# EUnetHTA Technical Review of methodological standards in comparative effectiveness & clinical HTA

## Final Report of EUnetHTA Guideline reviews and Survey

Prepared for the EFPIA HTA WG

v 3.0


**Prepared For:**

Edith Frénoy
Director Market Access, HTA Policy lead
T: +32 2 62 62 548
M: Edith.Frenoy@efpia.eu

**Questions To:**

Manpreet Sidhu
ICON PLC
Divisional Principal GHTA-HERO
M: +1 347 241 9094
E: Manpreet.Sidhu@iconplc.com



**Date:** 15 August 2019
**Report:** Updated final version
**Version:** 3.0

# Version History

| Version Number | Summary/Reason for Changes | Date Issued |
|---|---|---|
| 1 | First draft | 5 June 2019 |
| 2 | Final version | 5 July 2019 |
| 3 | Updated final version | 15 August 2019 |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# Table of Contents





© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# List of Tables





© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# List of Figures





© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# Abbreviations

| Abbreviation | Description |
|---|---|
| AE | Adverse event |
| AMSTAR | A MeaSurement Tool to Assess systematic Reviews |
| CAT | Computerised adaptive test |
| ClinRO | Clinician-reported outcome |
| CNS | Central nervous system |
| COA | Clinical outcome assessment |
| EBM | Evidence-based medicine |
| EC | European Commission |
| ED | Early dialogue |
| EFPIA | European Federation of Pharmaceutical Industries and Associations |
| EMA | European Medicines Agency |
| ENT | Ear, nose and throat |
| EQ-5D | EuroQoL five dimensions |
| EU | European Union |
| EUnetHTA | European Network for Health Technology Assessment |
| FDA | Food & Drug Administration |
| G-BA | Gemeinsamer Bundesausschuss (Federal Joint Committee) |
| HAS | Haute Autorité de Santé (French National Authority for Health) |
| HRQoL | Health-related quality of life |
| HTA | Health technology assessment |
| ICH | International Council for Harmonisation |
| IQR | Interquartile range |
| IRT | Item Response Theory |
| ISPOR | International Society for Pharmacoeconomics and Outcomes Research |
| ITC | Indirect treatment comparison |
| JA | Joint action |
| JCA | Joint clinical assessment |
| MAIC | Matching adjusted indirect comparison |
| MCID | Minimal clinically important difference |
| MS | Member state |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| Abbreviation | Description |
|---|---|
| NICE | National Institute for Health and Care Excellence |
| NMA | Network meta-analysis |
| NRS | Non-randomised study |
| OS | Overall survival |
| PBM | Preference-based measure |
| PFS | Progression-free survival |
| PRO | Patient-reported outcome |
| RCT | Randomised clinical trial |
| RD | Responder definition |
| REA | Relative effectiveness assessment |
| REQueST | Registry Evaluation and Quality Standards Tool |
| RoB | Risk of bias |
| RREA | Rapid relative effectiveness assessment |
| RWD | Real-world data |
| RWE | Real-world evidence |
| SoC | Standard of care |
| STC | Simulated treatment comparison |
| TA | Therapy area |
| TLV | Dental and Pharmaceutical Benefits Agency |
| WG | Working group |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# 1. Executive summary

## 1.1. Introduction

In January 2018, the European Commission (EC) proposed a regulation on Health Technology Assessment (HTA) with the aim of improving the availability of innovative health technologies for European Union (EU) patients, ensuring efficient use of resources, strengthening the quality of HTAs across the EU, and improving business predictability. A key pillar of the proposed EC regulation on HTA regulation is focused on conducting Joint Clinical Assessments (JCAs), which aims to standardise and centralise HTA procedures for clinical evidence beyond 2020, provided the EU regulation comes into effect (i.e., the legislative process is still ongoing). The JCA methodology approach is supported and underpinned by a series of guidelines published by the European Network for Health Technology Assessment (EUnetHTA).

The objective of this work was to develop materials that will enable the European Federation of Pharmaceutical Industries and Associations (EFPIA) HTA Working Group (WG) to engage in the debate and process related to HTA methodology within the context of the proposed EC regulation on HTAs. This report outlines the key evidence generation challenges facing pharmaceutical manufacturers in Europe that are prompting manufacturers to seek early scientific advice from HTA bodies. The report also provides a critical review of the current EUnetHTA methods standards and recommendations on future methodological requirements for JCAs within the EU.

## 1.2. Methods

The methods outlined in this report are two-fold. Firstly, an online survey related to prior experience with Early Dialogues (EDs) was conducted among pharmaceutical companies. Secondly, a critical review of selected EUnetHTA methodology guidelines was undertaken.

Pharmaceutical companies were surveyed to identify key evidence requirements for which they may seek ED/ HTA scientific advice. The aim here was to identify evidence generation challenges that require state-of-the-art HTA methods.

In parallel to the online survey, a de-novo Critical Review Framework was developed based on an iterative process. The Critical Review Framework was used to critically review the following selected EUnetHTA methodological guidelines: endpoints, comparator selection, direct and indirect treatment comparison (ITC), health-related quality of life (HRQoL) and utility measures, internal validity of non-randomised and randomised studies, and safety. Subject matter experts with expertise in HTA methodology across France, Germany, the UK, ICON plc, and the EFPIA HTA WG) reviewed the guidelines to provide in-depth insights on key gaps and potential future improvements, and to inform their supportive use in JCAs.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

## 1.3. Key findings

EFPIA has published its position on the proposal for a Regulation on EU HTA including the process of JCA (EFPIA 2018). EFPIA's position is that JCAs are limited to medicinal products with a centralised marketing authorisation and mandatory uptake in the national assessments is implemented. This process could allow for predictable evidence generation requirements at the development stage, and thus transparent, timely, consistent and high-quality JCAs are key to expediting patient access to new medicines. Furthermore, the process would need to allow for key stakeholderinput into the scope of the JCA, provide HTA agencies with further clarification to resolve uncertainties, highlight discrepancies/inaccuracies in the interpretation of the data submitted, and appeal if the agreed process has not been fully adhered to.

On the joint scientific consultations pillar of the EU joint work, the EFPIA welcomes the opportunity for early discussion of the clinical data requirements for regulatory approval and JCA, while also facilitating the selection of the relevant comparator(s) for both regulatory approval and HTA purposes.

The ED online survey results indicated areas in which European alignment on clinical evidence generation requirements is supported by the EFPIA member companies, with the three most common reasons for seeking early scientific advice from HTA bodies related to comparator selection, study endpoints and patient population.

A critique of selected methodological guidelines showed that the majority of these documents meet their stated objectives and clearly present the problem statement and guideline scope. Nevertheless, a critical review of the guidelines further found key gaps relative to the current internationally recognised best practice (state-of-the-art) HTA methodologies. Areas for improvement in the nine EUnetHTA guidelines are presented below ordered by the frequency of reasons to seek ED/scientific advice:

***Comparators and Comparisons guideline – identified methodological gaps:***

Methodological gaps identified in this guideline pertain to the establishment of a consensus on defining and selecting the standard of care (SoC) to be used as a comparator amongst EU Member States (MSs). Further guidance is needed on the level of evidence of efficacy and effectiveness required by the assessors, and how different sources of SoC are assessed for quality and appropriateness (i.e., comparator(s) choice should follow the principles and hierarchy of evidence-based medicine (EBM) with additional guidance from the timely use of European clinical guidelines). It is acknowledged that different comparators are warranted due to varying treatment landscapes, lines of treatment, clinical guidelines, and varying subpopulations. A pragmatic approach should be endorsed for the final basket of relevant comparators to ensure that manufacturers can undertake the necessary analyses for the JCA (e.g., the scope should include one comparator that is considered relevant by all MSs; this should be used as the base-case for JCA).

Lastly, the guideline does not take into consideration the early alignment of comparator selection in terms of early scientific advice from regulatory and HTA agencies.

***Direct and Indirect Comparisons guideline – identified methodological gaps:***



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The main methodological gaps pertain to the following areas: inclusion of state-of-the-art indirect comparison methods such as population-adjusted indirect comparisons (Matching Adjusted Indirect Comparison (MAICs) and Simulated Treatment Comparison (STCs)), provision of detailed guidance on the use of observational or non-randomised studies (NRSs), on when to employ qualitative versus quantitative summaries of data, and on how to account for heterogeneity. With regard to the heterogeneity of efficacy (e.g., measuring survival benefit in the presence of non-proportional hazards) especially, further guidance is needed on the most appropriate methods to account for time-varying treatment effect in clinical HTAs. Overall, the limited consideration and discussion of methods in this guideline restricts its applicability across assessments and HTA bodies. It should be noted that in some situations, the evidence found to inform a network meta-analysis (NMA) may be so disparate (i.e., sparse and/or heterogeneous) that the output would be uninformative and/or result in erroneous conclusions.

This guideline is currently under revision by the EUnetHTA, and the key issues noted in this critique of the guideline should be addressed in the next version.

***HRQoL and utility measures – identified methodological gaps:***

The guideline appears to lack structure in the presentation of methods for comparative clinical effectiveness assessments and cost-effectiveness analysis. Guidance on the importance and relevance of qualitative data to underpin the validity of HRQoL instruments is insufficient. Further elaboration would be required to account for the insensitivity of some generic HRQoL instruments. The guideline lacks recommendations on which instrument would be preferred for JCAs. Additional guidance on the available types of HRQoL analyses (e.g., time to deterioration) and how these should be reported is also lacking. Furthermore, details of country differences are missing, despite the importance of taking these into consideration in the course of evidence collection and presentation; there isl likely to be variation in the extent of adherence across the EU HTA bodies due to the variability in country requirements and preferences.

***Clinical endpoints guideline – identified methodological gaps:***

Key identified methodological gaps pertain to provision of further guidance on analysis of time-to-event data, which would promote consistency and predictability of HTAs. Moreover, detailed guidance is missing in the evidence requirements for therapies that provide long-term clinical benefit and/or the potential for cure, as well as endpoint selection for pragmatic trials. In addition, further discussion on estimands, sensitivity analysis in randomised clinical trials (RCTs), and determination of minimally clinically significant difference for endpoints were considered warranted.

***Composite endpoints guideline – identified methodological gaps:***

Critique of the document noted gaps in details regarding the weighting of composite endpoint components and the statistical methods for recurrent events of composite endpoints. Thediscussion of competing risks was also inadequate.

***Surrogate endpoints guideline – identified methodological gaps:***



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

In this guideline, the key gaps related to the provision of recommendations for the reporting of surrogate endpoints, clarity on what circumstances the use of a surrogate endpoint would be acceptable for JCAs, and the relevant limits for adequate surrogacy. Furthermore, guidance on the requirements for therapies that provide long-term clinical benefit and/or the potential for cure was missing. Otheridentified gaps were the lack of advice on validation and acceptable analyses (e.g., correlation with a prespecified correlation coefficient or surrogate threshold)to evaluate adequate surrogacy, and guidance on the use of validated surrogates by drug class, disease type and disease stage.

***Internal validity of iseNRS guideline – identified methodological gaps:***

Most reviewers of the guideline noted that guidance is needed on the assessment of internal validity for innovative study designs in real-world evidence (RWE)-based evidence generation research. Furthermore, what would be beneficial is clarification on the circumstances under which NRSs are acceptablethe identification of study subjects, and acceptable outcomes and analytical methods.

***Internal validity of RCTs guideline – identified methodological gaps:***

The guideline does not provide sufficient evidence relating to the reasons why the Cochrane risk of bias (RoB) tool was selected over other existing tools. Furthermore, there is lack of guidance on how the Oxman and Guyatt index and the A MeaSurement Tool to Assess systematic Reviews (AMSTAR) instrument tools should be used, as well as when bias should be judged as “low”, “high” or “unclear” to avoid subjectivity. Several reviewers also noted that further guidance is needed on the assessment of the internal validity of RCTs with treatment switching at progression.

***Safety guideline – identified methodological gaps:***

Lack of guidance on the synthesis of summarised safety information (quantitatively or qualitatively), how to determine “clinical relevance” of adverse reactions, and how to weight relative safety against relative benefit were noted by most reviewers. Furthermore, the guideline needs to provide clarity on how to account for different follow-up periods and what method to use to quantify relative safety.

In comparative safety assessments, all interventions (pharmaceutical, medical devices, and non-drug therapies) should be considered, especially given the increasing development of e-Health applications. Gguidance should be provided on how to assess safety with connected devices alongside medications, and how to use data from tools that allow the severity of adverse events (AEs) to be reported directly by patients.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

## 1.4. Recommendations to inform the EFPIA HTA WG position

The key recommendations discussed in this report pertain to guidelines which were identified as having critical gaps or areas for improvement, and should support the formulation of the EFPIA HTA WG position on the development of methodological guidelines for JCA (Table 1).

