019	Draft report published Aug 2019 (recommended use)	Recommended as per indication 9 Oct 2019
Volanesorsen (Waylivra) Sc injection once weekly for 3 months, then once fortnightly with stopping rule UK EAMS June 2018- May 2019[6] Adjunct to diet in adults with genetically confirmed familial chylomicronaemia syndrome and at high risk for pancreatitis in whom response to diet and triglyceride lowering therapy has been inadequate	HST	28 Nov 2019	26 Feb 2020 2nd Appraisal Committee to discuss responses to draft report	Recommended as per indication 21 Oct 2020
Emapalumab (Gamifant) iv infusion every 3-4 days until stem cell transplant EU PRIME Primary paediatric haemophagocytic lymphohistiocytosis with refractory, recurrent or progressive disease or intolerance to conventional HLH therapies	Technology Appraisal (TA)	(Technical Engagement 6 Mar 2020) 6 May 2020	23 Jul 2020 EMA refusal, so suspension of appraisal	Suspended
Onasemnogene abeparvovec (Zolgensma)* One-off gene therapy EU PRIME 5q SMA with bi-allelic mutation in SMN1 gene and clinical diagnosis of SMA Type 1, or up to 3 copies of SMN2	HST	8 Oct 2020 public information only	10 Feb 2021 2nd Appraisal Committee	*Draft Guidance 8 March 2021* *Symptomatic* *- 6 months or younger,* *- 7 to 12 months and agreed by national multidisciplinary team via auditable criteria, (excluding those who require permanent ventilation for >16 hrs/day or tracheostomy)* *Pre-symptomatic* *- via conditions in OBMEA*

* *observation non-focal and/or confirmatory*

Table 3: Ethnographic Observations of CADTH, Canada

Treatment indication and regulatory authorisation	Process	Appraisal	Process after observation	Recommendation
Voretigene neparvovec (Luxturna) One-off gene therapy patients with vision loss due to inherited retinal dystrophy (IRD) caused by RPE65 mutations	Canadian Drug Expert Committee (CDEC)	16 Sep 2020	Interactions with MAH re draft recommendation	Recommended, subject to • initiation criteria • prescribing conditions • pricing conditions (reduction in price) November 2020 Reports available January 2021

The interdisciplinary collaboration among the WP's members over the three years of this WS2 research enabled us to consider how a multi-stakeholder committee deliberates on evidence and uses stakeholder inputs to inform their value judgements and come to an HTA recommendation or decision. Each observation, interview and the emerging findings are documented in confidential files at the University of Edinburgh and MS41 provides a comprehensive overview of the methods and findings. As most members of the research team have been involved in these observations and interviews over the three-year period of the research, sharing of key insights and exploration of emerging issues has informed the work of the other workstreams and in return the emerging ideas from the workstreams has guided the questions to explore further in WS2.

WS3 and WS4 have undertaken a range of research activities (including multi-stakeholder workshops, systematic literature reviews, HTA case studies and a variety of consultation exercises) to develop tools that support use of PROs and OBMEA. These activities are reported in full in other milestone reports and D10.2 and D10.3. This Deliverable 10.1 presents the overarching appraisal framework that is suitable for RDTs drawing together all our research.

II. The focus of the appraisal framework

Non-cancer rare diseases

Our observations and interviews have expanded our understanding of the challenges related to appraisal of RDTs.

Non-cancer rare diseases are different to rare cancers because the paucity of knowledge with non-cancer rare diseases is much greater.

In cancer there are often clear and rapid diagnostic and treatment pathways, data available on natural history and prescribing and an understanding of the important clinical outcomes.

Non-cancer rare diseases not only have small populations but their multi-system clinical manifestations can lead to a wide range of impacts on the patient, meaning the optimal endpoints for study are unclear. Many non-cancer rare diseases occur in childhood and so create major carer burden. Many are hereditary and so have a range of implications for the nuclear and wider family. Non-cancer rare diseases are difficult to diagnose (with many mis-diagnoses over several years). Progression can be rapid and many are severely life limiting, others are chronic but may progress to a terminal stage (e.g., organ failure). Non-cancer rare diseases are not well documented, some don't have ICD-10 codes to allow chart reviews to determine natural history. With limited clinical expertise, different approaches to treatment may be taken within one health system, so the optimal standard of care and usual care pathway are not clear. Current treatments may only treat specific symptoms and off-label treatments may be used, particularly as many diseases occur in children.

For all these reasons our focus has been to study non-cancer rare diseases, with the one exception of tisagenlecleucel in the treatment of refractory acute lymphocytic leukaemia, to give us insights into the issues with cell therapies. As outlined in Tables 1-3 we observed a range of other treatment RDTs, for gene therapies, for conditions that occur in adulthood and are life limiting, for chronic rare diseases, for conditions that onset in childhood. This has provided us with rich insights to inform this appraisal framework.

Appraisal of RDTs, not pricing

As a result of the smaller number of patients enrolled in clinical studies, clinical development costs may be lower for RDTs. However, the return on this investment is often lower than for conventional treatments because of the much smaller patient population that will be treated in clinical practice. Therefore, RDTs can be associated with very high prices per patient (Berdud, Drummond, and Towse 2020). We recognise the issues about the high prices associated with RDTs but have not addressed these in our work. They need to be addressed, but as these are intertwined with political issues and confidentiality, we leave those aside as they were not the intent of our research. We just promote that MAHs request a “fair price” (Scottish Medicines Consortium terminology) and do all they can to support demonstration of best value of their RDT. What HTA and healthcare Payers can do, is create a transparent, consistent and fair appraisal system to determine the value of an RDT that draws on all available evidence and expertise recognizing the specificities of each RDT. This is what we try to support in this Deliverable.

An economic setting

This research has been undertaken to respond to a call to improve economic methods and so our focus has been on systems that undertake an economic evaluation, but much of the framework is relevant also for those systems that focus on determination of added benefit and then price negotiation.

The purpose of most HTA bodies is to inform decisions about fair allocation of resources for the entire population in the health system. This requires identification of treatments that provide patient benefit and are good value, considering best use of all health services within the budget available. Consideration of opportunity costs is important and so cost utility analyses are often used with willingness to pay thresholds (WTP) set informally or in legislation. Some HTA bodies recognise that in the interests of fairness, the needs of particular groups may override those of the broader population. Rare diseases may be one such special case as stated in the NICE Interim Methods Guide for the Highly Specialised Technologies Programme (2017). This document stated that a utilitarian approach to cost effectiveness (in which greatest gain for the greatest number is valued highly) is unlikely to produce guidance which would recognise the particular circumstances of very rare diseases (vulnerability of small groups with limited treatment options, nature and extent of evidence, challenges for companies to make a reasonable return on their research and development investment), but there must be some consideration of costs and benefits. It is with this backdrop we present our framework.

III. Methods to develop this appraisal framework

Workstream 5 drew insights from all other WP10 workstreams, particularly WS1 (as reported in MS40) and WS2 (as reported in MS41). In addition, we organised a range of public multi-stakeholder engagements at HTA conferences to learn more from stakeholders, start sharing findings and gain feedback. The key engagements relating to the entire framework are described here. Each workstream had its own research programme and engagement strategy and that is reported in the individual deliverables.

In the first 18 months of the project, we observed four RDTs going through the SMC’s Patient and Clinician Engagement and appraisal meetings. We received a range of returns to the WS1 template outlining country appraisal processes and undertook the first interviews with SMC stakeholders. We then created a first draft of the recommendations for appraisal of

RDTs that would become the structure for this appraisal framework. This was presented in a full-day workshop we organised at the HTAi Annual Conference in June 2019. The workshop was attended by 52 international participants, representing all stakeholder groups. The meeting opened with presentations of the emerging results from WS1, reflections from stakeholders, and presentations of appraisal systems in Scotland, Sweden, New Zealand and Canada. Then in small group work we discussed the following issues:

- How can we deal with gaps in clinical evidence?
- Should there be special HTA processes for RDTs?
- What tools do stakeholders need?

After feedback from the breakouts, our emerging recommendations, and areas we thought would need to change following the earlier workshop sessions, were discussed. In an afternoon session, issues related to OBMEA were explored. We wrote a detailed report of the workshop, reflecting as a team on all the inputs and agreeing next steps in each workstream.

After the workshop we modified the recommendations based on learnings from the workshop and presented them in the Project month 18 report.

In the following year, we undertook two observations in the NICE Highly Specialised Technologies (HST) programme and one in the Technology Appraisals (TA) programme, as well as an observation of a gene therapy in the new ultra-orphan pathway at SMC and further interviews. This further informed development of the recommendations. We discussed these with the final results of WS1 in an HTAi 2020 panel session involving HTA leaders and a patient expert. At this stage the recommendations were:

1. Flexibility is needed in appraisal frameworks, but also consistency ⇒ fairness
2. Inputs from clinical and patient experts are more important to explain the evolving understanding of the disease, who would be treated, what outcomes matter, how clinical trial effects can be interpreted in real-life, and advise on treatment stopping protocols
3. Additional criteria/alternative processes/modifiers need to be built into the entire HTA process - evidence submissions for each stakeholder, critical assessment, deliberative discussion (with appropriate frameworks, training and support for all stakeholders)
4. Need structure for decision-making beyond cost-effectiveness with clarity about how the decision-making works (stakeholders, invited experts, committee members)
5. Any modifiers to traditional processes should be explicitly presented for each product at each meeting so that all understand what flexibility is possible
6. Need to characterise different forms (and levels?) of uncertainty and what this means for decision-making
7. Stakeholders could help resolve uncertainties if they were notified in advance and able to submit evidence related to specific questions
8. Guidance is needed on modelling methods that are feasible for rare diseases – when are models too unstable to be the basis of decision-making?

Towards the end of 2020, we were able to undertake another NICE HST observation and had the opportunity to observe a gene therapy at CADTH in Canada that we had seen at NICE and SMC. We also undertook a supplemental observation of a case at SMC that we had observed at NICE. These final observations were particularly informative to hone the recommendations. We separated out the consistency and flexibility element that had previously been part of the recommendations to be an introduction to the rationale for the framework and refocussed the other elements. In January 2021, we discussed this final draft in our own WP10 webinar with the chair of an HTA committee, staff from two HTA bodies and an industry representative.

Appreciation of nature of condition	1. Early and targeted clinical inputs are required throughout the process
	2. Stakeholders are notified in advance of key uncertainties and able to submit evidence related to specific issues
Adaptable value assessment process	3. The context in which value is considered (and flexibility on standard rules permitted) is made explicit (modifiers)
	4. Specific modifiers are explicitly presented for each product at each meeting
	5. Additional criteria/modifiers/alternative processes are built into the entire HTA process
	6. Guidance is developed on modelling methods that are feasible for ultra-rare diseases
Structured appraisal framework	7. Uncertainties are characterized in terms of form and extent and implications for decision-making
	8. The domains of decision-making beyond clinical or cost-effectiveness are clearly delineated
	9. Outcomes-Based Managed Entry Agreements may be implemented

This discussion at the WP10 webinar was influential in helping us identify what needed to change and what required further explanation. In particular, we realised that patient and clinical expert involvement needed to span all aspects – so this was extracted from the recommendations.

IV. The appraisal framework

The ethos of consistent flexibility for any HTA system

An HTA process is shaped by the legislative remit or organisational strategy of the establishment within which it sits (such as evidence-based guidance, institute for healthcare improvement, public health, or a health insurance fund). This may lead to a clear-cut definition of what the HTA remit is – e.g., to evaluate added benefit or clinical and cost effectiveness (henceforth we call this “value”). Whatever the scope of the HTA body’s remit, its processes cannot be simply quantitative, both scientific and social value judgements are inherent in the process of interpreting the evidence (Nicod et al. 2017). Scientific value judgements are generally driven by well established processes of critical appraisal originating from evidence-based medicine. Social value judgements are less tangible and dependent on the jurisdictional context of the HTA system. In any HTA body, the way in which these judgements are applied is likely to have evolved over the years. Processes will have been influenced by health service policies (such as patient-centred care), principles (such as respect for autonomy, non-maleficence, beneficence and distributive justice) and priorities (for treating particular groups), stakeholder feedback on processes and development of international best practices.

Our work does not seek to alter the fundamental principles on which an HTA body works. However, we have undertaken robust multi-methods, international research that has allowed us to reflect on how HTA appraisal could be improved to better determine the value of RDTs. The resulting framework could be adapted for use by any individual HTA body or HTA collaborative group.