A traffic light/RAG indicator is shows identified methodological gaps

- **Red:** **Major gaps**
- **Amber:** **Some gaps**
- **Green:** **No/minor gaps**



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

**Table 1 Key gaps identified in the EUnetHTA guidelines that require action from the EFPIA**

| | Gaps | Recommendations |
|---|---|---|
| Comparators & comparisons |  <br>• Process to establish consensus on SoC selection amongst EU MS<br>• Further guidance on the SoC selection is needed in terms of level of evidence of efficacy and effectiveness required, and how different SoC sources (e.g., registries) are assessed for quality and appropriateness<br>• Guidance on the timely use of European clinical practice guidelines appropriate for future JCAs<br>• Alignment of early scientific advice from regulatory and HTA bodies is not mentioned in the guideline | • A transparent, timely and consistent ED and scoping process, by which EU MS agree on (a basket of) comparators, is considered essential. Also, an early definition of what constitutes SoC treatment(s) for future JCAs is needed. The comparator selection should be in consultation with industry<br>• Comparator selection should be driven by EBM principles, clinical practice (e.g., physician choice/selection/guidance, patient engagement) and unmet need when scoping to identify a basket of relevant SoCs<br>• Provide guidance when there is no SoC (e.g., multiple comparators available) to define a basket of treatments, considering local restrictions in the selection of comparators<br>• Upon agreement on a manageable set of comparators (e.g., treatment most likely to be replaced), a hierarchy of considerations should influence the assessment, in accordance with EBM standards (e.g.,approvals with RCT data on top and registry/real-world data (RWD) to be lower in ranking)<br>• Provide guidance on the level of evidence of efficacy and effectiveness required for SoC selection and how different sources (e.g., registries) are assessed for quality and appropriateness<br>• Learnings from regulatory guidance (European Medicines Agency (EMA) and Food & Drug Administration (FDA)) in comparator selection should be integrated in the JCA and similar criteria to reach consensus should apply in HTAs |
| |  **STRATEGIC MESSAGE**: Advocate comparator selection according to the principles of EBM to establish a pragmatic basket of SoC with at least of one the comparators agreed by all EU MSs and following consultation with industry. | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Direct and Indirect Comparisons** |  <br>• Newer population-adjusted indirect comparison methods (e.g., MAIC or STC) are not included<br>• Detailed guidance on selecting the most appropriate type of evidence for use in observational studies or NRS, and when to employ qualitative vs. quantitative summaries of data (i.e., in cases where analysis is not feasible)<br>• Discussion on different variables such as ordinal and time-to-event (including those beyond binary and continuous outcomes)<br>• No clear recommendations on how to assess heterogeneity are provided | • Agreement is needed across the EU jurisdictions that indirect comparison is an acceptable method for joint relative effectiveness assessment (REA) as well as guidance on where an indirect comparison may not be appropriate (e.g., when there is significant heterogeneity then the ITC results would be of low value); the EFPIA position is that ITC is a necessary method for JCA and could addressed in the scoping process when selecting an appropriate comparator.<br>• Guidance is needed on the most appropriate methods to use when non-proportional hazards are observed in time-to-event RCT data and sub-groups considering the role of biomarkers (e.g., if trial not stratified by subgroups then the randomisation may be compromised in an ITC)<br>• Overall, the guideline should be revised in order to accommodate the following:<br>○ Comment on the need for scientific validity and predictability of ITC methods (e.g., stimulate research and report empirical verifications supporting the reliability of ITC to predict direct comparison results)<br>○ Provide a framework which researchers can use as guidance for the most appropriate analysis (e.g., a decision tree approach outlining required methods and limitations, describing the type of ITC by case)<br>○ Summary of the newer approaches including MAIC, STC<br>○ The use of observational studies (iseNRS) as a key evidence base and how to use RWE (e.g. hierarchical Bayesian framework)<br>○ Account for additional variables like time-to-event survival (outcomes) beyond binary and continuous outcomes<br>○ Advise on when to employ qualitative vs. quantitative summaries of data (i.e. in cases where an analysis is not feasible)<br>○ Additional statistical modeling approach to account for heterogeneity |
| |  **STRATEGIC MESSAGE**: EFPIA will provide technical recommendations to the EUnetHTA guideline development teams with the expectation that the proposed recommendations will be considered for the next version; EFPIA affirms that while indirect comparisons are essential for JCAs, it recognises that these analyses will be inappropriate in certain cases. | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **HRQoL and utilities** |  <br>• Importance and relevance of qualitative data to underpin the validity of the HRQoL instrument<br>• Lack of sensitivity of some generic HRQoL instruments<br>• Advice on statistical approaches to handle missing data should be included (e.g. mixed effects models, imputation and sensitivity analyses)<br>• More consideration of meaningful change definitions are needed. Responder definition (RD) and minimal clinically important difference (MCID) should be estimated using anchor-based approaches in the context of use (and supported by distribution-based approaches) or use those in the literature (e.g., EORTC-QLQ-C30)<br>• Detail on country-differences to consider in evidence collection and presentation | • The guideline should be split into two distinct sections with a clear definition of data and methods:<br>○ Part A – HRQoL, for comparative clinical effectiveness assessments<br>○ Part B – Utilities, geared toward the cost-effectiveness analysis<br>• Available types of analyses relevant for HRQoL should be included in the guidelines (e.g., time to deterioration) and to how these should be reported<br>• Guidance required for HRQoL assessment in acute disease, rare disease, and novel therapies<br>• Advise on need for evidence of clinically meaningful change to aid interpretation of scores and other considerations in the analysis/interpretation of data (e.g., norms for HRQoL and adjusting for sociodemographic factors)<br>• Guidance needed on time points of administration of HRQoL instruments where considered feasible, participant burden, and handling of missing data; consider inclusion of long-term data capture (e.g., post-progression) and exit interviews to establish treatment impact and to interpret meaningful clinical outcome assessment (COA; including PROs) scores<br>• Provide clear advice on the value of data from mixed-methods approaches in the context of clinical trials; qualitative evidence is imperative to ensure selected instruments are appropriate and relevant for the patient population<br>• Include discussion and recommendation on the use of computerised adaptive tests (CATs), item banks and meta-analyses<br>• Mapping should remain optional due to practicalities from introduction of new COAs, but inclusion of a generic instrument (e.g., EQ-5D) in clinical trials should be recommended; recommendations for statistical approaches for mapping or mapping to domains / responses should be made |
| |  **STRATEGIC MESSAGE**: Recommend and support revisiting the state-of-the-art by adoption of new methods and guidance in an updated guideline that sets apart HRQoL from utilities; furthermore, position the need for utilities in JCAs. To ensure comparability and consistency across EU MSs, the EuroQoL instrument (EQ-5D) is recommended. Guidance should be provided on the country index values that should be used and what is considered a meaningful change in baseline score. | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Clinical Endpoints** |  <br>• Evidence requirements for therapies that provide long-term clinical benefit and/or potential for cure; guidance on extrapolation beyond the end of a trial and imputation of missing data (e.g., address duration of follow-up period)<br>• Discussion of estimands and sensitivity analysis in RCTs<br>• Determination of specific requirements for health applications (e.g., digital endpoints)<br>• Guidance for selection of endpoints and rationale for deviation from regulatory endpoints<br>• Guidance on determination of thresholds for meaningful change or differences in endpoints<br>• When discussing the three broad groups of endpoints, it is important to reference COA endpoints in general, rather than just HRQoL, as this is just a subset of COAs | • Provide detailed guidance on the level of evidence, acceptability of analytical methods and endpoint validation, beyond OS<br>• Further clear guidance is needed on methods to adjust overall survival (OS) if cross-over from the control to the experimental arm occurs. Moreover, guidance is needed on how to estimate OS in trials where non-proportional hazards are observed<br>• Further advice – for example, on the definition of long-term assessment in terms of observation time – would be valuable<br>• Include guidance on how endpoint selection should be informed by their value to patients; currently, the methods guideline focuses on statistical aspects and less on how to incorporate the patient voice in order to be accepted by HTA bodies<br>• Holistically cover all aspects of endpoint definition and measurement, including concepts of estimands from the updated International Council for Harmonisation (ICH) guideline<br>• Guidance on the determination of thresholds for meaningful change or differences in endpoints and how endpoints should be weighted in JCAs |
| |  **STRATEGIC M ESSAGE**: Advocate for clear guidance on validation of endpoints, relevance to patients, and accepted analytical methods, including methods to estimate treatment effect when cross-over/treatment-switching occurs | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Composite Endpoints** |  <br>• Discussion of weighting of components of the composite<br>• Statistical methods for recurrent events of composite endpoints<br>• Further discussion of competing risks | • Focus more on interpretation and presentation of composite endpoints to provide practical support for HTA assessors, as the current perspective is the regulatory standpoint<br>• Provide clarification on the methods that should be used to validate a new composite endpoint for new technologies (e.g., CAR-Ts)<br>• Recommend using criteria from research to weight components of a composite in clinical endpoints, and improved understanding of competing risks and recurrent events to improve the acceptance of composite endpoints for clinical HTA and allow for more streamlined clinical trials<br>• Composite endpoint selection should be informed by their value to patients; currently the guidance is too statistical and the patient voice is not sufficiently heard; further focus should be on how to incorporate the patient voice in order to be accepted by HTA bodies<br>• Composite endpoints accepted by EMA should be equally accepted by HTA, without the need to present results for the different components separately; early Parallel Consultation with Health Authorities and HTA bodies is key to avoid discrepancy in constructions of composite endpoints for regulatory and HTA purposes |
| |  **STRATEGIC MESSAGE**: Recommend revisiting the guideline from the HTA and not the regulatory perspective, and provide clear guidance on the validation of composite endpoints to ensure consistency with regulatory requirements | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Surrogate Endpoints** |  <br>• Recommendations on the reporting of surrogate endpoints<br>• Clarification needed under what circumstances the use of a surrogate endpoint would be acceptable for the JCA and accepted threshold for surrogacy<br>• Requirements for surrogate endpoints for therapies that provide long-term clinical benefit and/or the potential for cure<br>• Advice on validation and thresholds to assume adequate surrogacy<br>• Guidance on the use of validated surrogates by: drug class, disease type and disease stage | • Provide detailed advice on the adequacy of surrogate markers, the validation process, and statistical methods<br>○ Handle the current scientific perspective and criteria between options:<br>1. First in class<br>2. Curative therapies<br>3. Therapies in an established disease area<br>○ Upon the introduction of a new technology (e.g., CAR-T), guidance should be provided on how to re-validate the surrogate endpoint<br>• A pragmatic approach is required for acceptance and validation of surrogate endpoints; when surrogate are considered established, crucial or important endpoints they should be adopted, unless there is strong evidence suggesting they are not appropriate<br>• Refrain from setting thresholds without a commonly agreed standardised methodology to validate surrogacy<br>• Recommendations on standardisation of reporting requirements for surrogate markers data from clinical trials (e.g., references to biological plausibility and evidence of validation)<br>• Discuss adequacy of surrogate markers for therapies that provide long-term clinical benefit, including early disease markers and/or the potential for cure<br>• Endpoints should be informed by their value to patients (e.g., how progression-free survival (PFS) is valued in terms of no tumor growth or the absence of metastasis)<br>• Underline the need for early Parallel Consultation with Health Authorities and HTA bodies to address the variability in requirements (e.g., on the levels of evidence needed for surrogate markers in an HTA context) |
| |  **STRATEGIC MESSAGE**: EFPIA's position is that surrogate endpoints, when considered established, crucial or important endpoints, should be adopted unless there is strong evidence suggesting they are not appropriate; clearer guidance and additional research on methodology to validate surrogacy is advocated | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Internal validity of NRS** |  <br>• In RWE-based evidence generation, guidance is needed on assessing the internal validity of innovative study designs<br>• The EUnetHTA guideline does not deep dive into the details of how currently available RoB tools apply to different study designs and which are the most commonly used by HTA agencies<br>• Regulators are developing a framework for RWE programs (e.g., FDA)<br>• No consensus between HTA bodies on the preferred NRS appraisal tool<br>• Many of the existing tools for RoB assessment fail to address all relevant domains or confuse the assessment of internal validity with that of the quality of reporting<br>• Clarify under what circumstances NRSs are acceptable, and provide clarity on how to identify study subjects and what the acceptable outcomes and analytical methods are | • Adapt the guideline scope not to include NRS as the sole but an additional source of information on effectiveness and safety in cases where evidence of higher level is not available<br>• With a pan-European template for the submission dossier/package, provide examples of bias assessment<br>• Provide guidance on understanding circumstances where NRS evidence is acceptable, the most appropriate methods to identify study subjects, acceptable outcome measures and analytical methods<br>• Recommend acquiring a consensus on the definitions of NRS designs to enable the use of the appropriate RoB tool<br>• Include guidance on the assessment of internal validity for single-arm trials that use a RWD control arm, cross-sectional studies, historical comparisons, and innovative study designs (e.g., Simon two-stage optimal design)<br>• Provide examples and present case studies of widely accepted RoB Tools used by the majority of HTA agencies, especially in RWE studies and study designs that may be problematic to assess with the recommended ACROBAT / ROBINS-I tool (e.g., explore tools and techniques to estimate HR from historical comparisons/NRS)<br>• Request further evidence relating to the validity and feasibility of the ROBINS-I tool<br>• Advise that pragmatic trials should be considered as RCTs and are, therefore, not included in the scope of NRSs |
| |  **STRATEGIC MESSAGE**: Advocate for a consensus among HTA agencies to utilise standard, state-of-the-art methodology in assessing the internal validity of NRSs to facilitate manufacturers on the selection of appropriate NRS designs and corresponding analytical methods for addressing bias, and to inform future JCAs. | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Internal validity of RCTs** |  <br>• Evidence on the appropriateness of the Cochrane RoB tool over other existing tools is missing<br>• Guidance lacking on how the Oxman and Guyatt index and the AMSTAR instrument tools should be used<br>• Guidance lacking on when a bias is seen as “low”, “high” or “unclear” to avoid subjectivity<br>• Guidance on the internal validity of RCTs with treatment switching at progression is missing<br>• Post-randomisation confounding and selection bias are not explicitly mentioned | • Provide guidance on how to deal with studies which have a high or unclear RoB to allow uniform assessments, in line with the warning against multiple rating scales (e.g., Jadad).<br>• Provide a clear description of the different approaches that EU MSs currently have in dealing with bias<br>• Use of a standardised RoB assessment should become mandatory for all HTA bodies to ensure predictability and consistency. It would be beneficial for HTA bodies to agree on a standardised way to use and apply RoB tools in order to align outcomes in terms of the rating of bias in RCTs (high, low, or unclear bias)<br>  ○ EU HTA bodies would be expected to use and apply the same tool and have the same rating systems for RoB<br>• Need for mechanism to ensure adequate training of HTA agency staff to avoid inconsistencies when assessing RoB in RCTs<br>• Advise the EUnetHTA to update the guideline upon publication of the new version of the Cochrane Handbook (v.6) later in 2019<br>• Treatment switching, as well as statistical methods used to adjust for treatment switching, and pragmatic trials, need to be addressed in a revised EUnetHTA guideline |
| |  **STRATEGIC MESSAGE**: The current guideline is appropriate for use and focus should be on the implementation of commonly accepted standards across all EU MSs, including the use and application of a standardised RoB tool; guidance on how to analyse data from treatment switching is also required | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **Safety** |  <br>• Limited guidance on:<br>○ Synthesis of summarised safety information (quantitatively or qualitatively)<br>○ How to determine 'clinical relevance' of adverse reactions<br>○ How to weight relative safety against relative benefits<br>• Unclear whether only AEs that are at least possibly related to the study drug should be assessed, or whether all treatment-emergent AEs are to be considered<br>• No recommendation on how often the safety assessments should be conducted<br>• Lack of guidance on how to evaluate clinically significant difference in adverse reactions between interventions | • A clear description of the remit and scope of the future JCA relative safety assessment would be beneficial to avoid overlap or repetition with the regulatory safety assessment. Provide guidance on cross-functional collaboration and sharing of methodologies/best practices on safety assessment between the regulatory and HTA bodies<br>• The guideline should consider including recommendations on:<br>• A hierarchy of safety evidence<br>• Systematic process and search strategies to identify all relevant safety data<br>• Process or methodology of the comparison of safety data between technology and comparators, in particular where observational or single arms studies are being utilised<br>• Consistency in safety reporting (i.e., who determines the severity of the AE (patient or physician))<br>• Clearer separation between initial and repeat safety assessment<br>• Recommend considering all interventions (pharmaceutical, medical devices, and non-drug therapies) in comparative safety assessments and how to evaluate the relative safety of interventions with data from different groups<br>• Advocate for guidance on the assessment of safety, with connected e-Health devices alongside medications, and how tools allowing reporting of AE severity by patients should be used<br>• Advise on comparison of number needed to treat (NNT) and number needed to harm (NNH) for both the trial(s) and real-life populations |
| |  **STRATEGIC MESSAGE**: The current version is not a fit-for-purpose guideline for JCAs. It needs to be repositioned for comparative safety of all interventions with the aim of providing clear guidance on repeat relative safety assessments and reporting of SAEs to appropriately inform the future JCAs. | |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | Gaps | Recommendations |
|---|---|---|
| **All Guidelines (overarching themes)** | • General gaps identified across most critiqued guidelines and overarching | • Advocate for and support the early Parallel Consultation with Health Authorities and HTA bodies to build a learning environment and align, and inform clinical development plans in:<br>○ Comparator Selection<br>○ Direct and Indirect Comparisons<br>○ HRQoL and utilities<br>○ Clinical, Composite and Surrogate Endpoints<br>○ Safety<br>• Advocate the adoption of guidelines that reflect the internationally recognised methodological tools and practices to be utilised in the future JCA. Provide examples from different approaches that EU HTA bodies are aligned in using to substantiate the state-of-the-art methods. For example:<br>○ In comparator selection, a summary table of methods applied by MSs such as approved therapies, SoC, market share, most used<br>○ For surrogate endpoints, list examples or case studies of good practice considering apparent differences in acceptance of surrogate outcome measures among HTA bodies<br>○ For ITC methods, indicate that agreement is needed across the MSs on an acceptable method for JCA and guidance where an indirect comparison may not be appropriate<br>• It is notable that there is considerable overlap in the issues that need guidance relating to COA endpoints (Clinical Endpoints guideline) and HRQoL assessment (HRQoL and Utility Measures guideline), as the latter is a type of COA. This overlap could be addressed in one of two ways, either:<br>○ Include full guidance on HRQoL assessment in general in the Clincial Endpoints guideline and altering the HRQoL and utility measures guidelines to focus only on utilities; or<br>○ Expand the HRQoL and Utility Measures guideline to cover COAs and Ultiities in three smaller more focussed parts (as below) and cross-reference Part A in the Clinical Endpoints guideline as appropriate:<br>▪ Part A – COAs<br>▪ Part B – HRQoL, for comparative clinical effectiveness assessments<br>▪ Part C – Utilities, geared toward cost-effectiveness analysis |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

## 1.5. Conclusions

The outcomes of this study centre aim to generate greater understanding of the current consensus in considered methodologies when approaching JCAs in Europe. To ensure that the guidelines include state-of-the-art methods and are suitable for innovative technologies, future updates must address potential issues that manufacturers may face during the centralised clinical HTA process (e.g., SoC selection, analysis of time-to-event data, long-term extrapolation from key pivotal studies, variability on surrogate endpoints assessment, and acceptance of RWE).

Furthermore, it is considered of high importance that EDs are conducted in a transparent and collaborative manner to provide consolidated guidance for manufacturers in regard to clinical trial development, with ideally minimal divergence between the regulatory advice versus HTA advice, especially in the selection of comparator(s) and endpoint(s).

The EFPIA HTA WG is expected to disseminate this deep-dive critique of existing EUnetHTA guidelines along with recommendations for their future content development via different channels (e.g., position paper, publications, issue panels in major conferences) in order to inform the health policy debate across Europe.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# 2. Background

The EC proposal for a Regulation on HTAs to amend Directive 2011/24/EU builds on many years of voluntary cooperation in HTAs between MSs at the EU level, and aims to ensure quality in the JCA process and methodology. The operational objectives of the proposal are to, firstly, promote convergence in HTA tools, procedures and methodologies; secondly, reduce duplication of efforts by HTA bodies and industry; thirdly, ensure the use of joint outputs in MSs; and finally to ensure the long-term sustainability of EU HTA cooperation.

The proposal presents a unique opportunity for greater collaboration and alignment on clinical evidence generation requirements, ensuring consistency, transparency and synergy in clinical assessments by MS and evidence that is relevant for Europe. From the manufacturers' perspective, the proposal would move towards more predictable evidence generation requirements at the development stage.

In February 2019 the European Parliament proposed its vision for JCAs: they constitute scientific analysis of clinical outcomes evaluated in relation to appropriate comparative indicators and are expected to provide predictability on pan-EU evidence requirements. The proposal stipulates that MSs do not duplicate JCAs at the national level, but complementary assessments may still be conducted (e.g., to account for different SoC across MS), provided they are justified, proportionate and the EC is notified (article 8).

The legislative process involving the EC, EP and European Council is still ongoing and did not conclude before the May 2019 EP elections. The EUnetHTA has been piloting JCA activities via the rapid relative effectiveness assessment program (RREA) (Kleijnen S, et al 2015 and 2016) aiming to create an effective, sustainable framework for collaboration among European HTA bodies.

As of June 2019, the EUnetHTA has published 15 guidelines that form the basis of a framework for JCA methodologies to support RREA in the EU. The production of the EUnetHTA Guidelines has been coordinated by EUnetHTA Joint Action (JA)1 WP5, EUnetHTA JA2 WP7 SG3 and EUnetHTA JA3 WP6B.

## 2.1. Project objective

The objective of this work was to develop materials that would enable the EFPIA HTA WG to engage in the political debate and process related to HTA methodology, in the context of the proposed EC's regulation on HTA.

Supporting objectives were:

- To review the methodological standards and challenges faced by manufacturers in developing JCAs under the new proposed EC's regulation on HTA by drawing on experiences of subject matter experts and ICON plc, based on the clinical evidence available at or around the time of MA



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- To provide a critical review of nine EUnetHTA methodological guidelines developed for performing a RREA of pharmaceuticals:
  1. Comparators and Comparisons
  2. Direct and Indirect Comparisons
  3. HRQoL and utility measures
  4. Clinical endpoints
  5. Composite endpoints
  6. Surrogate endpoints
  7. Internal validity of NRSs
  8. Internal validity of RCTs
  9. Safety
- To develop a final report and position statement for the EFPIA HTA WG to facilitate political engagement and advocacy on methodological standards with clear recommendations for industry to assist in engaging the policy debate:
  - Advocates for state-of-the-art “clinical” HTA, balancing the concepts of scientific rigour and flexibility as well as predictability versus evolution of science
  - Accounts for the levels of uncertainty that may be faced by HTA agencies and payers
  - Supports timely patient access and national decision-making based on the best available evidence



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

## 2.2. Approach

The project comprised two work streams:

- Review ED / scientific advice insights gathered from responses of the EFPIA-member companies to an online survey
- Critically review the nine selected EUnetHTA methodological guidelines

### 2.2.1. ED / Scientific Advice Online Survey

An online survey comprising 13 questions was developed in an iterative process with feedback from the EUnetHTA HTA WG steering committee members (see Appendix A: Online survey questionnaire). The format of questions ranged from numerical (defined as questions with integer responses), to scalar, qualitative and multiple choice. Multiple-choice questions had an option for the respondent to select 'Other' and provide a text entry for context.

This questionnaire was administered online using Qualtrics and required an average of 30 minutes to complete. Respondents were representatives nominated by each of the EFPIA member companies.

Responses were analysed and reported according to numerical, multiple-choice or free-text categorisations (Table 2).

**Table 2 Analytical approach**

| Type of question | Analysis outputs |
|---|---|
| **Numerical** | Summaries with n, range, median, interquartile range (IQR), mean, and standard deviation reported |
| **Multiple choice** | Summaries with n and % of respondents selecting each option, with the total number of participants completing the survey as the denominator and represented with graphs or summary tables |
| **Qualitative with free-text responses** | • Data-driven qualitative analysis for all questions, and handled differently depending on their length and complexity (except question E)<br>• For question E, free-text responses were mapped onto the following domains: Patient population, Comparator selection, Endpoints, Study design, HRQoL/utilities, General evidence generation, Analysis and controlling uncertainty<br>• Single-term answers (e.g., 'Neurodegenerative diseases') and free-text answers were recoded to new variables acting as response options to the relevant question<br>• Alternative spellings or terms for the same concept were accounted for |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

Respondents were asked to report key evidence requirements that they might seek as part of the ED/ HTA scientific advice process, and to identify evidence generation challenges associated with with state-of-the-art HTA methods.