This appraisal framework is a set of recommendations with detailed guidance underpinning each recommendation that is supported by the tools presented in D10.2 and D10.3. The aim is to support any appraisal committee in making the best decision possible about an RDT given the unique circumstances of each rare condition and uncertainties that are bound to

arise. It assumes that economic evaluation is part of the process, but these elements can be ignored in systems based on evaluation of added benefit.

We originally aimed to develop an appraisal framework for RDTs, but as a result of consultation with HTA leaders we have developed recommendations for appraisal that are suitable for RDTs. These recommendations could be used to develop a bespoke appraisal process for RDTs within an organisation or across organisations or to enhance existing appraisal processes to ensure that they are fair for RDTs.

The essential elements required to create such a fair appraisal process for RDTs are:

- leniency in critical assessment of evidence relating to RDTs that recognises the inherent limitations
- flexibility in process to support determination of the value of each RDT
- consistency in application of leniency and flexibility.

Whatever system of appraisal is used, we recognise that "fairness" must be considered in terms of

- the overarching health system (e.g., seeking to maximize population health), and
- those with rare conditions, where there is paucity of knowledge, evidence and alternative.

Hence, we make recommendations for an appraisal process that ensures every RDT is appropriately critically assessed and in which gaps in clinical evidence are complemented by other forms of evidence and inputs. The recommendations aim to ensure that leniency and flexibility are applied consistently in the appraisal of RDTs.

The appraisal framework recommendations are:

Evidence Submission and Critical Assessment processes address all dimensions of value and identify uncertainties	**Appraisal Deliberation considers all dimensions of value**
1. The entire HTA process is shaped around clearly defined decision-making domains and any decision modifiers	5. Appraisal committees are bespoke for rare disease treatments, or general appraisal committees include several rare disease specialists
2. All relevant evidence is obtained for each domain of decision-making and all decision modifiers	6. The deliberative appraisal discussion is driven by the domains of decision-making, and use of modifiers is clearly understood
3. Critical assessment of clinical evidence explicitly considers what evidence could have been generated in the rare condition	7. Uncertainties are characterized in terms of form, extent and implications for decision-making
4. Critical assessment of economic models takes account of paucity of knowledge in rare diseases and judges whether the model is sufficient for decision-making	8. Outcomes-Based Managed Entry Agreements may be used to resolve decision-relevant uncertainties, if collection of sufficient data is feasible
Clinical and patient experts are involved iteratively throughout the appraisal process to explain context of condition, existing care pathway and help resolve uncertainties related to determination of treatment value	
Delivering fair appraisal of Rare Disease Treatments through consistent flexibility	

A. Evidence Submission and Critical Assessment processes address all dimensions of value and identify uncertainties

Our research shows that there is a paucity of clinical evidence for RDTs, not just from clinical trials, but also in terms of disease pathophysiology. This means that traditional appraisals of RDTs may not fully capture natural history, disease impacts, treatment benefits and disadvantages. This can lead to major uncertainties in determination of clinical effectiveness and economic models.

Our observations and interviews indicate that for RDTs:
the most common clinical effectiveness issues that arise are:

- studies involving small numbers, non-comparative and open label evaluations leading to potential biases
- challenges matching external comparators and undertaking indirect treatment comparison
- study of outcomes that do not capture the most important elements of the condition in terms of patient benefit
- discussion about what is a clinically relevant effect and what the effect means in daily life
- long-term effectiveness and safety
- generalizability from the trial to implementation in clinical practice;

and the most common cost effectiveness issues that arise are:

- model – construction of health states and estimation of transition probabilities
- extrapolations for BSC and treatment – often using complex modelling, with different models leading to different results
- derivation of utilities and differential use by treatment group
- costings – high costs for some states but low costs related to side effects.

We have observed stakeholder inputs that have helped elucidate the context of the rare diseases and resolve uncertainties, but we have also seen and heard how appraisal committees and their members struggle with the dissonance between weak clinical evidence and patient and clinician input that presents a much stronger case for benefit based on their own experiences.

The following recommendations and guidance seek to address some of these challenges by recommending ways to produce the best possible clinical and economic evidence, with appropriate critical assessment that takes account of the limitations in RD research, but also highlights potential biases. Importantly it goes beyond these traditional domains of rapid HTA, to encourage fuller HTA (Watt et al. 2008) to support the complex decision-making that is required with these treatments. Along with focussed inputs from stakeholders this creates a mosaic of evidence that is better suited to determine the value of an RDT.

1. The entire HTA process should be shaped around clearly defined decision-making domains and any decision modifiers

We recommend that RDT appraisals go beyond the traditional domains of clinical effectiveness (and cost effectiveness), to include

- nature of the condition
- patient, carer and family impacts
- organisational issues
- ethical issues.

This is similar to the domains of the HTA Core Model ® (EUnetHTA Joint Action 2 2016) outside the relative effectiveness assessment and economics domain, with an alteration to the patient and social aspects heading.

The entire HTA process should be shaped around these domains and should drive the deliberative appraisal process. HTA bodies may select only some of these domains or use them all but indicate they are not relevant for a particular RDT. The essential element is to have a transparent process that enables consideration of the issues that arise in these domains and to gather as much evidence as possible about them, so that the need for challenging value judgements is reduced.

In addition to the decision-making framework, any decision modifiers that will enable flexibility in the deliberative process should be outlined and considered consistently in every appraisal. This might include issues such as

- (extremely) severe
- rapidly progressive
- absence of suitable alternative treatment that is authorised/reimbursed (unmet need)
- disease predominantly affecting children
- expected to lead to premature death
- burden on carers
- innovative nature of treatment
- rarity
- equality.

Note that although we did not find any HTA body that explicitly stated they made modifications for diseases affecting children, in many interviews it was clear that not only appraisal committee members, but also those assessing the evidence, made more allowances when the condition affected children (interview #14, interview #25). We also increasingly heard that rarity in itself should not be a modifier, but that severity is more important. For us this depends on whether the rare disease is well characterised, if it is not, it is still disadvantaged beyond being a severe disease.

To decide within an HTA body, what decision-making domains and modifiers are appropriate for their setting, an organisational development approach is recommended with HTA staff and committee members in consultation with stakeholders. The agreed framework should then be clearly communicated and all aspects of the HTA appraisal process shaped accordingly.

All HTA staff and committee members should receive training about the entire decision-making framework for RDTs and any decision modifiers.

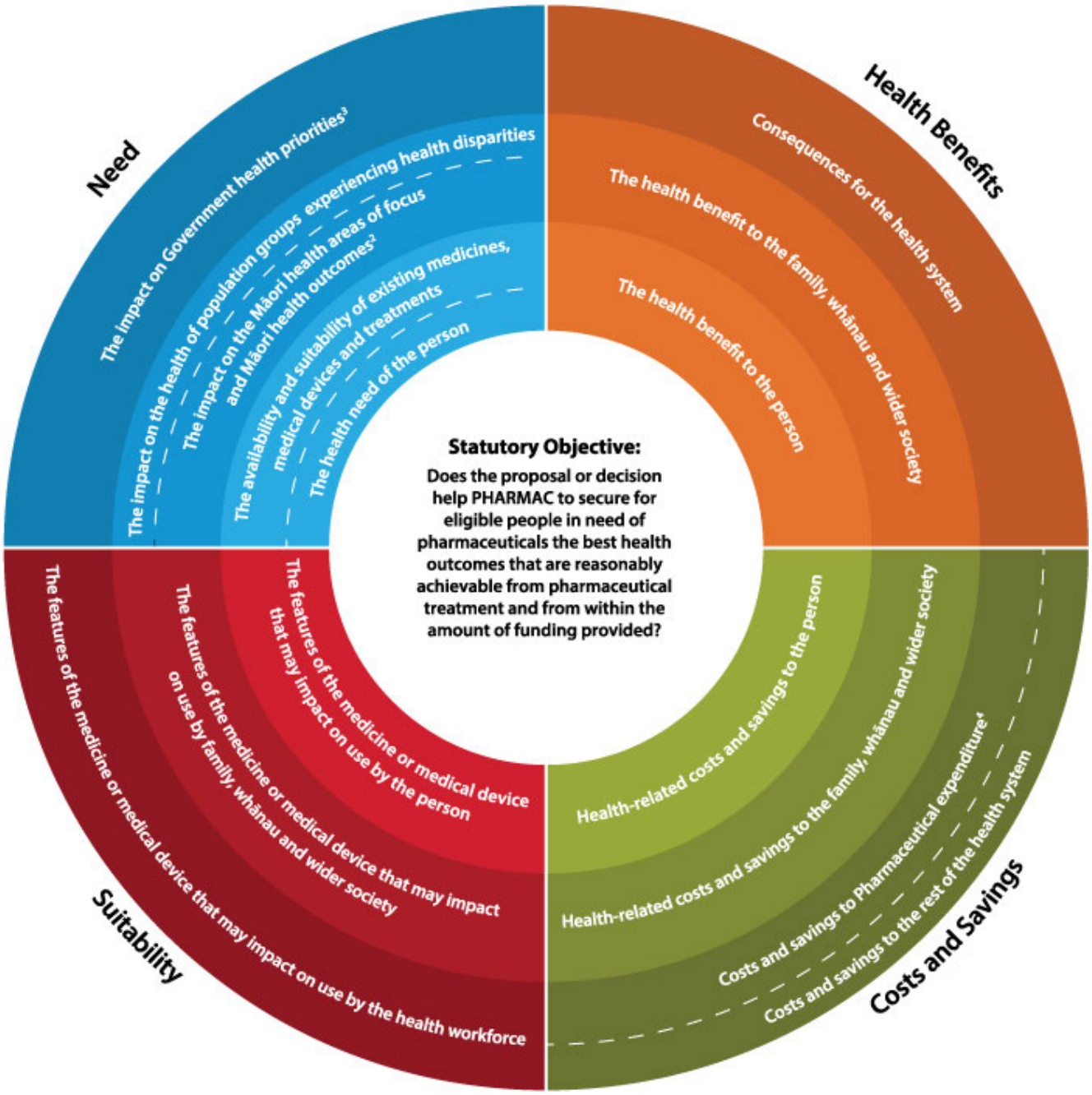
Stakeholder submission templates, expert review forms, critical assessment processes and the structure of deliberative discussion should have separate sections that relate to each decision domain and can capture evidence and insights related to decision modifiers.

An example from New Zealand

New Zealand has undertaken an extensive consultation exercise about their process for making funding decisions for new medicines. This resulted in "The Factors for Consideration" framework shown below that was published in 2020.

On the website this interactive diagram shows the elements the HTA body, PHARMAC, will consider in its appraisal deliberations, with more information associated with each element.

We are not recommending this framework per se, but we feel the transparency and presentation of different elements is helpful.



2. All relevant evidence is obtained for each domain of decision-making and all decision modifiers.

HTA decisions should be based on the best available evidence. This premise should apply to all domains of the decision-making framework.

Clinical and cost effectiveness

Early (and iterative) dialogues between the MAH, HTA bodies/Payers and other stakeholders are valuable to agree evidence sources that could contribute to determination of value of an RDT according to the TRUST4RD taxonomy (Annemans and Makady 2020) – population, disease, current therapy, new therapy, health system. This should include identification of the quantitative data that cannot be obtained in time for the appraisal submission and how such gaps could be resolved (expert input, qualitative research etc) or what considerations might be important from other domains.

Randomized Controlled Trials (RCTs) (even small ones) are encouraged, as uncontrolled trials require good understanding of natural history, which is often not the case for rare diseases. This is feasible as shown in a review of submissions for RDTs to CADTH from 2004 to 2015. Forty-seven of the 63 submissions (75%) relating to RDTs included at least one double-blind RCT and three included at least one open-label RCT. The average study size was 190 patients (range: 20 to 742) (Janoudi et al. 2016).

We observed that, even for ultra-rare diseases, treatments such as voretigene neparvovec for inherited retinal disorders and patisiran for amyloidosis could be studied with RCTs, which then demonstrated highly significant effects. In the case of the gene therapy, a double-blind approach was not feasible given the surgical procedure required, but the primary outcome was assessed blinded. However, questions were raised by CADTH about the blinding of the secondary outcomes that were used in modelling. It is important to ensure all outcomes that will be pivotal to the determination of benefit and construct of the economic model should be assessed in a blinded manner where possible.

The common alternative to an RCT is the construction of an external comparator, but in critical assessment many issues are found with these comparators given the small patient numbers and poor knowledge about natural history. Other more novel approaches to trial design, such as adaptive trials, crossover trials and n-of-1 trials have been suggested in Australia (Appendix 1) and Facey et al. (2014). Adaptive trials have been increasingly used in rare cancers, but we have not seen them for other rare diseases, MAHs could consider using more novel trial designs for RDTs.