### 2.2.2. Critical Review of EUnetHTA guidelines

#### 2.2.2.1. *Critical Review Framework Development*

A critical review framework was developed according to the following steps:

- Conducted a pragmatic search of literature for a pilot topic (i.e., NMA) to answer: (a) what is the recommended HTA methodology (state-of-the-art), (b) what is the currently accepted HTA methodology, and (c) what are the future trends for HTA in Europe?
    - A hand search of grey literature and also "pearling" reference lists of relevant articles was undertaken to gather evidence relating to best practices, checklists, frameworks and instruments to appraise guidelines. Ten key sources were selected for "pearling and snowballing"
- A de-novo critical analysis framework was developed based on an iterative process (inception-elaboration-transition)
    - Retrieved evidence relating to frameworks, concepts and principles were synthesised in a fit-for-purpose analytical tool that were reviewed and adapted by subject matter experts

The final Critical Review Framework included two key domains (Figure 1):

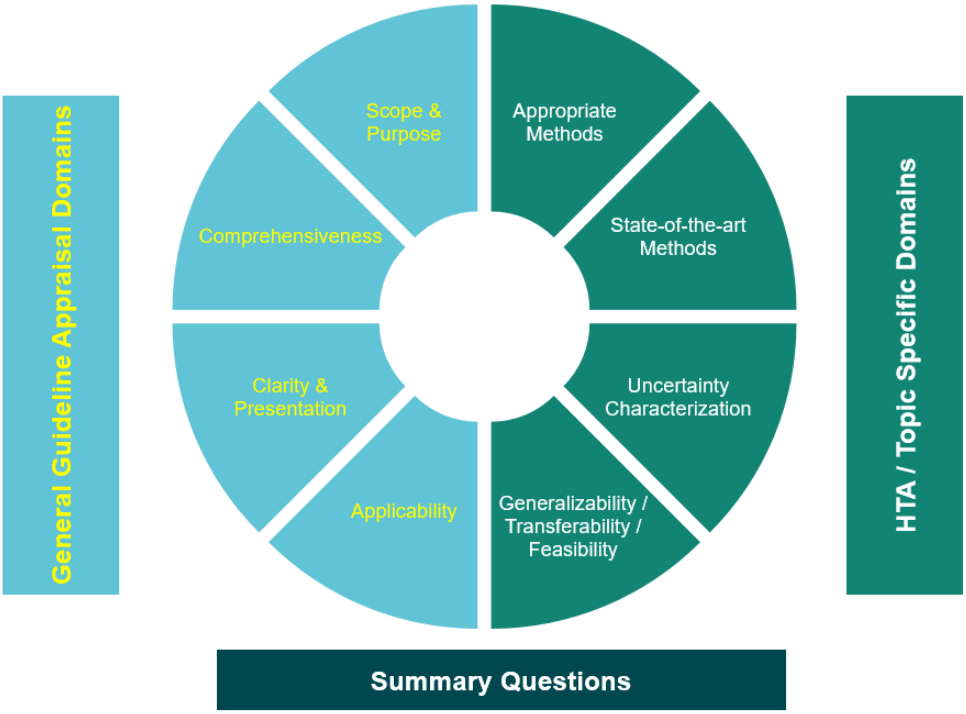
- **General guideline appraisal**: Informed by guideline development methods (Schünemann HJ, *et al*. 2006), assessment instrument of the guideline development process (The AGREE Collaboration, 2003), and analysis of EU HTA systems (Angelis A, *et al*. 2018)
- **HTA methods/topic specific**: Informed by Drummond et al. (2008) key HTA principles (Drummond MF, et al. 2008), EUnetHTA HTA adaptation toolkit (EUnetHTA 2011), and International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Practices for Outcomes Research (ISPOR 2019).

Questions relating to each domain of the Critical Review Framework were developed and are listed in Appendix B: Summary questions in Critical Review Framework.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

**Figure 1. Critical Review Framework**




### 2.2.2.2. *Selected EUnetHTA guidelines for review*

Nine out of 15 published EUnetHTA methodological guidelines (Table 3) were selected for review by a panel of subject matter experts.

**Table 3 Selected published EUnetHTA methodological guidelines subject to critique**

| Title | Version | Updated | Link |
|---|---|---|---|
| Comparators & Comparisons: Criteria for the choice of the most appropriate comparator(s) | 2.0 | 2015 | PDF Version |
| Comparators & Comparisons: Direct and indirect comparisons* | 2.0 | 2015 | PDF Version |
| Endpoints used for Relative Effectiveness Assessment: HRQoL and utility measures | 2.0 | 2015 | PDF Version |
| Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints | 2.0 | 2015 | PDF Version |
| Endpoints used for Relative Effectiveness Assessment: Composite endpoints | 2.0 | 2015 | PDF Version |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | | | |
|---|---|---|---|
| Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints | 2.0 | 2015 | PDF Version |
| Internal validity of RCTs | 2.0 | 2015 | PDF Version |
| Internal validity of NRSs on interventions | 1.0 | 2015 | PDF Version |
| Endpoints used in Relative Effectiveness Assessment: Safety | 2.0 | 2015 | PDF Version |

* Guideline under revision

The scope of the reviewed material was partly defined by the results of the online survey questionnaire.

#### 2.2.2.3. *Expert critique of EUnetHTA guidelines and insights*

Methods experts who contributed to the critique of the scope methodological guidelines are listed in Table 4.

**Table 4 Subject matter experts who critiqued the scope EUnetHTA guidelines**

| Guideline | Reviewer (affiliation) |
|---|---|
| **Comparators and Comparisons** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Kaisa Miikkulainen (ICON plc)<br>• EFPIA project steering committee* |
| **Direct and Indirect Comparisons** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Prof Neil Hawkins (University of Glasgow, UK)<br>• Prof Gianluca Baio (UCL, UK)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Victoria Paly (ICON plc)<br>• Vanita Tongbram (ICON plc)<br>• EFPIA project steering committee* |
| **HRQoL and utilities** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France) |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | |
|---|---|
| | • Prof Uwe Siebert (UMIT, Harvard University, Germany)Prof Andrew Briggs (University of Glasgow, UK)<br>• Dr Hannah Lewis (ICON plc)<br>• Kellee Howard (ICON plc) |
| **Clinical Endpoints** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)Prof Andrew Briggs (University of Glasgow, UK)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Axel Svedbom (ICON plc)<br>• EFPIA project steering committee* |
| **Composite Endpoints** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Prof Andrew Briggs (University of Glasgow, UK)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Axel Svedbom (ICON plc)<br>• EFPIA project steering committee* |
| **Surrogate Endpoints** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Prof Andrew Briggs (University of Glasgow, UK)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Axel Svedbom (ICON plc)<br>• EFPIA project steering committee* |
| **Internal validity of RCTs** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, , France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Iain Fotheringham (ICON plc)<br>• Kaisa Miikkulainen (ICON plc)<br>• EFPIA project steering committee* |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | |
|---|---|
| **Internal validity of NRSs** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Iain Fotheringham (ICON plc)<br>• Kaisa Miikkulainen (ICON plc)<br>• EFPIA project steering committee* |
| **Safety** | • Dr Carsten Schwenke (SCO:SSiS, Germany)<br>• Prof Isabelle Durand-Zaleski (University of Paris XI, France)<br>• Prof Uwe Siebert (UMIT, Harvard University, Germany)<br>• Prof Andrew Briggs (University of Glasgow, UK)<br>• Dr Manpreet Sidhu (ICON plc)<br>• Victoria Paly (ICON plc)<br>• Vanita Tongbram (ICON plc)<br>• EFPIA project steering committee* |

*The EFPIA project steering committee team members (affiliation) who provided insights were: Adam Parnaby (Celgene), Adrian Griffin (J&J), Ansgar Hebborn (Roche), Charlie Nicholls (Sanofi), Gesa Pellier (Novartis), Jake Lebiecki (Pfizer), James Ryan (AstraZeneca), John Borrill (BMS), Monica Daigl (Roche), Stephane Régnier (Novartis) and Tom Vanderbrouck (NovoNordisk)



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# 3. Results

## 3.1. Survey results

### 3.1.1. Study sample – participating companies

Twenty-five out of 28 contacted companies completed the survey questionnaire during the period 18 February to 25 March 2019 (Table 5).

**Table 5 List of participating companies (EFPIA members) in the online survey**

| Responding Companies | | |
|---|---|---|
| Amgen | Chiesi | Merck Sharp & Dohme |
| Astellas | Eisai | Novartis |
| AstraZeneca | Eli Lilly | NovoNordisk |
| Bayer | GlaxoSmithKline | Pfizer |
| Biogen | Ipsen | Roche |
| Bristol Myers Squibb | Janssen/J&J | Sanofi |
| Boehringer-Ingelheim | Leo Pharma | Servier |
| Celgene | Menarini | Takeda |
| | Merck KGaA | |

### 3.1.2. Engagement with HTA scientific advice over the past five years

Among the 25 respondents who responded to question A (*How many HTA scientific advice/early dialogue engagements have you personally or your company experienced within the past 5 years?*) of the Appendix A: Online survey questionnaire, a significant variance was observed in the number of HTA scientific advice/ED engagements (Table 6).

**Table 6 Number of HTA scientific advice/ED engagements (n=25)**

| | Personal experience | Company-wide experience |
|---|---|---|
| **Mean** | 4.6 | 10.8 |
| **Standard deviation** | 5.1 | 10.2 |
| **Median** | 3.0 | 8.0 |
| **Range** | 0–25 | 0–40 |
| **IQR** | 3.0 | 8.0 |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The mean and median number of company-wide experiences are more than double that of the personal experiences, and experience with HTA scientific advice/ED engagements is generally limited across individuals and companies, with outliers (companies/individuals) with significant experience.

### 3.1.3. Types of scientific advice previously needed/requested

All respondents (n=25) of question B (*Which of the following types of scientific advice/early dialogues have you previously needed/requested? Select all that apply from a predefined list*) provided input from the company perspective and related to advice for different assets. The majority of respondents reported seeking single-country HTA (88%) and regulatory/HTA (64%) scientific advice (Figure 2). Eight out of 25 (32%) sought two different types of early advice, 24% three types, 20% four types and only 8% all five types of ED/scientific advice listed.

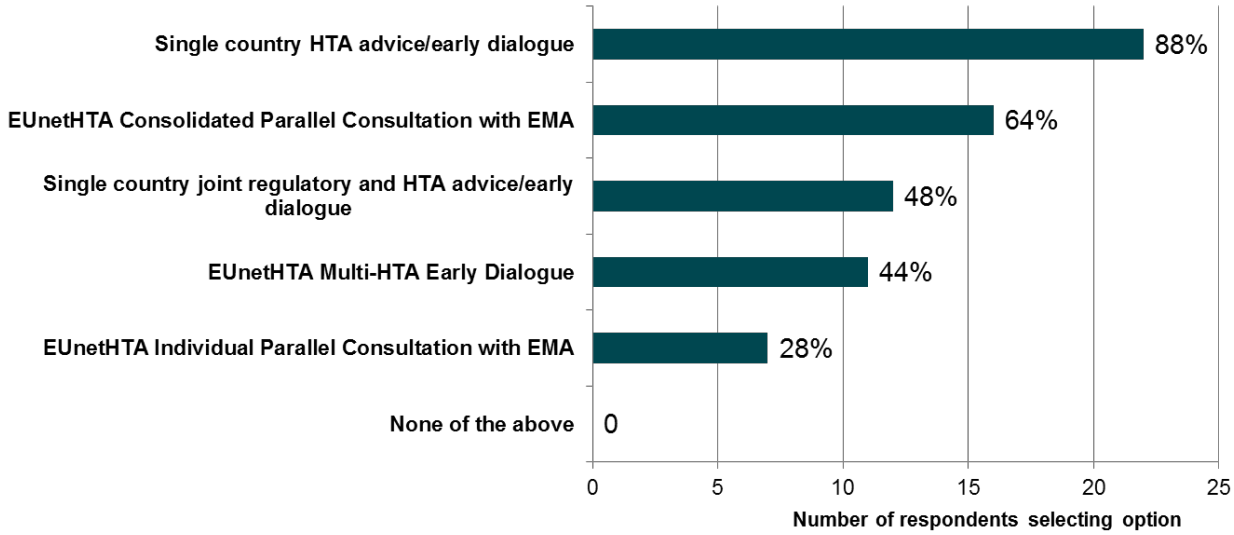



**Figure 2 Types of scientific advice needed or requested (n=25)**

The company size was not associated with the number of types of scientific advice reported to be sought, since one respondent was appointed from each company with potentially limited access to all ED/scientific advice activities occurring for all pipeline assets.

### 3.1.4. European HTA bodies where single country HTA advice was requested during the past 5 years

Among the 23 participants who reported participation in single-country HTA advice, their responses in question C (*In which European HTA bodies have you sought single country HTA advice during the past 5 years? Select all that apply from a predefined list*) indicated that advice was most frequenly requested from the Gemeinsamer Bundesausschuss (G-BA) in Germany (22 out of 23 respondents, 96%), National Institute for Health and Care Excellence (NICE) in the UK (78%), and Haute Autorité de Santé (HAS) in France (57%) as shown in Figure 3.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

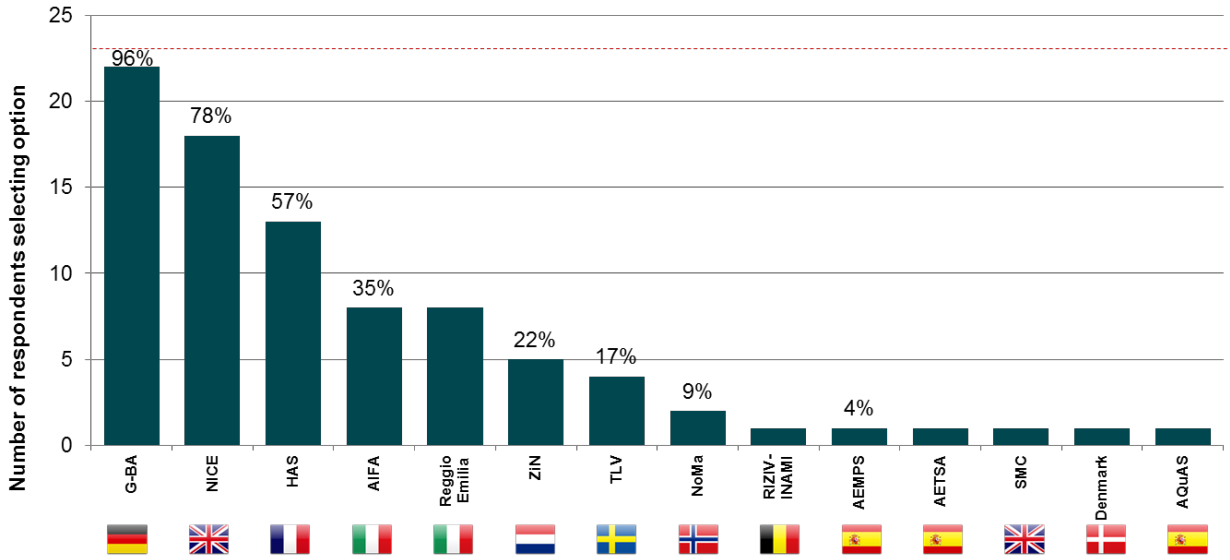



**Figure 3 National HTA bodies where single-country scientific advice was requested (n=23)**

None of the survey participants sought advice from HTA bodies in Finland, Austria, Czech Republic, or Portugal. The two survey respondents who did not request country specific advice in the past 5 years were not asked to elaborate on the reasons.

### 3.1.5. Therapeutic areas where single-country HTA advice was requested during the past 5 years

All respondents (n=25) indicated the therapy areas (TAs) in which their companies sought advice, (Question D: *Which of the following therapeutic areas have you requested HTA scientific advice/early dialogue? Select all that apply from a predefined list of 27 TA options*), with the majority of them being in oncology (17 of 25, 68%), Central Nervous System (CNS) (36%) and dermatology with hematology (32% each) (Figure 4). No advice was reported as being sought for endocrinology, ear, nose and throat (ENT), orthopedics, surgery or "other" TAs such as endometriosis, neuromuscular and rare diseases.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

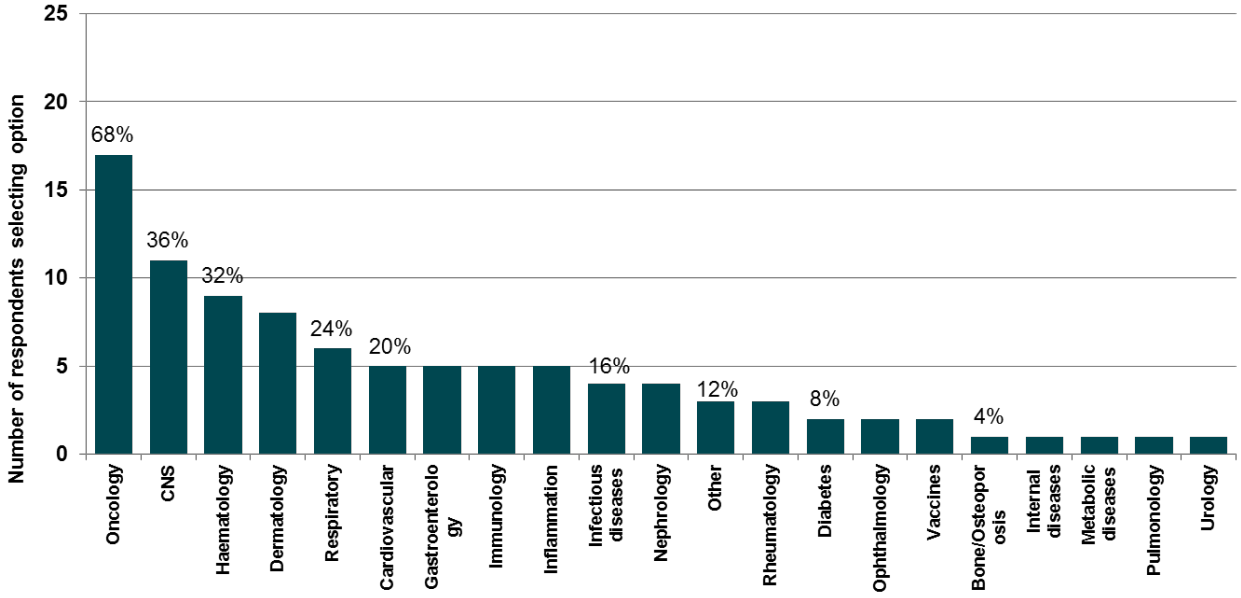



**Figure 4 Therapy areas where scientific advice was sought over the five past years (n=25)**

The reported top-5 TAs show the uncertainty around evidence generation uncertainty from clinical programs in this space, likely because of the evolving technologies and treatment algorithms. Companies seek advice on the acceptable evidence requirements to support relative value/clinical benefit increments.

### 3.1.6. Key evidence generation challenges that required clarity on HTA methods when developing the company position

In question E (*What were the three key evidence generation challenges that required clarity on HTA methods when developing the company position?*) respondents were prompted to provide their answers in free text . All responded with 24 recorded free-text answers that were mapped to the existing eight domains shown in question F (see section 3.1.7).One respondent indicated that the question was not understood and did not provide further details.

Comparator selection (72%), endpoints (56%) and study design (36%) were the most frequently occurring themes in respondents' answers (Figure 5). With regard to responses in the "other" option (7 of 25, 28%), participants indicated that clarity on HTA methods was required on the economic model (3 of 7), and patient-reported outcomes (PROs), direct vs. indirect evidence, country-specific advice and methods for vaccines (one each).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

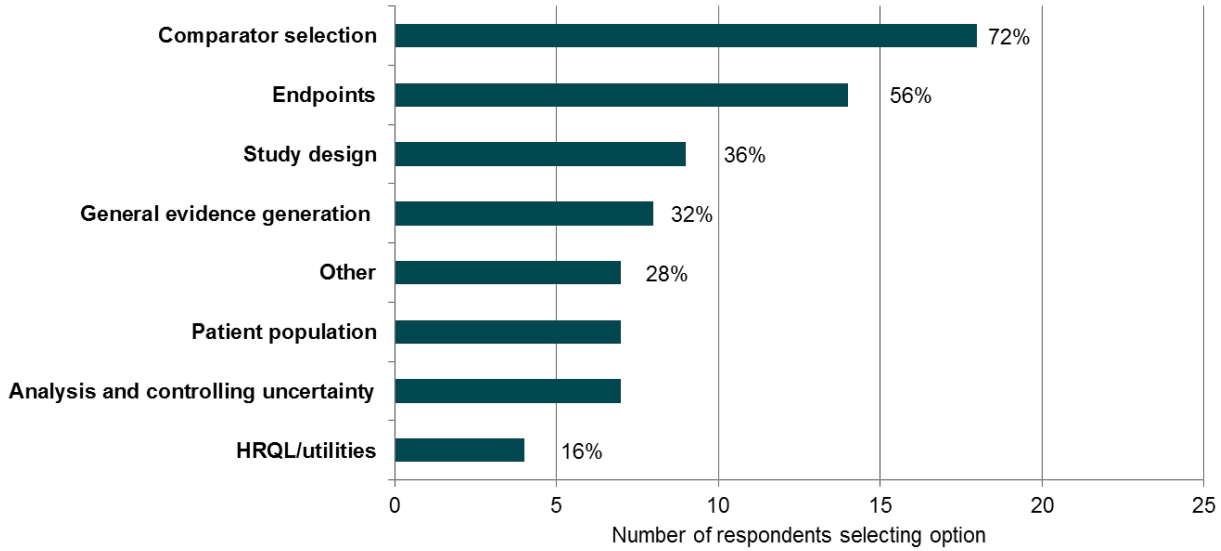



**Figure 5 Top-3 key areas of evidence generation challenges (free-text)**

### 3.1.7. Key issues that manufacturers seek scientific advice for

All respondents (n=25) completed question F (*What were the key issues that you sought advice for, among the following domains (i.e., not limited to HTA methods). Select from list of eight pre-defined option*) and indicated that comparator selection (24 of 25, 96%), patient population (92%) and endpoints (92%) were the most frequently occurring issues respondents sought advice for (Figure 6).

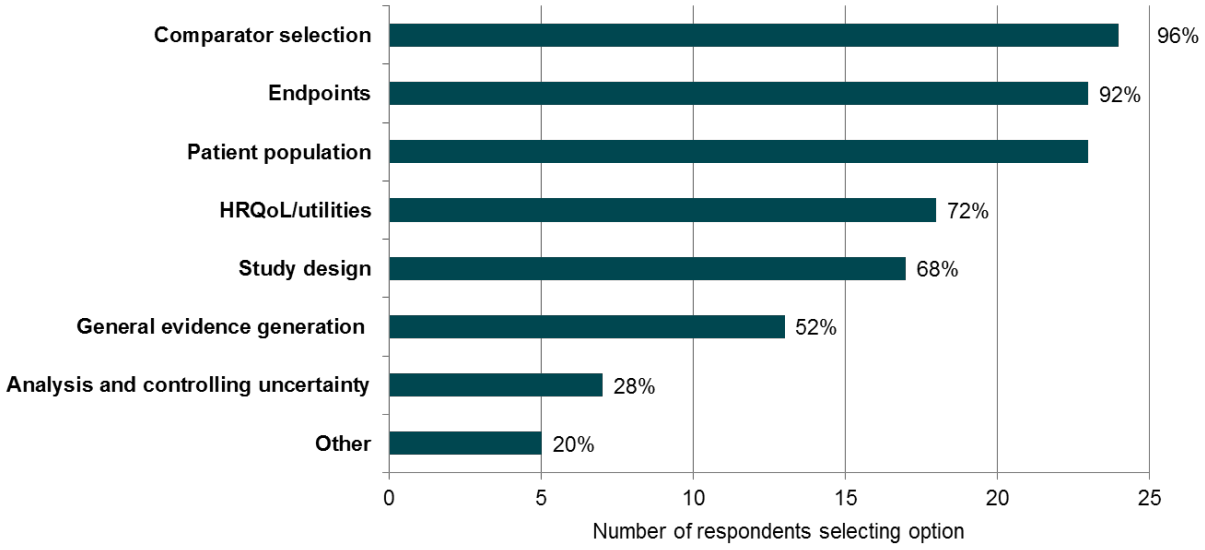



**Figure 6 Key issues to seek scientific advice (pre-defined list of answers)**

Respondents were required (mandatory) to insert free-text clarifications for each of their selections and upon qualitative analysis the recurring themes were identified and are listed below:



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- **Comparator selection**: Appropriate comparators including country and subgroup-specific, when to use placebo vs. SoC from country to country, and off label comparators
- **Endpoints**: Relevance and acceptance of surrogate and composite endpoints, selection of secondary endpoints, patient-relevant endpoints
- **Patient population**: Subgroups, identifying and demonstrating unmet need, clinical vs. reimbursed population
- **HRQoL/utilities**: Relevance and acceptance of PRO and HRQoL measures, when to use generic vs. disease-specific measures, acceptance of mapping methods to EQ-5D
- **Study Design**: Single arm trials and selection of a synthetic control group, study duration

### 3.1.8. Level of satisfaction on clarity and usefulness of additional information from HTA scientific advice

Respondents were asked to select from a scale of 1 (extremely unsatisfied) to 5 (extremely satisfied), the following:

- What is your level of satisfaction with the **clarity** of the additional information provided from the HTA scientific advice/ED for the key methodological issues you selected? (Question H1A)
- What is your level of satisfaction with the **usefulness** of the additional information provided from the HTA scientific advice/ED for the key methodological issues you selected? (Question H1B)

Patient population, followed by HRQoL/utilities, were the issues with the highest reported satisfaction for both clarity and usefulness of additional information from HTA scientific advice/ED (Table 7).

**Table 7 Satisfaction ratings for additional information received in scientific advice by methods domain**

| Methodological Domain | Satisfaction with clarity | | | | | Satisfaction with usefulness | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (No of Respondents) | Mean | SD | Median | IQR | Range | Mean | SD | Median | IQR | Range |
| **Comparator selection** (N=24) | 3.3 | 0.9 | 3.0 | 1.0 | 1.0–5.0 | 3.5 | 0.9 | 3.5 | 1.0 | 1.0–5.0 |
| **Endpoints** (n=23) | 3.4 | 0.8 | 4.0 | 1.0 | 1.0–4.0 | 3.4 | 0.8 | 3.0 | 1.0 | 1.0–5.0 |
| **Patient population** (n=23) | 3.8 | 0.9 | 4.0 | 1.0 | 2.0–5.0 | 3.9 | 0.7 | 4.0 | 1.0 | 3.0–5.0 |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **HRQoL/ utilities** (n=18) | 3.7 | 1.0 | 4.0 | 1.0 | 2.0–5.0 | 3.6 | 0.7 | 4.0 | 1.0 | 2.0–5.0 |
| **Study Design** (n=14) | 3.5 | 0.9 | 4.0 | 1.0 | 1.0–5.0 | 3.6 | 0.9 | 4.0 | 1.0 | 1.0–5.0 |
| **General evidence generation** (n=13) | 3.3 | 0.6 | 3.0 | 1.0 | 2.0–4.0 | 3.4 | 0.7 | 3.0 | 1.0 | 2.0–4.0 |
| **Analysis/ controlling uncertainty** (n=7) | 3.1 | 0.7 | 3.0 | 0.5 | 2.0–4.0 | 3.1 | 0.7 | 3.0 | 0.5 | 2.0–4.0 |
| **Other** (n=5) | 3.2 | 1.3 | 4.0 | 1.0 | 1.0–4.0 | 3.4 | 1.3 | 4.0 | 0.0 | 1.0–4.0 |

### 3.1.9. Assessing support received from HTA scientific advice

#### 3.1.9.1. *Divergence of advice between the regulatory and HTA bodies*

Respondents were asked if they *"identified any divergences in the advice received from the EMA and HTA bodies for your HTA scientific advice briefing package (i.e., parallel consultation of EUnetHTA with the EMA, https://www.eunethta.eu/services/early-dialogues/parallel-consultations/)?"*; upon a positive response, to provide clarity on the "*key underlying evidence and methodological issues in the feedback received that were conflicting or divergent between the EMA and HTA bodies (e.g. use of RWE, MAIC to combine RCT with RWE)?*" (Questions G1 and G2).