Given the complexity of rare diseases and manifestations in different body systems, there should be clear justification for the outcomes studied – with description of the basis for the sample size calculation, any hierarchical testing policy and pre-specified sub-group analyses. As shown by Patisiran in amyloidosis, demonstrating significant effects on such a sound statistical footing is possible, it demonstrated significant effects on neuropathy and QoL, with the indication of effects in a sub-group that presented with cardiomyopathy.

The impossibility of measuring certain outcomes on some patients in the trial should be explained – e.g., six-minute walk test for those in wheelchairs (Duchenne Muscular Dystrophy), Forced Expiratory Volume in babies and toddlers (cystic fibrosis), visual function in infants (eye disorders).

Given the heterogeneity in presentation and progression of some diseases, variability should be reported with explanation of incidents that might cause spikes in outcome and trial measures to control those (volanesorsen and triglycerides affected by eating).

Attention should be paid to studying endpoints that will be needed in the construction of the economic model and that are associated with major resource use, such as certain types of serious adverse events and disease exacerbations leading to hospitalisation.

As clinical trials are likely to be small, there should be careful oversight of trial conduct and all efforts taken to reduce missing data. Attention should be paid not just to the primary endpoint, but to all endpoints that might be important in determination of added benefit or to construct the economic model. Collection of data after treatment discontinuation is particularly valuable.

If there is more than one clinical trial, a meta-analysis plan is helpful for an RDT to maximize use of all data. The IQWiG methods guidance (Appendix 1) suggests this and notes that for individual studies, a higher than usual significance level may be permitted, but this should be pre-specified.

For clinical and cost effectiveness, the MAH is expected to submit all relevant information to HTA. When there is a paucity of evidence in clinical trials all efforts should be made by the MAH to maximize data from pre-authorisation use (such as early access schemes), to provide data from longer-term follow-ups in clinical trials and real-world evidence if the treatment has been reimbursed in another country.

Real-world data from audits or registries arising from special access schemes in the jurisdiction should be requested from clinical experts.

We observed important discussion about animal studies for voretigene neparvovec in inherited retinal disorders to provide reassurance about duration of effect. This is unusual in HTA, but is likely to become more commonplace with advanced therapeutic medicinal products, so explanation of relevant pre-clinical data should form part of argumentation for determining the duration of the treatment effect.

The assessment report from the European Medicines Agency can be a particularly valuable resource for an HTA assessor. For an RDT this might help critique the pre-clinical data and documents the post authorisation efficacy and safety studies that have been requested.

Economic models should not be overly complex. All inputs to the model should be justified with efforts taken to show pre-specification and not cherry picking of the best assumptions, but those that are most clinically plausible. We recommend a transparent and reproducible model, such as that created by Discretely Integrated Condition Event (DICE) simulation in WP2.

The face validity of assumptions should be considered and extensive sensitivity and scenario analyses to determine which uncertainties drive the model.

Costings may be obtained using the WP3 European database and for sensitivity analyses with a wider perspective the work of WP4 taking account of patient lost productivity and carer costs should be considered.

Other decision-making domains

Evidence for decision-making domains outside clinical and cost effectiveness should use structured evidence collection and might include sources such as MAH and stakeholder submissions, systematic literature reviews, expert meetings/focus groups, interviews, expert consensus surveys, questionnaires to understand issues relating to the nature of the condition and organisational issues. Whatever methodology is used, there should be pre-specification of the study, rigour in conduct and unbiased reporting to provide robust evidence.

There should be a clear presentation of the **nature of the condition** relating to the population proposed for reimbursement. This will include elements such as:

- Epidemiology– prevalence and incidence by relevant sub populations in the jurisdiction
- Pathophysiology the disease and genetic implications of the disease (who else could be affected in a nuclear or wider family)
- Biological mechanism of action of treatment
- Characteristics of patients in the jurisdiction
    - age of onset
    - journey to diagnosis (which can be long for many RDs)
    - clinical manifestations of the disease (across all body systems)
    - what symptoms patients find most challenging
    - current treatment regimens (burden, effectiveness and unmet needs)
    - expected disease progression for the reimbursed population including life expectancy and relating to health states in economic model.
- Current treatments
    - their purpose (e.g., to treat specific symptoms)
    - the extent of treatments a patient may have had by the time they are considered eligible for the new treatment
    - the benefits and disadvantage of current treatments
    - an average day in the life of a patient – treatment regimens.

Patient, carer and family aspects may be a separate decision-making domain or integral to the other domains (such as clinical effectiveness). Whatever route is chosen, there should be clear reporting of activities to determine patients', carers' and family members' experiences and perspectives about the condition and treatments. This is particularly important in rare diseases.

- As rare diseases may have several clinical manifestations or affect body systems that are not often studied (like metabolism), clinical trials that study common outcomes are unlikely to capture the full impact on the **patient.**
- Many rare diseases occur in children and young adults and can be severely debilitating, this can have a major impact on **carers**: ability to work, physical and emotional strains, financing of equipment, lost social time etc. Hence it is increasingly common that carer impacts are considered. However, this is often done through an ad hoc adjustment to utility. It is recommended that the MAH undertakes specific work in their clinical development programme to determine carer impacts as part of the clinical trial and in other stages of the disease.
- In addition to the impact on carers, separate consideration may need to be given to **family** issues when treatment is burdensome and restricts family life, with particular consideration of the impacts on siblings given the hereditary nature of many rare diseases.

Furthermore, for rare diseases, the psychological issues faced by patients, carers and families might be more predominant than in other diseases as outlined in the underpinning section.

All these aspects may be studied through the MAH's clinical research programme by use of quality of life measures or qualitative sub-studies in clinical trials, or in separate research to develop robust patient-based evidence (Staniszewska and Werkö 2017) or surveys to understand burden of illness on patients, carers and families. Patient groups can also provide important input on these issues at the time of appraisal as outlined in the underpinning recommendation.

Organisational issues commonly involve issues relating to staffing and service provision. Consideration of organisational issues can help system readiness for a technology, which is particularly important for rare diseases where current treatment pathways are not uniform in a jurisdiction. Issues to consider might include:

- Current care pathway for patients, number of clinical experts in the jurisdiction, location of treatment centres, ability of patients to travel and provision for accompanying families of children, coordination of care across different sectors and impact on families
- Availability of any tests necessary to determine eligibility for treatment (particularly genetic tests)
- Provision of specialist/accredited services to ensure appropriate expertise, sufficient volumes/outcomes and training to ensure quality when a surgical procedure is required, centralisation of equipment, referral pathways that provide equitable access to all in the jurisdiction.
- Need for re-estimation of eligible patients when service established (often knowledge of prevalence is poor if there is no treatment for the RD, but when a treatment becomes available more people may come forward)
- For longer-term conditions, prioritisation processes may be needed to manage prevalent vs incident populations.

For **ethical issues**, a rapid literature review may be undertaken using the critically important elements of the EUnetHTA HTA Core Model v3.0 ® (page 254 onwards) such as:

- Benefit-harm balance
- Autonomy
- Respect for person
- Justice and equity
- Legislation
- Ethical consequences of HTA.

For an example of an ethical issues report, see the CADTH report for voretigene neparvovec.

Collaboration – an example

In rare diseases, we strongly recommend collaboration amongst all stakeholders to develop new tools that can support data collection in HTA. During our project we have been on the advisory board of Project Hercules and recommend their approach for other rare diseases.

Project Hercules

Project Hercules is a multinational collaboration set up by the patient organisation, Duchenne UK, to develop tools and evidence to support HTAs and reimbursement decisions for new treatments for Duchenne Muscular Dystrophy (DMD). It brings together patient organisations, clinicians, academics involved in HTA, leading pharmaceutical companies, HTA bodies and other advisers to build a better evidence base for DMD. Its outputs include:

Analyses of all available clinical data to build a **disease model** that reflects the progression of DMD, predictive capability of endpoints and transitions between stages of the disease. The analysis reflects meaningful disease stages to DMD patients, families and to HTA agencies (and added in an important new disease state compared with previous models).

An **economic model** that can be adapted by any company for the HTA of their treatment.

A **new Quality of Life measure** that will better capture important elements of DMD and which will generate a utility measure for use in QALY generation. This has been fully validated and translated into many languages.

A **Burden of Illness** study to provide a comprehensive measure of the impact of DMD on patients and their families, and in the social and healthcare sectors.

3. Critical assessment of clinical evidence explicitly considers what evidence could have been generated in the rare condition

Most HTA bodies have methods guides that form a starting point for critical assessment of clinical and cost effectiveness and some have detailed checklists that support the critical assessment, which are published. However, as shown in Appendix 1 and reported for England and Scotland in MS41, few of these guides take account of the issues that might arise when assessing clinical studies arising from the small, heterogeneous populations expected for an RDT.

Our research suggests that the following issues should be considered.

The regulatory procedure for an RDT can affect the evidence available for HTA. An RDT may go through an expedited regulatory process, be authorised on the basis of interim analysis, or be authorised for a restricted indication to ensure a favourable benefit-risk profile. This yields a small sample size but for an RCT can also impact the benefits of randomization, leading to imbalances in baseline characteristics. Thus, the implications of the regulatory process should be considered before starting a critical assessment.

The best evidence submissions from a MAH are those that have a clear "value thread" through all sections and honest critique of the evidence. This requires a flow of information across sections to present consistent arguments and ensure that the limitations which arise due to the nature of a condition are considered when determining what evidence it was feasible to collect. This is recognised by HAS in France who judge the quality of the evidence according to the prevalence of the disease and recruitable patients (Nicod et al. 2017). From the methods guides in Appendix 1, for ultra-rare diseases, New Zealand indicates that if data are not available they may use rapid assessment, recognising that more detailed assessment may not resolve key uncertainties. Meanwhile, Norway has a lower

requirement for documentation – indicating that it should be "as good as it can be", particularly noting that there will be considerable uncertainty.

From the country vignettes in WS1, 69% of the 32 countries report that they take a more lenient approach in their evidence requirements to demonstrate added benefit, recognising that there may be non-randomised evidence or use of short-term outcomes (Nicod et al. 2020). If this is the case, to ensure consistency and clarity of process, it would be helpful if HTA methods guides documented the leniency that is allowed for RDTs.

The outcomes studied in clinical trials should be compared to patients' perspectives on what aspects of the disease they most want to see changed with treatment, and based on views of clinical experts about key markers in disease progression. A clear statement should be made about what is NOT measured in the clinical evidence or obtained other submissions.

> Observed longitudinally, at a population-level, via the UK CF [cystic fibrosis] Registry, it will be possible to understand the changes to rate of incidence of these complications and decline in health in the treated population versus matched comparators. However, the cumulative benefit to quality of life – of opportunities taken rather than forgone, anxiety and stress avoided rather than endured, and time spent living a life on one's own terms rather than stalked by frequent, unpredictable episodes of ill health – is not possible to quantify directly. Current tools used to estimate Health-related Quality of Life (HRQoL) cannot sufficiently capture these nuanced but more fundamental aspects of health benefit, beyond the clinic setting. Nor can they allow a person with CF to observe their health status and circumstances dispassionately, stripped of crucial coping mechanisms and practices, in order to describe their quality of life with something approaching objectivity – a recognised limitation of studying HRQoL in people with lifelong or chronic ill health.
>
> Cystic Fibrosis Trust patient group submission for SMC, reproduced with permission

It is important that the link between the main outcomes in the clinical trial and outcomes indicative of clinical and patient benefit in the longer term are established (challenge with volanesorsen linking triglycerides to risk of acute pancreatitis). Some HTA bodies call this connection "surrogacy"[3], but the true demonstration of surrogacy is more challenging. German methodological guidelines, based on the original Prentice criteria (Prentice RL 1989) require surrogate endpoints to be biologically plausible and statistically validated, whereby there is demonstration of a treatment effect on the intermediate outcome and a treatment effect on the clinical outcome. Thus, it is unlikely that surrogacy will ever be proven in these terms for a chronic RDT. To show the link between clinical trial outcomes and longer-term outcomes, discussion of prognostic value, surrogacy of treatments in other similar conditions and generalizability of those results to the condition and treatment being studied is helpful. Consideration of this link is requested in Norway and they note they can involve Norwegian clinical experts in such evaluations.

The extent of missing data for endpoints used to determine added benefit and in the economic model should be clearly documented and consideration given to whether the data are missing at random (or if there could be informative censoring) and the impact this has on the analysis determined through different imputation methods.

There should be a clear schematic of all the evidence relevant to the value proposition including clinical trials (ph II and Ph III), long term studies, natural history studies, early access schemes, registries, other RWE. For an example of this see the NICE onasemnogene neparvovec slide 26. If the submission is the result of an interim analysis,

[3] Surrogate Endpoints | Evidera

date of availability and quantity of final data should be documented. A rationale should be given for any information not included in the submission

There are often challenges with MAIC due to difficulty in matching populations (particularly if the indication is for later line therapy) and obtaining comparative data, particularly in longer term (tisagenlecleucel) or for outcomes of interest (voretigene neparvovec – novel outcome about functional vision developed for the trial, so not data on this in natural history). This means that naïve indirect comparisons may be needed, which is not ideal.

HTA bodies need to develop clear guidelines about non-randomized comparisons, so that MAH's can provide clear, detailed and structured reporting of non-randomized comparisons as outlined in IMPACT HTA WP6.

Any checklists used to support critical assessment should be reviewed to determine any areas that need to be dealt with more leniently in rare diseases.

PROs

Considering the burdensome nature of RDTs and the fact that only few of the available treatments have a curative intent, it is important to account for the impact of disease and treatments on the QoL of patients and carers.

In HTA, QoL impact is generally captured through use of PROs. An understanding of the nuanced challenges around the development of PROs and use in HTA of rare diseases is crucial.

Our work from WS3 suggests that standard PRO methodologies may not always be suitable in rare diseases, and may result in inconclusive estimates and a range of uncertainties. Our research suggests this arises because:

- the small and heterogeneous population being studied is less likely to reach a minimally important difference defined in another disease, or effects are not significant due to underpowered analyses
- there are no validated PROMs for the disease, which may be a consequence of the challenges in demonstrating psychometric properties with small samples
- of difficulty in collecting PRO data from children, either because adult-specific measures are used, or because children-specific measures are challenging to administer, or because parents provide proxy assessments
- QoL evidence failed to capture a number of aspects of disease and treatment impacts that are important to patients in their daily lives. This could be because of the complexity of these conditions impacting multiple domains.

When critically assessing PRO evidence, it is important to consider the possible limitations of the measures presented.

Consideration of generic PROMs (such as EQ5D) alone are often not sufficient to capture the complexity of what makes good quality of life in a long-term, progressive, life-limiting condition (Powell et al. 2019).

Encourage use of a PRO, or a combination of PROs, that are holistic to pick up wider impacts than studied in primary endpoint – e.g., in cystic fibrosis, a respiratory questionnaire somewhat duplicates the effect seen in clinical trials, whereas other benefits to quality of life are stated were not quantitatively demonstrated (cystic fibrosis treatments observation).

Patients and clinicians should comment on whether quality of life measures studied adequately capture the key issues they face.

In open-label trials, recognise that biases may arise in reporting of PROs as a result of the patient's expectation of benefit.

Encourage patient input that elucidates quality of life (e.g., patient surveys, focus groups, submissions) to support the interpretation of PRO data that is often associated with a range of uncertainties. This helps provide a more holistic picture of the impact of disease and treatment on patients, their carers and families and what matters most to them.

Some rare diseases have a major impact on carers and families. This may arise not just in rare diseases in children, or adults at end of life, but also in other long-term progressive conditions (e.g., visual impairment). As highlighted in a NICE DSU report (Pennington and Wong 2019) further research is needed into family/carer outcomes across disease areas, with different interventions, and change in outcomes over time.

4. Critical assessment of economic models takes account of the paucity of knowledge in rare diseases and judges whether the model is sufficient for decision-making

As shown in Appendix 1, few national HTA methods guides make any dispensation for economic modelling of RDTs. Some note that cost effectiveness is not required (Belgium). From our country vignettes, we also know that cost effectiveness is not required for ultra-OMPs in Slovakia. Other methods guides indicate that ICERs may be higher than usual thresholds (Hungary, Norway). Only the Institute for Clinical and Economic Review (I.C.E.R.) in the USA provides specific guidance.

Critical assessment of the economic model should not be an academic exercise to unpick the model. Models are always an imperfect simplification. The question is, is it good enough to use as the basis for the HTA decision? To determine this, the assumptions and uncertainties that have a major impact on the ICER should be clearly explained (so-called decision-relevant uncertainties). This is crucial for the deliberative process, so that committee members can explore the most important aspects and agree the most plausible assumptions.

When rare diseases can result in many clinical manifestations, an economic model may not fully capture the impacts of the disease. The construct of the model over the entire time horizon should be discussed with clinicians to identify whether it is a sufficient representation of the disease. There should be triangulation with the important impacts identified by patients and clinicians to see if they have been captured in the model [redacted] There should be explicit documentation of what the economic model is NOT capturing so this can be considered in the deliberation.

Natural history data are an important part of the economic modelling of long-term effects, but in addition to the usual challenges of estimating natural history from registries or retrospective chart reviews, with some rare diseases complications often arise as there is no diagnostic code (ICD-10) and limited registry data. Careful scrutiny of natural history studies is required and a checklist should be developed to support this. This will be considered further in the RWE4Decisions initiative (Facey et al. 2020).

The primary and secondary outcomes in the clinical trials and a table showing the sources of inputs to the economic model should be presented. Any mismatch between the focus of the clinical study and construction of the health states in the economic model should be justified (voretigene neparvovec used a new outcome in the clinical trial but based health states on traditional visual measurements due to lack of long-term data on the novel outcome).

In rare diseases, some model inputs may need to be taken from other similar disorders; the clinical justification for this should be clearly explained. For example, for RPE65 inherited retinal disorder, it may be suitable to take utilities for each visual impairment health state from another form of progressive visual impairment as it's not a multi-systemic disorder, but transition probabilities may not be transferable as the rate of disease progression may differ.

Survival, duration of treatment effect and treatment waning over time are often estimated via sophisticated models that can produce very different estimates depending on assumptions used in the underlying statistical distributions. Furthermore, there are increasingly complex approaches, such as mixture cure models, which are often difficult for committees to fully understand. WP2 has studied survival extrapolation and indicates that a systematic approach is needed to justify the approach taken and its limitations. The framework proposed by NICE's Decision Support Unit (Latimer 2013) is recommended for extrapolation of survival, but WP2 also indicates that expert opinion and clinical plausibility should be used more frequently. This aligns completely with our findings, as discussed in the stakeholder involvement underpinning section. WP2 also promotes that use of external data can improve extrapolation and model choice. This is rarely presented in HTA, but is an area worthy of consideration.

Clear presentation of timing of benefits accrual is helpful and can show how much of the treatment effect relies on the extrapolation modelling (CADTH CDEC pharmacoeconomics report showed that 96% of the incremental benefits with voretigene neparvovec accrued beyond the timepoints for which clinical data were available).

Health State Utility Values

It is important to critically assess the possible limitations of using health state utility values (utilities) in rare diseases. Deriving utilities from PRO data requires large datasets, which is not always feasible in small and heterogeneous conditions. Also, it may not always be possible to generate utilities for each of the health states from these smaller datasets.

EQ5D values may be high at baseline for a condition that is said to have a major impact on QoL, this may be due to "response shift phenomenon". This occurs in chronic conditions where patients adapt to their long-term illness – it's their "normal". In such cases utilities may be higher than expected (orkambi, cystic fibrosis) and higher than general population (volanesorsen, familial chylomicronaemia syndrome) at baseline.

In relation to rare diseases, discussion often arises about the issue of public vs patient trade-offs.

> *"You've got the additional complication in this with the adaptation so with the EQ5D, what you want to know is what the public would trade to avoid being in that state. And sometimes they'll trade quite a lot to avoid being in that state, say being blind or being wheelchair bound, but if you ask those people who live in those states who have always been blind or who have adapted to being blind or who have always had a wheelchair, then they don't value those states like that. And there's a bit of tension because we are supposed to be looking at public preferences, public evaluations of these states, because then you avoid the adaptation of the individual minimising the impact. But equally that goes against the likelihood of getting the drug because the individual says it's not that bad because I'm used to it and you've got to kind of balance those two things."*
>
> *Interview #11, HST committee member*

Vignettes are often used to derive utilities for ultra-rare diseases, but often criticised. If vignettes are used to derive utilities, patient and clinical input should be obtained to ensure that the utilities reflect the circumstances of each stage of the disease.

An agreed methodology for development of vignettes to derive utilities is required that

- takes account of the limited knowledge about disease states and small number of clinical experts able to support development of vignettes
- carefully reviews health state descriptions to ensure they are fair
- recognised the need for resulting utilities to be discussed with patients in lay terms to check their face validity (e.g., blindness being worse than death, willingness to trade half your life to achieve…)

In our observations, we saw substantial discussions about whether a utility decrement should be applied for carer burden, often with ad hoc approaches be suggested in submissions, referring back to precedents in other conditions or treatments considered to be similar. As highlighted in a NICE DSU report (Pennington and Wong 2019) more work needs to be undertaken to agree when and how carer health effects should be included in economic evaluations.

In the I.C.E.R. Value Assessment Framework for Ultra-RDs guidance is given on flexible judgement of utilities, inclusion of wider costs and consideration of a societal perspective.

> When there are challenges translating the outcome measures used in clinical trials and available patient-reported data into QALYs, ICER will conduct a search for "mapping" studies that may allow translation of surrogate outcomes into quality-of-life measures. The validity of these mapping studies will be discussed with manufacturers, clinical experts, the patient community, and other stakeholders in order to get their input on the most feasible way to translate these other measures of patient outcome into QALYs.
>
> When the impact of treatment on patient and caregiver productivity, education, disability, and nursing home costs is substantial and these costs are large in relation to health care costs, ICER will present its base case health system perspective model results in tandem with the results of a scenario analysis inclusive of broader societal costs. This will most often occur in cases where the incremental cost-effectiveness ratio changes by greater than 20%, greater than $200,000 per QALY, and/or when the result crosses thresholds of $100,000-$150,000 per QALY.
>
> Similarly, a health-benefit price benchmark (HBPB) linked to the societal perspective analysis will be presented alongside the standard HBPB.
>
> I.C.E.R. (2017, updated 2020) Modifications to the ICER value assessment framework for treatments for ultra-rare diseases

Other aspects of modelling

Discounting counting of costs and benefits to present value may disadvantage treatments in situations where all of the costs are upfront, but the long-term benefits of enabling an individual to live a normal life are heavily discounted. Therefore, consideration of a lower discount rate than the usual established in an economic base case may be considered. NICE has clear criteria that need to be fulfilled for this to be considered but these are not straightforward to apply (TA Methods guide 2013, section 6.2.19). For example, there has been debate about whether the treatment really achieves "normal life" and what that means (Strimvelis in severe combined immunodeficiency syndrome and voretigene neparvovec, NICE).

Focussing on the key drivers of the economic model, each element should be critiqued indicating any concerns, weighing options about what is feasible in the RD, what is clinically plausible and then stating whether an assumption can be found that is "sufficient for decision making" (NICE, CADTH).

A probabilistic sensitivity analysis that characterizes uncertainties in terms of probability distributions around parameters may be helpful. However, the PSA does not capture the correlations between parameters and can be difficult to explain. So, in addition, a range of sensitivity and scenario analyses should be performed for alternative assumptions so that the appraisal committee can determine the most plausible ICER and consider the associated uncertainty.

A sensitivity analysis with a perspective wider than the NHS is also suggested to consider wider costs and benefits relevant to the patient and carer such as out of pocket expenses, lost earnings and carer quality of life gains from new treatment.

For key uncertainties, a threshold analysis to show at what level cost effectiveness is achieved could be helpful.

If a cost utility model cannot be created that is it sufficiently reliable for decision-making, alternative models should be considered such as cost effectiveness (based on a natural outcome, such as life-years) or cost consequence (balancing multiple relevant outcomes).

Where possible, the economic model should be explained in as plain language as possible to enable clinical experts to contribute to discussions about assumptions and key uncertainties.

Given the complexity of models, clinical engagement may need to be an iterative process - scoping, after submission, after initial critical assessment, before appraisal etc.

Areas that could be improved in the economic model should be clearly identified so that research can be undertaken to improve assumptions in the models of future treatments for the rare disease.

B. Structured Appraisal Deliberation considers all dimensions of value

HTA appraisal is a deliberative process to identify all aspects relevant to value, align on assumptions and resolve as many uncertainties as possible to come to an informed and fair judgement about whether to recommend/decide if an RDT can be used/reimbursed in a health system. This cannot be done formulaically, but processes can be put in place to ensure that an appraisal committee systematically addresses each element of the decision-making domain (or indicates why it is not relevant), receives information on the modifiers and other considerations relevant to the individual RDT, leaving space for wider social value judgments.