Among the 18 respondents who sought parallel advice (72% of the total), half (9 of 18, 50%) identified divergence from the advice they received between European regulatory (EMA) and HTA bodies. Six of nine (67%) survey respondents who reported divergence indicated comparator selection followed by study design being as the main underlying issues.

#### 3.1.9.2. *Advice received among HTA bodies*

Participants were asked to provide their input on whether "*HTA scientific advice/ED helped to plan and/or develop the required evidence needed for the HTA submissions, provided that the clinical program was successful and advice was implemented?*" (Question H2) for each of the eight domains considered. As shown in Table 8, the HTA scientific advice/ED was reported to support companies (i.e., help develop and



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

confirm the evidence planning) mostly in selection/definition of the comparator(s) (61%), endpoints (61%) and patient population (54%).

**Table 8 Assessment of support from the HTA scientific advice**

| Methodological Domain | Yes | | No | | Confirmed optimal evidence plan | | Cannot answer | |
|---|---|---|---|---|---|---|---|---|
| (Number of Respondents) | n | % | n | % | n | % | n | % |
| **Comparator selection** (N=24) | 12 | 52% | 2 | 9% | 4 | 17% | 5 | 22% |
| **Endpoints** (n=23) | 11 | 48% | 3 | 13% | 3 | 13% | 6 | 26% |
| **Patient population** (n=23) | 11 | 46% | 2 | 8% | 3 | 13% | 8 | 33% |
| **HRQoL/utilities** (n=18) | 7 | 41% | 1 | 6% | 5 | 29% | 4 | 24% |
| **Study Design** (n=14) | 7 | 39% | 1 | 6% | 5 | 28% | 5 | 28% |
| **General evidence generation** (n=13) | 5 | 39% | 2 | 15% | 3 | 23% | 3 | 23% |
| **Analysis/controlling uncertainty** (n=7) | 1 | 14% | 1 | 14% | 1 | 14% | 4 | 57% |
| **Other** (n=5) | 3 | 60% | 1 | 20% | 0 | 0% | 1 | 20% |

Among respondents who thought that the scientific advice positively supported their evidence planning (i.e., table 6 column "yes"), they were asked to elaborate on "*How did the HTA scientific advice/ED on these key issues you identified earlier, listed below, help you to plan and/or develop the evidence for the HTA submissions? Where discussed as a need please provide feedback on recommendations for methodologies to handle post-licensing evidence generation (PLEG) data*" (question H3). The responses collected are presented according to the key themes:

- **Patient population** (n=12): Validation and acceptability of study population and subgroup analysis



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- **Comparator selection** (n=11): Selection of additional comparators, including plans for ITC and data generation beyond the clinical program, validation of selected comparators
- **Endpoints** (n=11): Endpoint selection and validation (including PROs and analysis hierarchy for secondary endpoints), alignment of clinical development and additional supportive data
- **Study design** (n=7): Optimisation and validation of study design, including duration and randomisation schedule
- **HRQoL/utilities** (n=7): Endpoint selection and consideration for additional PROs
- **General evidence generation** (n=5): Data gaps and RWD for conditional and post-approval requirements
- **Analysis and controlling uncertainty** (n=1): Scientific advice helped to inform the analysis plan
- **Other** (n=3): Shaped company's strategic thinking beyond the clinical development program, estimate potential upside in access from extending trial readouts

Furthermore, participants were queried on "*Other useful methodological topics not in the guidelines developed by HTA bodies, and not discussed in previous questions, that the HTA scientific advice/ED report provided*" (question H4). Reported answers were qualitatively analysed and the key themes that emerged along with quoted examples are shown in Table 9.

**Table 9 Additional method insights from ED/scientific advice**

| Theme (Number of respondents) | Examples of comments from responses |
|---|---|
| **None** (n=5) | [No responses] |
| **Interpretation of guidelines** (n=4) | "*Sometimes advice gives insights about HTA bodies' current thinking which may not yet be reflected in the guidelines*", "*Discuss borderline or difficult to interpret with guidelines cases*" |
| **Analytical/statistical approaches** (n=4) | "*Information on addressing longer term data uncertainty*", *"Validation of the SAP", "Statistical strategy Subgroups analysis and stratification method"* |
| **PRO** (n=4) | "*Valuable advice was provided on the acceptance of specific PRO instruments*", "*Optimal placement of PRO instrument before- during - after treatment*" |
| **Subpopulations** (n=3) | "*How to handle… biomarker-defined patients populations*", "" |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | |
|---|---|
| **External experts opinion** (n=3) | "*Real value is in the discussion during the meeting where you get insight from the experts, although the final report may not cover some of these*" |
| **Comparator** (n=2) | "*[Feedback on]… country-specific comparator regime… duration and dose*" |
| **ITC** (n=2) | "*Best approach for undertaking an indirect treatment comparison* " |
| **Other** (n=2) | *"Did not find the interaction useful in any manner unfortunately* " |
| **Evidence requirements** (n=1) | "*Evidence requirements to support additional benefits beyond positive clinical outcomes data*" |
| **Economic model** (n=1) | *"Valuable input for the economic modeling"* |

A question on the divergence of recommendations from different HTA bodies during ED/scientific advice (Question I) focused on each of the themes that respondents selected in question F (section 3.1.7). Responses indicated that consensus was more frequent for patient population and study design (Table 10).

**Table 10 Consensus among HTA bodies and manufacturers' action in case of disagreement**

| Methodological Domain | Disagreement among HTA bodies? | | Action taken |
|---|---|---|---|
| (Number of respondents) | **No (%)** | **Yes (%)** | **Yes (% of respondents who indicated disagreement)** |
| **Patient population** (n=23) | 12 (52%) | 11 (48%) | 5 (45%) |
| **Comparator selection** (n=24) | 9 (38%) | 15 (63%) | 8 (53%) |
| **Endpoints** (n=23) | 11 (48%) | 12 (52%) | 8 (67%) |
| **Study design** (n=17) | 10 (59%) | 7 (41%) | 7 (100%) |
| **HRQoL/Utilities** (n=18) | 9 (50%) | 9 (50%) | 6 (67%) |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | | | |
|---|---|---|---|
| **General evidence generation** (n=13) | 6 (46%) | 7 (54%) | 5 (71%) |
| **Analysis and controlling uncertainty** (n=7) | 2 (29%) | 5 (71%) | 4 (80%) |
| **Other** (n=5) | 2 (40%) | 3 (60%) | 3 (100%) |

Among those who reported lack of consensus in the advice received, the majority of them indicated that they took corrective action (see last column in Table 10). Indicative examples of the reported actions taken were:

- Looking into alternative sources of data
- Having informed discussion within the product team, agreeing on trade-offs, next steps, and priorities
- Optimising the study design to the country with the greatest evidence requirement need
- Further checked the validity of endpoints through a discussion with key opinion leaders
- Further stratified clinical trial patients at randomisation and conducted additional subgroup analyses.

Specifically for the EUnetHTA multi-HTA ED and consolidated parallel consultation, 18 respondents (of 25, 72%) indicated that consensus was reached amongst HTA bodies (Questions K1–2). Among those who reported that consensus was not reach (7 of 25, 28%), four indicated that it was not applicable (e.g., not participated in consensus-building exercise) or not clear, two were not able to comment/discuss, and one respondent stated consensus was not reached for surrogate endpoints, cross-over study design and comparators.

In multi-HTA EDs, only two out of 25 respondents (8%) indicated that their question was denied and national advice recommended instead (Questions L1–2); one noted EUnetHTA did not accept the indication as relevant for a single-arm trial and the other as not being part of a multi-national dialogue.

### 3.1.10. Manufacturers' recommendations for future/additional methodological guidelines

The final question asked participants to provide "*recommendations for future/additional methodological guidelines related to preparation for EU Joint Clinical Assessments? Specifically of the recommendations suggested, which are the most pertinent and achievable given the changes under consideration within a broader HTA framework (including ex-EU HTA bodies), as well as comments relating to methodological issues that were not covered by the previous questions (e.g., predictability/consistency of advice, state-of-the-art methods, digital endpoints)?*" (Question M). The answers varied significantly across respondents, with four identified themes emerging:



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- **Face-to-face meetings**: 8/25 (32%) of survey respondents provided recommendations for the process of receiving HTA advice or participating in EDs that included allowing for structured patient input and input for selection of HTA body representatives
- **Generalisability or specificity of guidelines**: 6/25 (24%) of survey respondents provided recommendations for how general or specific methodological guidelines should be. There was disagreement regarding the level of prescriptiveness and generalisability of the guidelines, and how tailored they wanted methodological guidelines to be for each country
- **Topics for discussion**: RWD, study design (single-arm trials and cross-over)

No further recommendations were given by four out of 25 (16%) survey respondents.

## 3.2. Critical review results

### 3.2.1. Comparators and Comparisons

The Comparators and Comparisons: Criteria for the choice of the most appropriate comparator(s) guideline (2015) outlines the criteria for the choice of the most appropriate comparators and provides a summary of current policies and best practice recommendations to support REAs.

#### 3.2.1.1. *Scope & Purpose*

The overall objective of the guideline is to provide advice on internationally agreeable best practice recommendations for the selection of the most appropriate comparator for both pharmaceuticals and non-drug interventions. In addition, it outlines potentially expected challenges when choosing the comparator for specific assessments.

It can be noted that the scope of the guideline should cover the willingness to agree on a compromise if the SoCs are different between countries. In the presence of varying legislation between EU MS, either the comparator needs to be a compromise or legislations and rules for the SoCs need to be adjusted across borders.

The scope of the guideline should also consider how different sources for SoC, such as registries, are assessed in terms of quality and appropriateness. Moreover, the acceptability of placebo as a relevant comparator for joint REA is not covered in the guideline. Additional guidance on the use of investigator's or clinician's choice should also be provided.

It is important to address the situation of different comparators in subpopulations, with the EUnetHTA guideline being moderately prescriptive of situations where different comparators are used for specific subpopulations (i.e., "define comparator interventions separately for all subgroups and sub-indications, depending on the evidence available").

The problem statement highlights the importance of recognizing the explicit aim of the assessment when selecting the most appropriate comparator. According to the guideline, important factors to consider are:



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- How should comparator selection be established (e.g., recently approved technologies, well-established treatments, non-approved treatments, non-drug interventions), and how to perform the analysiswhen more than one intervention is considered the appropriate comparator
- The timing of comparator selection should be considered within the evolving therapeutic landscape and shift of the SoC from the time a clinical study is designed to the time of the HTA

#### 3.2.1.2. *Comprehensiveness*

The guideline does not determine comprehensive or explicit decision criteria for comparator selection. However, it recommends evidence requirements to support the chosen comparator (e.g., a preference-organised list where evidence should be found to support the claim that the intervention is used in routine clinical care, and guidance in cases where national clinical guidelines contradict each other). Furthermore, elaboration on off-label therapies or therapies that have temporary approval as relevant comparators would provide guidance to the target audience.

#### 3.2.1.3. *Clarity and presentation*

The guideline outlines a list of 10 recommendations on how to choose the most appropriate comparator. Three of these specify “if required by national procedural rules”, while there is no information on which countries these apply to. Many aspects of the recommendations depend on national policies rather than scientific aspects; thus, they could be interpreted differently by different agencies. The guideline summarises current policies across Europe, but lacks detail and specificity.

The guideline also presents two options with supporting details for instances where there is no pan-European reference comparator; evidence needs to be available to show the chosen comparator is routinely used in clinical practice, and the comparator intervention should be validated for the respective clinical indication or population with available evidence.

#### 3.2.1.4. *Applicability*

The guideline is mainly an overview of national policies and does not identify any organisational barriers or potential cost and time implications. It can be assumed that HTA agencies’ budgets and national drug budgets may be limited, and differ between countries, leading to different types organisational barriers to adopting new medicines.

The guideline does not provide recommendations for comparator selection for innovative pharmaceuticals which have just entered the market or received a positive reimbursement recommendation, thereby changing the “best SoC”. This is typically tackled on a case-by-case basis, but it can be assumed that this may require the manufacturer to conduct additional research if the agencies suggest a different comparator.

A recent example of a scenario where selected comparators differed across HTA agencies is the case of the new migraine preventive product erenumab (Aimovig, Novartis), a calcitonin gene-related peptide (CGRP) monoclonal antibody. The comparators accepted by NICE in England (draft GID-TA10302) and the Dental and Pharmaceutical Benefits Agency (TLV) in Sweden (1558/2018) differed as the selection of



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

comparator for the HTA appraisals were informed by data of actual prescription and utilisation of existing treatments and routine clinical practice. (NICE ID1188 2019; TLV Aimovig 2019).