5. Appraisal committees are bespoke for RDTs, or general appraisal committees include several rare disease specialists

If a separate appraisal route is used for RDTs, the appraisal committee should include members that have experience in treating, commissioning, supporting or studying issues related to rare diseases. It is particularly important to include paediatricians who treat patients with rare conditions, specialists who treat adults, clinical geneticists, commissioners, two patient group representatives[4] and an individual with ethics/philosophy experience to help the committee balance different arguments. Given the consistent reporting of psychological issues for patients, carers and family members, a psychologist would also be helpful. Ideally, clinical experts should have a connection with international experts, for example via the European Reference Networks[5].

For general appraisal committees, there should be at least one adult and one paediatric rare disease specialist and one rare disease patient group representative. In such a committee, when appraising an individual RDT, other members should be asked if they have experience of treating patients with the condition as given the many clinical manifestations of a rare disease, specialists such as cardiologists, metabolic specialists etc can provide valuable insights (Patisiran – importance of cardiologist input).

Committee members draw on their own expertise and focus on different elements in submissions to prepare for committee discussion. They depend on the expertise of other members, for example “to interrogate the economics”. Therefore, it is important to have the right mix of expertise on an appraisal committee and have processes in place to overcome gaps in expertise if individuals with key knowledge are unable to participate (e.g., by allowing deputies to attend or asking for specific feedback on certain issues to input to the meeting).

Each appraisal committee member needs training when they start. This should go beyond the methodological basics of the clinical and cost effectiveness analysis techniques, to explain the decision-making framework, how decisions are made. This should include guidance in the application of decision modifiers, and training in the principles of accountability for reasonableness (Daniels and Sabin 2008) to support members in the exercise of discretion and judgement. Their role and how their knowledge and experience can be utilised optimally to contribute to the committee’s deliberations should be discussed.

To ensure that members have time to read the substantial set of papers associated with an RDT (could be 500-1000 pages), members should be given protected time or recompensed for preparing for and attending appraisal meetings.

Committees should be learning organisations that regularly review their decisions, any challenging situations that have arisen and improvements that could be made to process.

6. The deliberative appraisal discussion is driven by the domains of decision-making and use of modifiers is clearly understood

The chair of the appraisal committee should give a clear introduction to each RDT outlining in what category the RDT sits and any decision modifiers that apply or will be discussed. Given the complexity of issues that arise with RDTs and the wide-ranging nature of a good deliberative discussion, the chair and HTA staff should ensure all elements of the decision-

[4] e.g., one umbrella rare disease organisation such as for genetic diseases and one that covers a range of rare diseases, such as rare kidney or metabolic diseases

[5] Work of the ERNs | Public Health (europa.eu)

making framework and modifiers are discussed carefully at some point, before a decision is made. Some modifiers may feature in presented evidence (e.g., carer burden may be captured in the economic model) and care is needed to avoid double-counting (or double dispensation).

Senior members of the committee can play an important role, laying out any disparities between evidence and inputs and indicating what assumptions the committee needs to consider.

There needs to be clarity about how the deliberative appraisal process will consider each decision-making domain and if the domains are equally balanced. Our experience shows that even if domains beyond clinical and cost effectiveness are considered, one of these (such as clinical effectiveness, or cost effectiveness) has most weight and if the decision in that domain is in the balance, the other domains can tip the decision (favourably or unfavourably). The final determination of value in the HTA report should document the key considerations of the committee in relation to the domains and how they were balanced to reach the recommendation.

The elements of patient benefit that have not been studied in clinical trials or weaknesses in disease representation in the economic model should be clearly articulated so that committee members can understand the limitations of the evidence in relation to real-life. This enables consideration of the overarching adequacy of the evidence to demonstrate value, alongside the methodological critical assessments of the evidence. For example, these gaps were noted in our observations: avoiding fear of pain associated with acute pancreatitis (volanesorsen, familial chylomicronaemia syndrome), lost schooling (voretigene, sight loss). The committee can then judge how important these gaps are and consider how that impacts their view of the evidence.

Discussion of uncertainties and their implications will be at the heart of the deliberation. These are addressed in recommendation 7.

If other factors come into play, such as trade-offs, these should be indicated. For example, most committees focus on cost effectiveness, not affordability (budget impact). However, some of our interviews indicated that budget impact is traded off against uncertainty, i.e., if there is expected to be a large impact on health service resources, less uncertainty is accepted (interview #25).

In interviews, concerns were raised about setting precedents in appraisals particularly in relation to application of flexibility with methodological standards for RDTs ([redacted]). To address this, the reasoning for taking a flexible approach needs to be documented. This should not be taken as an absolute and narrow precedent for future decisions, as every RDT has a unique context.

The manner in which the committee makes a decision should be clear to the public (consensus, confidential or open voting, etc).

Appraisal reports should clearly show how appraisal committees considered the evidence and inputs in each domain and the decision-modifiers. Checklists such as those used by CADTH may be appropriate, or for each important issue there should be a statement about the committee's preferred assumption or consideration (see NICE Committee Considerations in their HTA reports).

7. Uncertainties are characterized in terms of form, extent and implications for decision-making

In health economic modelling different forms of uncertainty are characterised, as discussed in the NICE TA Methods Guide, 2013). We present these as:

Type of Uncertainty	Meaning
Decision	Probability a different decision would be reached if the true cost effectiveness of each technology could be ascertained before making the decision
Structural	Uncertainty around structural assumptions in the analysis - e.g., categorisation of health states, assumptions about extrapolation of costs and outcomes beyond trial follow-up) Handled with scenario analyses
Methodology	Uncertainty about appropriateness of methods Handled with sensitivity analyses
Parameter	Choice of data sources to provide values for key parameters such as different costs and utilities, estimates of relative effectiveness and duration of effect Handled with sensitivity analysis
	Parameter precision (after data sources decided) – uncertainty around mean health and cost inputs should be characterised by distributions and probabilistic sensitivity analysis used.

Uncertainties should be clearly characterised and explained. As outlined in recommendation 4, stakeholders should be engaged iteratively during the HTA process to resolve the decision-relevant uncertainties that have most influence on the model. Sensitivity, scenario analysis and PSA can be used to show the impact of different assumptions.

In our interviews, some committee members were keen to have a quantification of uncertainty (interview #23). There is a judgement about the level of uncertainty at CADTH for the clinical and cost effectiveness evidence ([redacted], but this categorisation is a high-level judgement that may not adequately capture the multi-dimensional and complex nature of uncertainty. Alternatively, it would seem better to characterise the uncertainties, identify those most relevant to the decision-making and gauge whether the MAH produced the best evidence possible within the context of the rare disease.

Furthermore, the uncertainty associated with an appraisal process arises in forms other than those defined in economic modelling.

> *"I do try to come with is a perception of the ongoing uncertainties. And by that I don't mean the uncertainties in the economics. The argument for the product may be. . . the upper end ICER might be horrendously uncertain. But that's a certainty….. It's just a shame that the two words are the same. It's coming … to the meeting, with what they're not quite sure about, what is the committee actually trying to communicate that is the more plausible ICER? How plausible or otherwise are some of the upper end ICERs? What could be the true risk to the Scottish system if this product were accepted?"*
>
> *Interview #2, SMC Member*

There are uncertainties that cannot be quantified but are known for RDTs, such as lack of knowledge about the condition (meaning that disease progression and best outcomes for study are unclear), the disease population in the jurisdiction and the implications of establishing a new care pathway. These considerations should contribute to the appraisal deliberation and stakeholder input on these elements may be particularly valuable.

Our limited analysis shows that common decision-relevant uncertainties for RDTs relate to:

- disease course (states of the model and transition probabilities)
- (sub-)population to be treated
- real-life effectiveness
- duration of treatment effect
- how long should treatment be given in a chronic condition } (utilities)
- how will treatment be delivered.

(The time horizon and discount rate for benefits always impact an economic model, but these are often stipulated by the HTA body.)

There can also be committee uncertainty, when members of the committee have different views when judging evidence or interpreting criteria for leniency. [redacted]

We also observed how miscommunication might lead to new "uncertainties", as when an element that was not explained well, or was not understood by the committee might be subsequently labelled as an uncertainty and treated as such (interview #17).

For patients and carers, uncertainty is a very different concept – often there is major uncertainty before a diagnosis, but perceived certainty of treatment effect for an individual patient or their carer (e.g., nusinersen). More needs to be done to consider the different aspects of uncertainty and we await the HTAi Global Policy Forum work on this topic.

Finally, given the high prices of some RDTs it must be recognised that even if all uncertainties could be resolved, difficult decisions would still have to be made to ensure fairness in the whole health system (interview #9).

8. Outcomes-Based Managed Entry Agreements may be used to resolve decision-relevant uncertainties, if collection of sufficient data is feasible

The purpose of OBMEA is to enable reimbursement of a promising health technology such as an RDT, whilst collecting data to resolve important uncertainties that can inform subsequent re-appraisal.

OBMEA have often been unsuccessful, not collecting sufficient data to inform re-appraisal. However, in our discussions with HTA/payer bodies, many lessons have been learnt. The most important is that there must be a purposeful approach to OBMEA with clear definition of the decision-relevant uncertainties to be resolved. This requires some form of "covenant" among all participating stakeholders that they will do their utmost to contribute to the data collection (including MAH, clinicians and patients). Then a document is needed that clearly outlines the purpose of the OBMEA including the uncertainties to be resolved, the data sources (such as registries, administrative data, prescribing data etc) and how the data will be gathered and analysed. In addition, attention must be paid to ongoing monitoring of the data collection to ensure sufficiency, completeness and quality.

To support this, D10.3 provides the following tools:

- Checklist to consider when an OBMEA is feasible for an RDT
- Template for the agreement including the data collection elements alongside any legal requirements and organisational considerations
- Terms of Reference for a monitoring committee
- Patient group submission for re-appraisal after OBMEA.

C. Underpinning: Iterative expert clinical and patient input throughout the process

Clinical and patient experts are involved throughout the appraisal process to explain the context of the condition, existing care pathway and help resolve uncertainties related to the determination of treatment value

Our observations show clearly that patient group and expert clinical input can help explain the nature of the condition and current pathways. Clinicians can help interpret clinical trial effects compared to what might be expected in clinical practice. Patients and patient groups can share experiences and impacts of treatments. This expert clinical and patient knowledge is often elicited through submissions or statements or by inviting individuals to meetings. There are more robust ways to do this – through developing robust patient-based evidence and we encourage MAHs and HTA bodies to do this.

Such stakeholder involvement is recommended internationally in all HTAs, but is particularly valuable for rare diseases, given the paucity of clinical evidence and committee knowledge already discussed.

There is a concern from some involved in HTA that clinical and patient experts may be biased (particularly by MAH influence), only advocate for the treatment and can only provide anecdotal input from their own perspective. These issues can be overcome by managing conflicts of interest according to organisational policies and by targeting their input to make best use of their knowledge.

In the following we consider where clinical and patient input might be most valuable for RDTs and then how to optimize each form of input.

Role at each stage of the appraisal process

<u>At the outset of the appraisal (scoping)</u>

Clinical and patient experts should be involved in HTA scoping discussions/education of HTA assessors to explain

- the nature of the condition (pathophysiology of condition, condition states and progression, mechanism of action of new treatment)
- the care pathway, unmet needs and important outcomes
- any experience of the treatment in a clinical trial or early access.

This information will help in critiquing the clinical evidence and economic model.

Ideally this input should be focussed on the specific population considered in reimbursement with a presentation from the patients and clinicians to inform the HTA parties, rather than being restricted to questions that HTA assessors want answered. This allows information to emerge that might not have been anticipated. The input could be a presentation from the clinicians about the condition, current management and important outcomes in the disease pathway. It could include patient stories of those who are in the specific population being considered. The patient stories and clinical input should preferably be videoed so that new assessors joining the team can review the video as a starting point.

During critical assessment of evidence

Clinical experts should be consulted at the time of critical assessment about

- how treatment effects in the clinical studies can be interpreted in real-life, for example
    - the prognostic value of an outcome and any evidence for surrogacy
    - whether an effect would be expected in the timeframe of study
    - which outcomes are most important (pulmonary exacerbations more important than small changes in lung function in cystic fibrosis)
    - what causes variability in outcomes, can they be used to determine response or as a discontinuation criterion
- the validity of important modelling assumptions relating to treatment effect, for example
    - explaining mechanism of action of treatment in relating to the condition (voretigene neparvovec – once the gene is in the viable retinal cell it's there for life, patisiran – halting the amyloid build-up)
    - duration of treatment effect
- the assumptions in the economic model including constructs and inputs.
- health service impacts in terms of treatment administration and patient monitoring requirements

During appraisal deliberations

Clinical and patient experts should be able to contribute to deliberative appraisal meeting discussions about the evidence to help resolve key uncertainties.

Clinical experts can provide important insights to questions such as

- likely number of eligible patients in different sub-populations,
- optimal treatment positioning,
- duration of treatment effect
- treatment continuation/stopping protocols
- organisational/infrastructure issues relating to health service readiness.

This might include balancing of clinical trial evidence with national experience, e.g., in early access (such as for volanesorsen).

The meaning of utility values for each state should be discussed with patients to ensure they have face validity (for inherited eye disorders blindness had a utility worse than death, this is not a view held by patients, for other chronic rare diseases baseline generic PRO measures are often better than the general population as patients have their own "normal" and learn to live with their condition).

Considerations to optimize process of clinical and patient involvement

In our observations, we heard from clinical experts who did not understand how appraisal decisions were made or what their role would be. Patient groups were generally better prepared about the process, but afterwards reflected that if they had understood what the discussion was going to focus on, they could have prepared better and provided more helpful input (interview #15).

Stakeholders participating in the appraisal need to be able to access training (perhaps online) about the rationale for appraisal, the process and most importantly how they can make most impact. The framework for decision-making should be explained and stakeholders should be encouraged not to advocate for a treatment, but to provide balanced insights – facts and experiences - to contribute to the discussion relating to each domain of the decision-making framework.

MAHs should work on creating a plain language explanation about the condition, natural history, the clinical trial that has been undertaken, assumptions that have been made about long term effects, the structure of the economic model and an honest presentation of the major assumptions and gaps. This would then help all stakeholders consider how they can help resolve the gaps.

Rare diseases may not have specific patient groups and may be covered by umbrella groups that have less specific knowledge/fewer contacts with specific patients. Consideration of the burden on patient groups should be considered and approaches for involvement streamlined (Facey et al. 2018). Ideally, the MAH or HTA body should undertake a literature review of patient issues (including patient group submissions in other countries/other related diseases) or conduct interviews or focus groups with patients. MAHs should also commission relevant research such as qualitative research studies within clinical trials to develop robust patient-based evidence (Rand et al. 2019) (Facey and Hansen 2015) that shows a range of experiences and views. HTA bodies should include staff who can critically appraise patient-based evidence and who can explain the quality of such evidence.

An HTA body should identify/gather evidence about patient, carer and family issues and seek patient group input to confirm/refute literature findings, address outstanding issues and discuss the key issues in the appraisal.

Clinical and patient experts should be given sufficient notice to contribute to all discussions, with as much information as possible and be given time to provide evidence/information from clinical practice/patient experience to help resolve uncertainties. For example, provision of local audit data (cystic fibrosis) or early access data. This may require a lengthening of the appraisal process, but could provide benefits if repeated appraisal deliberations are avoided.

Some HTA bodies present quotes from patient group submissions to the appraisal committee, but our research shows that these are seen as anecdotal by many appraisal committee members. A balanced presentation explaining the sources for the patient group submission and a diversity of views in the relevant patient population is seen as more informative.

As the deliberative appraisal discussion progresses, it can be helpful for a committee or staff member to take responsibility for checking whether the stakeholder submissions confirm or refute assumptions under discussion.

When stakeholders are involved in an appraisal meeting, they should be sent the redacted committee papers in advance, with some guidance on the flow of the meeting and where they will input. It could be helpful to have a nominated staff member to support them at the meeting (we have seen this for patient groups, but not for clinicians). There should be a clear slot in the meeting where they know they will be asked for their input. Then it needs to be clarified whether they can get involved in the appraisal committee discussion and ask questions/make comments or if they have to wait to another specific point to contribute further.

There should be measures in place to deal with the emotional issues that might arise, such as recounting the death of a child.

The appraisal committee chair may need guidance on how to sensitively engage patients/carers in the meeting.

We have seen some very long appraisal meetings, where there are limited breaks and a feeling that the table cannot be left. If patients are invited to such a meeting all steps should be taken to accommodate all their medical, accessibility and well-being needs (e.g. provision of papers in a format that patients can access in sufficient time in advance of the meeting, written directions on how they will be involved in the meeting, seating space for carers and assistance dogs, an accessible toilet (preferably a changing place), organising several breaks during the meeting, provision of snacks and food that meet dietary requirements and water for assistance dogs).

In the HTA report, the form of expert clinical and patient input should be clearly documented with its source. The influence of specific elements of the input on the critical assessment and final appraisal recommendation should be reported (see NICE and CADTH).

Patient Group Submissions

Several HTA bodies (in Canada, England, France, Scotland, Spain, USA) use a template to enable patient groups to submit information about living with the condition, experience with current and new treatments and expectations of new treatment. An example template for all medicines is available on the HTAi website in various languages, and is suitable for RDTs.

Patient groups should clearly explain the sources of information in their submissions (survey, questions to Facebook group, interviews with patients who received treatment (in clinical trials or via early access, focus group, registries etc). When statements are made, they should be referenced by their source.

Some patient group submissions have used photos (volanesorsen – amount of fat taken from the blood during plasmapheresis, healed lesion of a skin condition) and some videos (undertaking the functional vision test with voretigene neparvovec, a pre-symptomatic infant with spinal muscular atrophy walking at an early age after onasemnogene abeparvovec). We believe that, in some conditions, photos and videos could be helpful to explain the nature of condition and show impact.

If possible, information should be gathered from the specific patient population considered in the appraisal. If a wider population is used, comment should be made about the applicability to the population of interest. The extent to which the sources cover the diversity of the population should be stated (CF Trust registry has 99% coverage in UK) and how patients were identified for inclusion.

There is often a feeling from committee members that patient group submissions are “cherry picking” positive experiences. It is more convincing if a balance of views is presented, rather than being entirely positive about an RDT.

Often there are uncertainties about how long to treat patients and whether a stopping rule should be applied, so exploring issues with patients that discontinued treatment can help understand how to optimize treatment delivery.

It may be possible to take information from patient group submissions in other countries to describe the nature of the condition and existing treatments – adding elements that are different in the relevant jurisdiction.

The appraisal committee struggles with emotion, but facts can be very hard hitting and change views. It is important that patient groups explain what is particularly challenging about this disease. For example, “30% indicated that diarrhoea is so bad that they do not have confidence to leave the house.” An important input from patients is to understand what outcomes they would like to see with a new treatment.

There seem to be more psychological issues documented with non-cancer rare diseases. There could be many reasons for this such as:

- hereditary nature
    - means a patient can see another family member decline and predict their own future (amyloidosis)
    - multiple patients in one family can create competing requirements – e.g., two children requiring different hospital visits
    - may affect decisions about future pregnancies
- lengthy time living with a condition that is not diagnosed
- major unmet needs – no treatment
- in a long-term condition, fear of an acute episode that will lead to the next stage of the disease (familial chylomicronaemia syndrome - fear of eating due to risk of acute pancreatitis)
- treatment regimens that take many hours every day (cystic fibrosis) and thus lead to exclusion from normal social activities/impact work, that separate the individual from others (cystic fibrosis – infection risk, familial chylomicronaemia syndrome – eating restrictions)
- rapidly progressive diseases diagnosed in adulthood (such as motor neurone disease) that have major loss of function quickly leading to loss of “self” and feeling of burden on family.

In discussions and interviews, we have heard these psychological issues raised, but rarely have we seen these details articulated in submissions. Patient groups may wish to consider gathering this information and more research is needed to explore the extent of psychological issues and if they should impact decision-making.

Our research shows that patient quotes presented at appraisal committee have little impact, but we know that HTA bodies recommend their inclusion. An alternative would be the use of patient stories. Hearing from the patient in their own language about the impact on them and their family of the condition, the current treatments and new treatment can be impactful (Patisiran, Orkambi – Quest for CF Cure). One approach is to outline what a typical day in the life of a patient is like and how it changes if there is an acute exacerbation. This could be documented in the patient group submission or given as a verbal patient story to start an appraisal meeting. An example in cystic fibrosis is shown.

The following is a typical daily treatment routine for a cystic fibrosis patient. The regime is unrelenting and burdensome for both patients and parents/carers

(Excerpt from a case study).
7.30 Mucoclear neb 20 mins
8.00 Physiotherapy percussion with acapella and niv attached 20 mins
8.20 Tobi/colomycin neb monthly alternated 15 mins
8.30 Wash nebuliser pots & dry with hair dryer 10 mins
8.45 Check blood sugars & inject background insulin 3 min
9.00 Breakfast, I often feel sick in the mornings but I have to still eat for vital calories and to prevent a hypo, take enzymes, 3 x vitamin A&D capsules, omeprazole for acid
reflux. azithromycin prophylaxis antibiotic, iron tablet, sertraline for depression, uniphyllin for airways, diclofenac for arthritis 5 mins
11.00 Mid morning snack with enzymes
13.00 Lunch with enzymes, iron & diclofenac
13.10 Inject fast acting insulin 3 mins
15.00 Ventolin inhaler, physio with acapella and niv attached 20 mins
16.30 Mid afternoon snack with enzymes
17.00 Comcivent neb 10 mins
17.10 Pulmozyme neb 10 mins
Have to wait an hour before I can do physiotherapy after inhaling pulmozyme
18.20 Physiotherapy percussion with acapella and niv attached 20 mins
18.45 Colomycin/tobi neb 15 mins
19.00 Wash and dry neb pots with hair dryer 10 mins
19.15 Dinner with enzymes & milkshake for bones, diclofenac & iron tablet
22.00 Supper with enzymes.
In addition, I typically have 6-8 courses of intravenous antibiotics (IV's) per year alongside my usual medication and sometimes a third IV antibiotic is added. They are intended to be 2-week courses but I am typically on them for 4-5 weeks nowadays as they have little or no effect. I have developed resistance and allergies to all the IV's and have to be admitted to hospital to be desensitized as I still have to have them even though my body is rejecting them.

Patient Group Submission to SMC from Quest for a Cure for CF, Reproduced with permission.

The burden put on patients and patient groups in contributing to the appraisal process needs to be carefully considered. All those submitting should be asked what effort and expenses they have incurred to be able to contribute to the meeting (time, commissioning of research etc). The HTA body should evaluate what has made a difference to committee deliberations, give feedback to those who contribute and provide training to them. The HTA body should reimburse the cost of attendance at an appraisal meeting and ideally pay a fee for patients' and patient groups' contributions to the HTA.

Expert clinical involvement

Assessment should involve at least one, or preferably more, clinical experts throughout the process to advise those assessing the evidence. Ideally clinical expert should commit to the entire process to be able to respond to questions that arise in interpretation of the evidence, not just respond to questionnaires or be involved at the final stage.

For complex diseases, for which there is little experience, an expert panel could be created from across the jurisdiction that includes those with experience in different clinical settings (as is done in CADTH). It may also be necessary to involve experts from neighbouring jurisdictions or wider networks, such as European Expert Reference Networks. Conflicts of Interest should be documented and balanced. These panels could be used to describe the current clinical management of patients, characterize unmet therapeutic needs, explore potential place in care pathway of new treatment, discuss system readiness and potential implementation challenges (such as genetic testing, availability of experts to diagnose etc).

V. Discussion

Our recommendations outline considerations to ensure a thorough but fair appraisal process for RDTs that takes account of the paucity of clinical evidence and knowledge. It is an appraisal process that can be used with all treatments. This overcomes “chasms” where there are very different thresholds between different appraisal programmes (interviews #21, #22). It is more of a continuum with layers of flexibility that are particularly needed for rare diseases, particularly those with high unmet need and very low prevalence. Interestingly, the New Zealand decision-making framework was planned to be for RDTs, but after wide-ranging public consultation they ended up, like us, creating an appraisal framework for all treatments that is suitable for RDTs.

We believe that if a detailed decision-making framework is defined with modifiers to support flexibility in appraisal, and if these drive all parts of the process, then all those involved (stakeholders and appraisal committee members) can better contribute evidence and inputs. This will lead to a stronger evidence base for decision-making and best use of expert knowledge to resolve uncertainties and inform value judgements about an RDT. Different deficiencies in evidence, care pathways and access to expertise in a health jurisdiction, means this is not a rigid framework; deliberation and hard choices will still need to be made.

This framework will support consistency of flexibility in appraising RDTs to ensure fairness, within a framework of accountability for reasonableness (Daniels and Sabin 2008) – delivering a fair process – consistently including all relevant evidence, knowledge and considerations to make the best decision possible.

This guidance provides recommendations for HTA bodies MAHs, clinical and patient experts. Importantly, it can also inform evidence generation plans for the pharmaceutical industry far in advance of the appraisal, so that they bring the best evidence possible to HTA.