#### 3.2.1.5. *Appropriate methods*

The comparator approaches in the guideline align with currently used HTA methods with regard to the use of subpopulations, approaches in choosing a comparator in case of multiple alternatives, consideration of the lowest-cost treatment, reimbursed treatment, treatment of the same class, legal considerations and off-label use. However, it would be beneficial to place more emphasis on the inclusion of non-drug interventions when they are applicable to the same patient patient population and treatment line, and are relevant as comparators for drug interventions.

The guideline also considers that it may be necessary to define comparator interventions separately in case the new intervention has a wide therapeutic indication and is currently treated differently in subgroups in the MSs.

#### 3.2.1.6. *State-of-the-art methods*

The choice of comparator depends on the local health care system, HTA guidance and legal and cultural context rather than an explicit methodology. Ideally the comparator should be identified from clinical practice guidelines at the European or international level, from the published scientific literature, with EU marketing authorisation or other recognised regulatory authorisation in the assessed indication.

In many circumstances there is no SoC consensus across MSs since real-life populations inevitably vary. Country-specific guidelines need to be consulted to identify a relevant comparator (e.g. HAS, NICE, G-BA).

Survey results (Kristensen 2019) of MS HTA agencies indicated that the majority of the HTA institutions provided identical answers regarding the methodology on choosing a comparator. The main challenge in comparator selection is balancing between the comparator technology or technologies most likely to be replaced versus those supported by evidence of their efficacy and safety profiles. A systematic review of HTA methodologies across eight European countries identified that the comparator is usually the SoC and can consist of multiple options (Angelis A *et al*. 2018).

It is noted that future research should investigate which comparators were selected in the conducted RREAs (JCA pilot phase) in order to gain insight into how HTA agencies have collectively selected comparators.

#### 3.2.1.7. *Uncertainty characterisation*

The guideline notes that the overall uncertainty in choosing an appropriate comparator lies with the differences in national policies combined with the specific assessment question. The guideline aims to address this uncertainty by providing several options for the selection process. However, there is no clear path to follow and the choice of comparator would depend on the agencies involved.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

According to EUnetHTA, local restrictions occur in 79% of the countries, whichmay cause uncertainty in the acceptance of a chosen comparator at the local level (*EUnetHTA. An Analysis of HTA and Reimbursement Procedures in EUnetHTA Partner Countries: Final Report 2018)*. There is no method provided on how to address this uncertainty.

#### 3.2.1.8. *Generalisability, transferability and feasibility*

The choice of comparator depends on the local health care system, legal and cultural context; therefore, the recommendations provided in the guideline are not expected to be fully transferable across EU jurisdictions.

In fact, the Institute for Competitiveness (iCOM) states that the choice of comparator is the most debated issue for the JCAs at the transnational level, in part due to variations in the SoC across countries and different processes for choosing the right comparator (iCOM 2017). These differences are acknowledged in the guideline's problem statement and the recommendations are open to interpretation across the EU in the search for common ground.

#### 3.2.1.9. *Conclusions*

The main methodological gap identified in the Comparators and Comparisons guideline pertains to establishing consensus on the SoC selection amongst EU MSs and the lack of predictability that is caused by the imprecise method of determining comparators.. Moreover, further guidance on the SoC selection is needed in terms of the level of efficacy and effectiveness evidence required by the assessors, and how different sources of SoC are assessed for quality and appropriateness. The guideline does not take into consideration the early alignment of comparator selection in terms of early scientific advice from regulatory and HTA agencies, which is considered an important gap.

The guideline provides a variety of recommendations for comparator selection and provides guidance on navigating some of the uncertainty surrounding the choice of comparator. However, it is also recognised that the process is often driven by individual MSs, the treatment landscape, clinical guidelines, regional restrictions, policies and budgetary aspects and cultural context of the respective health care system.

The document may support predictability in certain countries when conducting relative effectiveness evaluations; however, the uncertainty of comparator selection will remain as long as MSs have varying policies.

Furthermore, no clear recommendations are provided regarding the timing and predictability of comparator selection. For example, the final decision on the comparator in Germany is made alongside the decision on the added benefit (i.e., the comparator may change during the assessment). This procedure may lead to a assessment of there being no benefit due to a lack of evidence because the dossier provides evidence against the "wrong" comparator.

The current and evolving treatment pathways should be considered in comparator selection; new drugs may replace or move current therapy to a later stage (e.g., ALKs, TKIs, PDL1s may shift chemotherapy to



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

later lines). It would be beneficial to provide guidance on the timely use of European clinical practice guidelines appropriate for future JCAs.

#### 3.2.1.10. *Recommendations for internal position paper*

The key recommendations addressing the identified gaps in methodology are:

- Establish a willingness to agree and the means to find agreement, through a transparent, timely and consistent process among EU MSs, on what constitutes SoC treatment(s)
- Establish the selection of an appropriate basket of comparators that would be informed by clinical guidelines, physician preference and their hierarchy based on the quality of evidence (e.g., RCT first)
- Discuss and establish alignment between regulatory and HTA ED guidance on comparator selection

A transparent, timely and consistent process by which EU MSs agree on a basket of SoCs is essential to inform future JCAs. Establishing the willingness to agree and the means to find agreement through a transparent, timely and consistent process among EU MSs on what constitutes SoC treatment(s) is recommended for future JCAs. The comparator selection, and the scoping process leading to the decision, should be driven by clinical experience such as physician preference to identify a basket of relevant SoC treatments. In the scoping process, stakeholders should discuss and align on the basket of treatments, their relevance, physician selection criteria with patients' unmet need(s) and direct patient engagement in the decision process.

Further guidance on the SoC selection is needed in terms of level of evidence of efficacy and effectiveness required and how different SoC sources (e.g., registries) are assessed for quality and appropriateness. Moreover, further detailed guidance would be beneficial when there is no clear main comparator, such as when more than one comparator is available in order to define a basket of treatments; this would also shed light on situations where there are local restrictions in the selection of comparators. It would be beneficial to consider a hierarchy of a manageable set of comparators with the ones approved with RCT data to be on top and those with registry data/RWD to be lower in this hierarchy.

Early comparator discussions would be beneficial for all stakeholders. Alignment in regulatory (EMA) and HTA ED guidance on comparator selection should be promoted and divergences discussed openly. Learnings from the regulatory guidance (EMA and FDA) should be integrated in the JCA and similar criteria to reach consensus should be applied in the HTA.

A clear description of the different approaches to comparator selection currently in place in EU MSs and determing which of these is the benchmark, would be beneficial to inform future JCAs. It is suggested that EUnetHTA presents a summary table of comparator selection methods by country (i.e., usual care, market share, most used, supported by information on the sources of obtaining this information in each country).

### 3.2.2. Direct and indirect comparisons

The Comparators & Comparisons: Direct and Indirect Comparisons guideline (2015) focuses on the methods available for treatment comparisons. The guideline covers the strengths and limitations as well as



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

recommendations to support Relative Effectiveness Assessors in their activity. Indirect comparison is defined in the guideline as the estimation of the relative effectiveness of two or more treatments in the absence of any head-to-head trials.

#### 3.2.2.1. *Scope & Purpose*

The overall objectives of the guideline are clearly described but limited to describing the main methods of direct, indirect and mixed treatment comparisons in terms of the "types of relationships they can model". It is not self-evident as to what exactly this refers and the assumptions inherent in them.

The problem statement is: "What methods for direct and indirect comparisons are used; are more advanced methods like Bayesian mixed treatment comparison used?" The problem statements are only partially addressed regarding what types of methods of direct and indirect comparisons are used as there are material omissions on newer approaches, and different variables beyond the basic types.

The usefulness of the guideline to decision makers and the intended audience is limited due to the lack of a comprehensive summary of newer approaches, including the use of NRSs. Furthermore, the guideline should give a comprehensive view on available methods for single-arm studies, which are quite common in rare oncology indications, and discuss optimal study planning (or adjustment) to make them more comparable (i.e., the application of direct or indirect comparisons relies on the assumption that only comparable studies should be combined).

#### 3.2.2.2. *Comprehensiveness*

The guideline provides a list of evidence requirements in section 2.1; however, this is not comprehensive as there are other forms of evidence that are not summarised or considered (Philippo *et al*., 2018).

The guideline focuses mainly on iseRCTs and briefly mentions the use of observational (NRS) data without further context of how these study designs can be incorporated. In fact, there is little or no discussion of other study types such as observational and non-comparative studies, different variable types such as time-to-event, ordinal and time-to-event variables (Faria *et al*., 2015). Currently, there is a sufficient body of research to inform the planning and analysis of observational studies (using the target trial approach) so that these studies may be used to emulate clinical trials (Hernán MA et al., 2016; Kuehne F et al., 2019; Lodi S et al., 2019).

The utilisation of individual patient data is mentioned in the guideline; however, this discussion does not incorporate additional methods (e.g., MAIC).

Overall, the guideline provides a very general overview of decision criteria for the different methods (i.e., direct, indirect, and mixed treatment); however, there is ambiguity in terms of which methods are more appropriate for which types of evidence. An understanding of study designs and the body of evidence to support which method is suitable or appropriate is not provided (i.e., performing an MAIC when there is an established network of evidence is not advised). The discussion of Bayesian and frequentist methods is



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

limited and arguably outdated and the comparison of the Bayesian vs. the “Bucher Method” is somewhat stereotyped.

#### 3.2.2.3. *Clarity and presentation*

The guideline provides recommendations that are very general. For example, recommendation 10 states that “indirect comparison should only be carried out if underlying data from comparable studies are homogeneous and consistent”. This does not highlight the importance of assessing baseline characteristics and inclusion/ exclusion criteria.

The guideline presents methods for each form of analysis (i.e., direct, indirect and mixed treatment) while the selection of which method is the most appropriate remains ambiguous. The strengths and weaknesses are considered among each approach, which is one of the main objectives of the document. However, this appears to contradict the recommendations as there is no clear guidance on the most appropriate method to employ.

#### 3.2.2.4. *Applicability*

Overall, the guideline does not reflect practical HTA decision making. Different EU HTA bodies prefer very different approaches for ITCs (e.g., NICE prefers MAICs, whereas G-BA the Bucher method if the studies are sufficiently similar). Summaries of different methods are highlighted in a very general manner followed by a comparison of the strengths and limitations of each, which may lead to confusion.

Furthermore, the guideline does not provide clear recommendations on when to consider each method. For example, different networks of evidence are presented; however, the methods to employ and special considerations are not clearly outlined or summarised.

The guideline does not discuss tools for application. Documents such as the NICE Decision Support Unit (DSU) are more practical for highlighting best practice and providing appropriate information to execute different analyses (i.e., sample codes).

#### 3.2.2.5. *Appropriate methods*

Appropriate methods for when to consider MAIC and the use of observational studies (NRS) are not covered by the guideline. Historical comparison of single treatment arms of different studies is not covered, although it is quite important in many orphan drug applications.

The guideline recommends (p. 9) for meta-analyses, besides others, the effect measure risk difference. However, in meta-analyses, only relative effect measures should be used. Absolute measures should be avoided as they cannot be interpreted in case the studies are not sufficiently similar. In addition, no clear recommendations on how to assess heterogeneity are provided.

Currently employed methods may be missing due to the age of the document (published in 2015) and recent development of more advanced approaches. While the guideline provides a very general overview of the quantitative approaches and a discussion of the strengths and weaknesses of each method, further



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

discussion of methods in detail is warranted given the exclusion of other forms of analysis other than direct, indirect and mixed treatment comparisons.

#### 3.2.2.6. *State-of-the-art methods*

The guideline does not elaborate on state-of-the art methods and the summaries lack significant context and rationalisation. Although the purpose of the document is not to provide technical guidance, the authors in some situations over-simplify the methods by providing blanket rationalisations without concise support.

There are major analytical methods that are not summarised but need to be considered (i.e., use of NRSs, new statistical methods such as MAIC and STC, and NMA of survival data using fractional polynomials). Further discussion would be beneficial of the implications of conditional approval drug therapies, and appropriate methods for summarising data in unique situations where there is a limited amount of evidence (e.g., orphan designation). Guidance should be also provided on how to perform indirect comparisons that include both randomised and observational studies in one meta-analysis (Schnell-Inderst P. *et al.*, 2017).

Future research should investigate best practice methods to employ for treatment comparisons among new modalities of treatment (e.g., cell and gene therapies, targeted therapies with a previously unstudied biomarker) and evidence synthesis to support orphan drug submissions.

#### 3.2.2.7. *Uncertainty characterisation*

Methodological uncertainty is addressed throughout the document and summaries, notably within the critical comparison of methodologies (Section 2.5) and discussion. Sources of uncertainty identified in the guideline include heterogeneity, fixed and random effects (which and when to choose), publication bias, outliers and influential studies, sensitivity to priors in Bayesian analysis and dose-response.

The objective of the guideline is to highlight the strengths and weaknesses of each method of analysis (direct, indirect and mixed treatment comparison). The guideline seems biased toward RCT evidence, the primary focus of the document, and it appears to favour the Bucher method, highlighting that it is “computationally straightforward” and more “transparent”. Although the authors advise against using an inappropriate method of analysis, they do not clearly define a framework from which to reliably determine which method is the most appropriate.