We recommend this framework be taken and adapted for use in an individual jurisdiction or collaborative HTA group. In particular we note the plans for the EU HTA collaboration’s work on joint assessments will focus on OMPs and so we ask for their consideration of this work. Furthermore, we note the voluntary actions possible on Post Licensing Evidence Generation and recommend D10.3 to support that work.

This deliverable WP10 has built on novel ethnographic research across a range of RDTs and HTA systems to better understand the deliberative appraisal process. This has led to a framework with recommendations and detailed guidance that proposes augmentation of the entire HTA process beyond clinical and cost effectiveness evidence, to include other forms of evidence and inputs. Then in appraisal, the framework guides consistent application of the appropriate expertise and flexibilities needed in assessment and deliberation when considering paucity of clinical evidence and limited understanding of the rare disease. If used, this will support robust and fair appraisal of the value of an RDT. This framework and guidance could be used by any HTA body appraising RDTs, or by a new pan-European HTA network.

Appendix 1. HTA Methods Guidance on Handling Appraisal of Treatments for RDs

1.1 Methods for RDTs (OMP definition)

Web search in 2020 based on ISPOR website of pharmacoeconomic guidelines, review of HTA body websites to find up-to-date material, links provided in WS1 country vignettes, excluding England and Scotland that are reported in MS41.

Country	Clinical Effectiveness	Economic Modelling
Belgium (2017 KCE report on improving MEA)		It should also be noted that for …orphan drugs, a cost-effectiveness evaluation is not legally required (*reimbursement conditional on approval of advisory clinicians*) and was usually not done… *(There was a recommendation that for orphan drugs, even if the uncertainty around the input variables might be much higher, the evaluation of the ICER also remains a key parameter, an economic evaluation should also be legally required and be used as an evaluation criterion. But this was not accepted.)*
Canada (2017)	Nothing	Nothing
EUnetHTA (2020)	*In relation to definition of comparator, they note that up to date international clinical guidelines may not be available for rare diseases.* The use of a surrogate endpoint may be acceptable in some cases such as rare diseases, where big samples are not available. However, the validity of the surrogate depends on the empirical evidence, not on the size of the target population. Therefore, also in this case, the use of surrogate endpoint should be justified and discussed.	
France (2020)		*Some statements about utilities and rare events (in French)*

1.1 *continued* Methods for RDTs (OMPs)

Country	Clinical Effectiveness	Economic Modelling
Germany - IQWiG (2017/2020 draft)	**3.2.5 Benefits and harms in small populations** In small populations (e.g., patients with rare diseases or special subgroups of patients with common diseases), there is no convincing argument to deviate in principle from the hierarchy of evidence levels. In this connection, it is problematical that no international standard definition exists as to what is to be understood under a “rare” disease. Independent of this, patients with rare diseases also have the right to the most reliable information possible on treatment options. Non-randomized studies require larger sample sizes than randomized ones because of the need of adjustment for confounding factors. However, due to the rarity of a disease it may sometimes be impossible to include enough patients to provide the study with sufficient statistical power. A meta-analytical summary of smaller studies may be particularly meaningful in such cases. Smaller samples generally result in lower precision in an effect estimate, accompanied by wider confidence intervals. Because of the relevance of the assumed effect of an intervention, its size, the availability of treatment alternatives, and the frequency and severity of potential therapy-related harms, for small sample sizes it may be meaningful to accept a higher p-value than 5% (e.g., 10%) to demonstrate statistical significance, thus increasing quantitative uncertainty. Similar recommendations have been made for other problematical constellations. Such an approach must, however, be specified a priori and well justified. Likewise, for small sample sizes it may be more likely that is necessary to substitute a patient-relevant outcome that occurs too rarely with surrogate endpoints. However, these surrogates must also be valid for small sample sizes. In the case of extremely rare diseases or very specific disease constellations, the demand for (parallel) comparative studies may be inappropriate. Nevertheless, in such cases it is also possible at least to document and assess the course of disease in such patients appropriately, including the expected course without applying the intervention to be assessed (e.g., using historical patient data). The fact that a situation is being assessed involving an extremely rare disease or a very specific disease constellation is specified and explicitly highlighted in the report plan.	

1.1 continued **Methods for RDTs (OMPs)**

Country	Clinical Effectiveness	Economic Modelling
Hungary (2017)		Some technologies are not cost-effective at this ceiling value (*3xGDP/capita)* (e.g. (orphan) therapies for the treatment of rare diseases). However, cost-effectiveness is one of the aspects in reimbursement decisions (e.g., efficacy, effectiveness, equity, budget impact) and needs to be considered together with these other aspects.
Ireland (2019)	Nothing	Nothing
Netherlands (2016)	Results of RCTs are not always available in the literature. In addition, an RCT also has it downsides and limitations. Certain situations may arise, therefore, in which non-randomized or non-comparative studies will suffice or even may be preferred, for example when there is a clear dose-response relation, the natural course of a condition is known or in the case of rare diseases.	

1.2 Methods for RDTs (Smaller populations than OMPs – "Ultra-Rare")

Country	Definition	Clinical Effectiveness	Economic Modelling
Australia (2016)	"rare" <1/50,000	Nonrandomised studies may provide useful information when randomised trials are not feasible (i.e., when the disease or condition is rare) Where the rarity of the disease or condition prohibits the use of a traditional parallel-group randomised controlled trial, alternative trial designs may be acceptable (e.g., randomised crossover trials, including n-of-1 trials and trials with a randomised adaptive design). Such trials require a protocol, a clinical trial registry number or identifier, and a design that involves a randomisation procedure. Where a submission is based on such a trial, risk of bias can be addressed as for randomised trials. The best approach to assessing the validity of single-arm studies will depend on the design of the study. Justify the approach (or modifications to the approaches below) taken to capture the key limitations of the study design.	
New Zealand	"rare" 1<1/50,000 (~90 people) in NZ For this and all authorised indications and those in phase III trials	Extent of information available for analysis: Pharmaceuticals for rare conditions are more likely to undergo rapid analysis due to unavailability of data. More detailed analysis may not resolve key uncertainties.	

1.2 *continued* Methods for RDTs *("ultra-rare")*

Country	Definition	Clinical Effectiveness	Economic Modelling
Norway (2018)	<u>Very small patient group with extremely severe condition and considerable expected benefit</u> It is not useful to set absolute conditions for evaluating whether the requirements for "very small patient groups ", "extremely severe conditions " or "considerable expected benefit ", is fulfilled. There should, however, be indicative criteria for decision-making. 1. Very small <1/100,000 on a global basis per pharmaceutical *(all indications)* AND <50 pats in Norway per pharmaceutical (steady state prevalence[6]) 2. Extremely severe Level of severity measured using absolute shortfall corresponding to at least 30 good life years (QALYs)[7] 3. Considerable expected benefit Expected benefit is considerable and a minimum of 2 gained good life years compared to standard (QALYs) All 3 indicative criteria should be fulfilled. The criteria are indicative in nature and must be assessed in accordance with an overall assessment in every specific case. In some cases, it may, at a later date, be relevant to re-evaluate how far the indicative criteria have been fulfilled.	Lower requirement for documentation = Documentation should be as good as it can be. Submission of documentation should to the greatest degree possible follow the recommendations in the general STA guidelines…. Considerable uncertainty in the documentation or calculation methods will lead to a lower prioritisation in decisions about new pharmaceuticals….There is, however, a requirement that the documentation presented is the best that can reasonably be expected, given that it is a very small patient group with extremely severe conditions. The link between outcome measures used in studies and effects of future morbidity, or death, must be sufficiently substantiated. NOMA can involve Norwegian clinical experts in such evaluations. Even if a pharmaceutical qualifies for consideration under this arrangement, the decision-maker can conclude that the pharmaceutical is not to be introduced on the grounds of documentation that is *too* poor or inadequate.	Higher level of resource use than normal may be acceptable

[6] Diseases can become less rare when an effective treatment emerges and diagnosis improves.

[7] Very stringent restriction, meant to cover children with congenital conditions, where higher use of resources could be justified by argument of greater shortfall, but review of 19 cases in Norway from 2014-2017 shows only 2 cases that reached 20 years shortfall, none reached 30 years shortfall.

1.2 *continued* Methods for RDTs *("ultra-rare")*

Country	Definition	Clinical Effectiveness	Economic Modelling
USA (2020) Issued for public consultation	Total population < 10,000 with no ongoing or planned clinical trials for a patient population >10,000	ICER will not change its approach to rating evidence according to the ICER EBM matrix, nor will there be different "standards" of evidence. Instead, ICER will provide specific context regarding the potential challenges of generating evidence for these treatments, including considerations of challenges to conducting RCTs, to validating surrogate outcome measures, and for obtaining long-term data on safety and on the durability of clinical benefit. The commonly used approach of evaluating treatments for ultra-rare diseases against historical controls will be highlighted. This added contextual language will be highlighted through special formatting in ICER reports and retained throughout press releases, executive summaries, and other versions of ICER reports.	For assessment of cost-effectiveness of a treatment for ultra-rare diseases, ICER will seek to produce a cost-effectiveness model for every new treatment, acknowledging and highlighting additional uncertainty in translating patient outcomes into quality-adjusted life year (QALY) or equal value of life year gained (evLYG) measures. For all treatments, including those for ultra-rare diseases, ICER will provide willingness-to-pay threshold results for from $50,000 per QALY/evLYG to $200,000 per QALY/evLYG. No special quantitative weighting system will be applied to different magnitudes of QALY gains or to baseline severity of the condition. ICER will calculate a health-benefit price benchmark for these treatments using the standard range from $100,000 to $150,000 per QALY/evLYG, but will add language in all report formats indicating that decision-makers in the US and in international settings often give special weighting to other benefits and to contextual considerations that lead to coverage and funding decisions at higher prices, and thus higher cost-effectiveness ratios, than applied to decisions about other treatments. When the impact of treatment on patient and caregiver productivity, education, disability, and nursing home costs is substantial and these costs are large in relation to health care costs, ICER will present its base case health system perspective model results in tandem with the results of a scenario analysis inclusive of broader societal costs. This will most often occur in cases where the incremental cost-effectiveness ratio changes by greater than 20%, greater than $200,000 per QALY, and/or when the result crosses thresholds of $100,000-$150,000 per QALY. Similarly, a health-benefit price benchmark (HBPB) linked to the societal perspective analysis will be presented alongside the standard HBPB. When there are challenges translating the outcome measures used in clinical trials and available patient-reported data into QALYs, ICER will conduct a search for "mapping" studies that may allow translation of surrogate outcomes into quality of life measures. The validity of these mapping studies will be discussed with manufacturers, clinical experts, the patient community, and other stakeholders in order to get their input on the most feasible way to translate these other measures of patient outcome into QALYs. <u>R&D costs template</u> ICER will postpone all efforts to develop a formal template for research and development costs. Further discussion will be sought with stakeholders on the prospects for forming a multi-stakeholder workgroup to evaluate the options for development of some kind of formal template. In lieu of a formal template, ICER will invite every manufacturer of a treatment under review to submit whatever information the company may wish to submit on development or manufacturing costs for inclusion in a new dedicated section of the ICER report. If the manufacturer believes that development or manufacturing costs are important considerations in justifying the price for their product, it is hoped they will submit information to support this assertion. No editing, judgment, or analysis will be performed by ICER on any information submitted.

1.3 Methods that may be relevant for RDTs

Country	Definition	Clinical Effectiveness	Economic Modelling
Italy		As established by the 2017 Budget Law[8], innovation status is judged for some products (including all OMPs). This can lead to a classification of "innovative" that is in place for a maximum duration of 36 months. For "conditionally innovative" or "potentially innovative" re-evaluation occurs after 18 months. In both cases there is immediate inclusion in regional formularies. The innovation status is evaluated based on 3 elements: - Therapeutic (unmet) need: absent, low, moderate, important, maximum - Added therapeutic value: absent, low, moderate, important, maximum - Quality of evidence (according to GRADE): very low, low, moderate, high To achieve the "innovative" rating, the unmet need and added therapeutic value must be maximum or important, generally with high quality evidence. However, for OMPs, there is an exception on the quality of evidence.	

[8] Law 11/12/2016, n. 232 (Legge di Bilancio 2017) website. https://www.gazzettaufficiale.it/eli/id/2016/12/21/16G00242/sg Accessed August 7, 2020

1.3 *Continued* Methods Relevant for RDTs

Country	Definition	Clinical Effectiveness	Economic Modelling
USA (2019)	High impact Single and Short-term Therapies (SSTs) Single intervention or short-term treatment (<1 year) that offers a significant potential for substantial and sustained health benefits extending throughout patient lifetime - Potential cures that can eradicate a condition - High impact Txs that can produce sustained major health gains or halt progression of significant illnesses subject to public consultation and discussion with stakeholders		***Assessing and describing uncertainty*** ICER will make cure proportion modelling its standard reference case for high-impact SSTs whenever relevant, but to address uncertainty we will also provide survival analysis based on other modelling approaches when feasible. In addition to the base case and associated sensitivity analyses, ICER will develop two specific scenario analyses to reflect an optimistic and a conservative assumption regarding the benefit of SSTs under review. Input for best approaches to modelling the optimistic and conservative scenarios will be sought beginning with the scoping phase and will be included as part of the model analysis plan. These scenario analyses will be presented in conjunction with the base case for consideration by the independent appraisal committees. When the SST price is known or can be estimated, assessments of SSTs will also include a scenario with a threshold analysis determining the duration of beneficial effect (e.g., cure) for those patients receiving short-term benefit that would be needed to achieve standard cost-effectiveness thresholds (e.g., $150,000/QALY). ICER will add a new section in the "Long-Term Cost-Effectiveness" section of ICER reports which will discuss "Uncertainty and Controversies" related to the economic evaluation. This new section will be added to all ICER reports, not just those for high-impact SSTs. ***Time divergence between costs and benefits*** ICER will make no change to its reference case 3% discounting to be applied to both health outcomes and costs. ***Sharing of health systems savings*** To stimulate further consideration of how the cost offsets generated by new treatments should be incorporated in calculations of the value and value-based price for a new treatment, ICER will develop two new hypothetical economic analysis scenarios that evaluate cost-effectiveness outcomes with a different approach to the cost offsets from a new treatment. Threshold analyses for treatment price will be presented but will not be suggested as normative guides to pricing. These two hypothetical scenarios will be generated for all high-impact SSTs under review, as well as other (non-SST) treatments with relevant and substantial potential cost-offsets. In most cases this will be situations in which potential cost offsets are greater than $1 million over a lifetime: 1. A 50/50 shared savings model in which 50% of the lifetime health system cost offsets from a new treatment are "assigned" to the health system instead of being assigned entirely to the new treatment; and 2. A cost-offset cap model in which the health system cost offsets generated by a new treatment are capped at $150,000 per year but are otherwise assigned entirely to the new treatment.

Appendix 2 Country Guidance on Deliberation and Data Collection Specific to Rare Diseases

Web search in 2020 based on ISPOR website of pharmacoeconomic guidelines, review of HTA body websites to find up-to-date material, links provided in WS1 country vignettes, excluding England and Scotland that are reported in MS41.

2.1 Deliberation Guidance for RDTs

Country	Definition	Deliberation
Australia (2016)	"rare" <1/50,000	The four factors described below apply in exceptional circumstances and are particularly influential in favour of listing. When all four factors apply concurrently, this is called the **'rule of rescue':** • **No alternative exists** in Australia to treat patients with the specific circumstances of the medical condition meeting the criteria of the restriction. This means that there are **no** • **nonpharmacological or pharmacological** interventions for these patients. • The medical condition defined by the requested restriction is **severe, progressive and expected to lead to premature death. The more severe the condition, or the younger the age at which a person with the condition might die, or the closer a person with the condition is to death, the more influential the rule of rescue might be in the PBAC's consideration.** • The medical condition defined by the requested restriction **applies to only a very small number of patients. Again, the fewer the patients, the more influential the rule of rescue might be in the PBAC's consideration.** However, the PBAC is also mindful that the PBS is a community-based scheme and cannot cater for individual circumstances. • The proposed medicine provides a **worthwhile clinical improvement sufficient to qualify as a rescue from the medical condition**. The greater the rescue, the more influential the rule of rescue might be in the PBAC's consideration. As with other relevant factors, the rule of rescue supplements, rather than substitutes for, the evidence-based consideration of comparative cost-effectiveness. A decision on whether the rule of rescue is relevant is only necessary if the PBAC would be inclined to reject a submission because of its consideration of comparative cost-effectiveness (and any other relevant factors). In such a circumstance, if the PBAC concludes that the rule of rescue is relevant, it would then consider whether this is sufficiently influential in favour of a recommendation to list that the PBAC would reverse a decision not to recommend listing if the rule of rescue were not relevant. This guidance on the rule of rescue is deliberately kept narrow. Although there are relevant arguments for broadening the guidance, the PBAC is concerned that doing so would reduce the relative influence of the rule of rescue if it is applied to a broader set of eligible submissions. In other words, the greater the proportion of submissions that the rule of rescue is applied to, the smaller its average impact in favour of listing across the identified submissions. One issue that has arisen concerning the rule of rescue is that a second medicine to treat the medical condition that is considered to meet the requirements of the rule is not suitable for this consideration. This is because, by definition, the second medicine does not meet the essential first factor (i.e., that there is currently no alternative intervention). This causes a difficulty if listing of the second medicine is sought on a cost-minimisation basis.

2.1 *Continued* Deliberation Guidance for RDTs

Country	Definition	Deliberation
New Zealand	"rare" 1<1/50,000 (~90 people) in NZ For this and all authorised indications and those in phase III trials	New framework for all deliberations following RDT pilot based on need, health benefits, costs/savings, suitability at levels of person, family/extended family, health system https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/factors-for-consideration/

2.2 Deliberation Guidance for Single and Short-Term Therapies (e.g., cell and gene therapies) that may be used in RDs

Country	Definition	Deliberation
USA (2019)	High impact Single and Short-term Therapies (SSTs) Single intervention or short-term treatment (<1 year) that offers a significant potential for substantial and sustained health benefits extending throughout patient lifetime - Potential cures that can eradicate a condition - High impact Txs that can produce sustained major health gains or halt progression of significant illnesses subject to public consultation and discussion with stakeholders	***Additional elements of value*** For all ICER reviews (not only those for high-impact SSTs), we will add three additional domains of "potential other benefits or disadvantages" for voting by independent appraisal committees: (1) A potential advantage for therapies that offer a new treatment choice with a different balance or timing of risks and benefits that may be valued by patients with different risk preferences; (2) a potential advantage for therapies that, if successful, offer the potential to increase access to future treatment that may be approved over patients' lifetime; and (3) a potential disadvantage for therapies that, if not successful, could reduce or even preclude the potential effectiveness of future treatments.

2.3 Deliberation and Further Data Collection for some RDTs

Country	Definition	Data Collection
Norway (2018)	Very small patient group with extremely severe condition and considerable expected benefit *[ultra-rare and severe and high expected benefit]* It is not useful to set absolute conditions for evaluating whether the requirements for "very small patient groups ", "extremely severe conditions " or "considerable expected benefit ", is fulfilled. There should, however, be indicative criteria for decision-making. • Very small <1/100,000 on a global basis per pharmaceutical *(all indications)* <50 pats in Norway per pharmaceutical (steady state prevalence[9]) AND • Extremely severe Level of severity measured using absolute shortfall corresponding to at least 30 good life years (QALYs)[10] AND • Considerable expected benefit Expected benefit is considerable and a minimum of 2 gained good life years compared to standard (QALYs) All 3 indicative criteria should be fulfilled. The criteria are indicative in nature and must be assessed in accordance with an overall assessment in every specific case. In some cases, it may, at a later date, be relevant to re-evaluate how far the indicative criteria have been fulfilled.	Monitoring is required of individual and aggregated data with consideration of start and stop criteria.

[9] Diseases can become less rare when an effective treatment emerges and diagnosis improves

[10] Very stringent restriction, meant to cover children with congenital conditions, where higher use of resources could be justified by argument of greater shortfall, but review of 19 cases in Norway from 2014-2017 shows only 2 cases that reached 20 years shortfall, none reached 30 years shortfall.

References

Annemans, Lieven, and Amr Makady. 2020. “TRUST4RD: Tool for Reducing Uncertainties in the Evidence Generation for Specialised Treatments for Rare Diseases.” *Orphanet Journal of Rare Diseases*.

Berdud, Mikel, Michael Drummond, and Adrian Towse. 2020. “Establishing a Reasonable Price for an Orphan Drug.” *Cost Effectiveness and Resource Allocation* 18(1):31.

Daniels, Norman, and James E. Sabin. 2008. “Accountability for Reasonableness: An Update.” *BMJ* 337:a1850.

EUnetHTA Joint Action 2. 2016. “HTA Core Model Handbook.” *The EUnetHTA JA2* 2.

Facey, K., A. Granados, G. Guyatt, A. Kent, N. Shah, G. J. Van Der Wilt, and D. Wong-Rieger. 2014. “Generating Health Technology Assessment Evidence for Rare Diseases.” *International Journal of Technology Assessment in Health Care* 30(4).

Facey, K., and H. P. Hansen. 2015. “The Imperative for Patient-Centred Research to Develop Better Quality Services in Rare Diseases.” *Patient* 8(1).

Facey, K. M., N. Bedlington, S. Berglas, N. Bertelsen, A. N. V. Single, and V. Thomas. 2018. “Putting Patients at the Centre of Healthcare: Progress and Challenges for Health Technology Assessments.” *Patient* 11(6).

Facey, Karen M., Piia Rannanheimo, Laura Batchelor, Marine Borchardt, and Jo De Cock. 2020. “Real-World Evidence to Support Payer/HTA Decisions about Highly Innovative Technologies in the EU - Actions for Stakeholders.” *International Journal of Technology Assessment in Health Care* Sep 3:1–10.

Janoudi, Ghayath, William Amegatse, Brendan McIntosh, Chander Sehgal, and Trevor Richter. 2016. “Health Technology Assessment of Drugs for Rare Diseases: Insights, Trends, and Reasons for Negative Recommendations from the CADTH Common Drug Review.” *Orphanet Journal of Rare Diseases* 11(1).

Latimer, NR. 2013. *Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data*. London.

Nguengang Wakap, Stéphanie, Deborah M. Lambert, Annie Olry, Charlotte Rodwell, Charlotte Gueydan, Valérie Lanneau, Daniel Murphy, Yann Le Cam, and Ana Rath. 2020. “Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database.” *European Journal of Human Genetics*.

Nicod, E., L. Annemans, A. Bucsics, A. Lee, S. Upadhyaya, and K. Facey. 2019. “HTA Programme Response to the Challenges of Dealing with Orphan Medicinal Products: Process Evaluation in Selected European Countries.” *Health Policy* 123(2).

Nicod, Elena, Karen Berg Brigham, Isabelle Durand-Zaleski, and Panos Kanavos. 2017. “Dealing with Uncertainty and Accounting for Social Value Judgments in Assessments of Orphan Drugs: Evidence from Four European Countries.” *Value in Health* 20(7):919–26.

Nicod, Elena, Amanda Whittal, Michael Drummond, and Karen Facey. 2020. “Are Supplemental Appraisal/Reimbursement Processes Needed for Rare Disease Treatments? An International Comparison of Country Approaches.” *Orphanet Journal of Rare Diseases* 15(1):1–14.

Pennington, B., and R. Wong. 2019. *MODELLING CARER HEALTH-RELATED QUALITY OF LIFE IN NICE TECHNOLOGY APPRAISALS AND HIGHLY SPECIALISED TECHNOLOGIES*.

Powell, Philip A., Jill Carlton, Donna Rowen, and John E. Brazier. 2019. “Producing a Preference-Based Quality of Life Measure for People with Duchenne Muscular Dystrophy: A Mixed-Methods Study Protocol.” *BMJ Open* 9(3):e023685.

Prentice RL. 1989. “Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria.” *Stat Med* 8:431–40.

Rand, Leah, Michael Dunn, Ingrid Slade, Sheela Upadhyaya, and Mark Sheehan. 2019. "Understanding and Using Patient Experiences as Evidence in Healthcare Priority Setting." *Cost Effectiveness and Resource Allocation* 17(1):1–13.

Staniszewska, Sophie, and Sophie Werkö. 2017. "Patient Involvement in Health Technology Assessment." *Patient Involvement in Health Technology Assessment* 43–50.

Watt, A., A. Cameron, L. Sturm, T. Lathlean, W. Babidge, S. Blamey, K. Facey, D. Hailey, I. Norderhaug, and G. Maddern. 2008. "Rapid Reviews versus Full Systematic Reviews: An Inventory of Current Methods and Practice in Health Technology Assessment (International Journal of Technology Assessment in Health Care (2008) 24, 2)." *International Journal of Technology Assessment in Health Care* 24(3).

Document History

March 2021	Submitted to EC
12 May 2021	Slight alteration of title and addition of EC logo and funding statement on front page Addition of citation and cc by Correction made to numbering in Appraisal Framework Table on page 16 and redactions added