#### 3.2.2.8. *Generalisability, transferability and feasibility*

The guideline is generalizable enough to be applicable across a wide variety of geographical locations. There are limitations as to how indirect treatment comparisons (ITCs) are regarded across EU HTA bodies, which has a major impact on how the data in submissions should be presented and assessed. Moreover, the methods are to a certain extent transferable across EU jurisdictions, although there may be organisational barriers to implementation. The recommendations remain general; however, the high-level framework of the advice implies that geographical differences may prevail even if the guidelines are implemented.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

HTA-related aspects not covered by the guidelines include: different types of data availability, whether or not ITCs are considered in new drug submissions and organisational barriers to implementation. The guideline does not support different types of evidence submissions (e.g. NICE highly specialised technologies, HST, which are special submissions for which general submission rules may not apply).

#### 3.2.2.9. *Conclusions*

The main methodological gaps in the guideline pertain to following areas; inclusion of key indirect comparison methods (e.g. MAIC or STC), detailed guidance on the use of observational or iseNRSs, when to employ qualitative versus quantitative summaries of data and further guidance to account for heterogeneity. Overall, the limited consideration of methods restricts the applicability of the guideline across appraisals and HTA bodies. The revised version of this guideline should consider the significant body of scientific and methodological literature published after 2015.

#### 3.2.2.10. *Recommendations for Internal Position paper*

The key recommendations addressing the identified gaps in methodology are:

- To underline the importance of early communication and alignment among manufacturers, regulatory agencies and HTA bodies
- To update the guideline to reflect newer state-of-the-art methodologies used in major HTAs
- To gain agreement across the EU jurisdictions that ITC is an acceptable method for joint REA
- To provide further guidance on situations where an ITC may not be appropriate
- To provide further guidance on the most appropriate methods to use when non-proportional hazards are observed in time-to-event RCT data (e.g., fractional polynomial and piece-wise constant models; Gsteiger S, *et al*. 2017)
- To provide further guidance on how to use RWE (e.g., hierarchical Bayesian framework)

The guideline should be revised in order to accommodate a revised framework which researchers can use as guidance for the most appropriate analysis (i.e., application of the methods in the form of examples and sample codes). It would also be beneficial to provide a decision-tree approach outlining the required methods and limitations (e.g., network feasibility [Yes/No], if Yes then considers effect modification, etc.). The decision tree should describe the type of ITC by case (e.g., if there is no RCT to account for a time-to-event endpoint or when to use qualitative and quantitative data). Details on how far to move toward the Bayesian, intensely complex, statistical modeling that requires expert knowledge to elucidate problems are also considered necessary for a revised version of the guideline.

Furthermore, the guideline could be revised to include a summary of the newer approaches such as MAIC and STC. Detailed guidance on the use of observational (NRS) studies as the key evidence base and additional variables such as time-to-event survival (outcomes) beyond binary and continuous outcomes should be accounted for.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

### 3.2.3. HRQoL and utility measures

The Endpoints used for Relative Effectiveness Assessment: HRQoL and utility measures guideline (2015) provides recommendations for the selection and assessment of HRQoL when completing an REA of pharmaceuticals.

#### 3.2.3.1. *Scope & Purpose*

The overall objective of the guideline is to address the suitability of different types of HRQoL instruments for the purposes of assessing relative effectiveness, both from a generic and disease-specific HRQoL perspective. These are addressed by discussing: a) types, and pros and cons, of relevant HRQoL instruments for REA and cost-utility assessment, b) potential issues with HRQoL data that should be considered in a REA, and c) what existing guidelines say about HRQoL measurements in the context of reimbursement request. The guideline also clearly states that the objective is not to discuss "development of HRQoL instruments, the use of instruments in clinical practice, for case-mix adjustments in financing health care services or for assessment of the health status of the population" (p. 14)

The findings and recommendations are applicable and relevant for both assessors and researchers, both of whom are clearly defined as the target audience. However, many of the recommendations in the guideline apply to protocol design and are, therefore, more relevant to researchers.

The problem statements are specifically described: a) variance in results from different HRQoL instruments reduces comparability of results across studies, and b) development of common guidelines is challenging due to varying use of REA in reimbursement decisions across countries. Thus, it would be useful to include an outline of country-specific differences and requirements (e.g., French guidelines state that HRQoL should be measured in diseases where QoL is compromised).

The guideline clearly addresses differences between HRQoL instruments and its potential impact on REA. However, it does not clearly address country differences in regards to REA, nor does it provide any guidance on ways to bridge the gap between the EU countries' requirements and their different practices. For example, some countries focus on cost-utility analysis and others on clinical decision making, both of which usually require data from the same trial and need to be considered simultaneously when selecting an instrument and analytic approach.

#### 3.2.3.2. *Comprehensiveness*

The guideline provides a comprehensive overview of HRQoL for REA to inform reimbursement decisions, cost-utility assessment, clinical decision making, and evidence requirements according to the REA purpose. Differences between country priorities and/or requirements for the assessment are noted but not elaborated on. Decision criteria according to the purpose of the REA are comprehensively discussed, but a visual overview (e.g., decision matrix) could be beneficial to include for additional clarity.

Generic HRQoL instruments and disease-specific HRQoL instruments are discussed in detail, and the importance of both types is highlighted. However, there is no clear guidance on preference for trial- vs.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

literature-based utility values beyond indications that the former are preferred. The guideline also incorporates methods into the review of literature and the decision criteria.

There are no major omissions with regards to methods. Information on the requirements for validity, reliability, and sensitivity of the HRQoL instruments is limited, but the guideline includes cross-references to other documents or guidelines containing further details.

#### 3.2.3.3. *Clarity and presentation*

The key methods are clearly presented. The guideline generally gives specific recommendations on HRQoL instruments to be used for the purposes of REA, and overall recommendations cover both generic and a disease-specific instruments. Clear caveats to each type of HRQoL instrument in the context of REA are also provided.

The recommendations call out the need for documentation of the validity, reliability, responsiveness and acceptability of the HRQoL instruments, though further guidance on appropriate methods to demonstrate these properties would be valuable. Moreover, the need to understand thresholds for meaningful change or differences (for example, through MCID, minimal important difference, or RDs), which is a critical element for the interpretation of scores, is noted but not highlighted. Finally, country-specific differences are mentioned as important considerations for evidence collection and presentation, but with little detail.

#### 3.2.3.4. *Applicability*

The guideline clearly addresses which HRQoL instruments should be considered appropriate for different purposes of REA, and could practically be utilised in the HTA decision making process. However, the guideline is not supported with any tools for application.

It would be beneficial to provide concrete details on the differences among EU countries and narratives on how to bridge these gaps. Presently, not all countries have their own EQ-5D tariffs; so, the guideline could include recommendations on which to use. For example, there is no French value set for the 5L despite the French guidelines supporting the use of EQ-5D-5L; thus, the 3L crosswalk value set is accepted.

Potential organisational or industry barriers are not discussed in the guideline. However, the difficulty of many decisions around HRQoL for REA needing to be made at the protocol and study design stage is mentioned.

#### 3.2.3.5. *Appropriate methods*

In general, the recommended methods are appropriate. Inclusion of both disease- or population-specific and a generic HRQoL instrument are principally recommended for most adequately capturing the impact of a disease.

This guideline largely presents a quantitative approach along with consideration of missing HRQoL data and should include advice on how to reduce missing data and how to handle missing data from an analytic perspective:



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- Careful consideration of participant burden in number of length of HRQoL instruments and also in frequency and nature (i.e., mode of administration) of assessment; participant reminders and site training
- Statistical approaches that are robust to missing data (e.g., mixed effects models) should be considered, and methods to impute missing data should be used; sensitivity analyses should be used to assess robustness
- Advice on how to approach pooling of data from trials would be warranted, and the appropriate use of survival analyses for HRQoL endpoints

The insensitivity of some generic HRQoL instruments (e.g., EQ-5D) in some conditions (e.g., ophthalmology) also warrants discussion; broad domains captured in generic instruments can fail to capture condition-specific QoL concepts. Although it mentions the importance of the validity of the HRQoL instrument, little information is given on how this should be assessed. The guideline notes that "disease-specific instruments might not capture unexpected changes in dimensions of HRQoL" (page 17-18), but this is unlikely to be an issue if the HRQoL instrument has undergone appropriate content validation with qualitative methods in the target population. Therefore, some additional discussion on the qualitative methods required to assess the content validity of the HRQoL instrument could be valuable.

Additionally, the guideline would benefit from a section on the importance of the timing of administration, the factors to consider when determining the assessment schedule, how to handle the effect of acute events on HRQoL, and how to consider participant burden when collecting patient-reported data; advice on need to collect beyond an event (e.g., progression) when time to deterioration must be established.

On page 12 it is stated that single item HRQoL questions are not considered valid for REA or cost-utility analysis; however these questions can still provide useful information and if part of a validated instrument (e.g., the VAS as part of the EQ-5D) can be utilised to examine the impact of a technology on HRQoL.

The guideline conclusions mention the importance of defining clinical meaningful change, but this should be discussed in more depth: meaningful change definitions should be estimated using anchor-based approaches in the context of use (and supported by distribution-based approaches) plus consideration of incorporating qualitative data from patient interviews (e.g., trial exit interviews).

#### 3.2.3.6. *State-of-the-art methods*

The guideline covers basic principles for determining the appropriateness of an HRQoL instrument based on context of use, but also highlights the lack of state-of-the-art methods. International Society for Quality of Life Research, ISPOR and the FDA guidelines are referenced for further information to assess validity, reliability and responsiveness and acceptability of an instrument.

Page 24 notes that "measurement equivalence testing should have been performed before application of the electronic version of an originally paper questionnaire in a clinical trial" per the referenced ISPOR Task Force paper on ePROs. However, the field has since moved on and no longer views the need for full equivalence testing as an absolute requirement (Muehlhhausen *et al*., 2015).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The guideline also mentions CAT (page 24) approaches are under development, and correctly notes that this is an approach in its infancy in health research. This is a useful approach to help reduce participant burden through reducing the number of items completed. The increased focus on item banks (e.g., PROMIS) also warrants some discussion as to the extent to which HTA bodies accept use of such systems, and the requirements in terms of evidence of content validity.

Traditionally, COA instrument development has used Classical Test Theory, which is also the approach to validation of instruments that is recommended by the FDA (2009). Nevertheless, uses of Modern Test Theory, such as Item Response Theory (IRT), are becoming more popular (particularly for the development of item banks – see above), in order to assess item performance. Guidance on the usefulness of IRT in the development of COA instruments is warranted in the guideline.

The guideline also mentions that "in REA the relative importance of the different HRQoL domains needs to be determined in order to draw conclusions with respect to the "net" relative effectiveness of an intervention" (page 16). Although the narrative that surrounds this discussion is sound, the weighting of domains in a multi-dimensional instrument relates to the development of the instrument and the scoring protocol. To alter this would require discussion with the instrument developer, and it is also noteworthy that regulators (as other parties interested in HRQoL data arising from a clinical trial) tend to prefer simple summary score methods.

Consideration of norms for HRQoL are very useful in the interpretation of the findings; however, these norms data are often quite outdated.

#### 3.2.3.7. *Uncertainty characterisation*

The guideline addresses and takes a clear and detailed position on methodological uncertainties related to HRQoL instruments, such as repeated measurement, cultural adaptation and translation, missing data, modes of administration, and evaluation by proxies. The document is broad in scope, does not take a stance related to a particular HTA agency, is not biased in terms of data assessment, and provides recommendations for REA in various contexts

Mapping of disease-specific to generic instruments or generic non-preference-based measure (PBM) instruments to generic PBM instruments have been identified as sources of uncertainty, and the guideline does not recommend mapping. However, in a landscape of hundreds of COA instruments, mapping algorithms can be an extremely useful tool to understand differences between results across studies given they are robustly developed and validated. This can be useful for estimating utilities in a scenario where, for example, utility data were not directly collected in the trial and the data in the literature are limited.

#### 3.2.3.8. *Generalisability, transferability and feasibility*

The proposed methods are both transferable and feasible across the EU jurisdictions, and take into consideration whether a country considers cost-effectiveness in their decision-making process. However, current country-specific HTA preferences for HRQoL instruments and current practices for REA in relation



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

to HRQoL are not discussed in detail. The guideline could also benefit from commenting on cases of very severe orphan diseases where the EQ-5D may not be suitable for capturing small improvements.

This guideline supports evidence requirements for HRQoL for REA to inform reimbursement decisions and cost-utility assessment. It discusses the value and use of both generic- and disease-specific HRQoL instruments. Overall, the guideline acts as a recommendation but does not ensure national uptake of the assessments given the variation between country HRQoL requirements and the lack of a no gold standard in this area.

#### 3.2.3.9. *Conclusions*

The guideline aims to support predictability in the methods used in REA evaluations and consistency of evidence reviewed; however, there is room for HTA-specific preferences to overshadow the recommendations.

The guideline could further address the potential for extreme lifetime health gains in HRQoL through further discussion of the ceiling effects on HRQoL instrments that may come as a result of new and innovative health technologies. The availability of population norms for scores on instruments is important for assessing whether patients have returned to a population average QoL is valuable; however, data on norms are often outdated.

The guidelines should note that the validity, reliability, sensitivity and applicability (and estimates of meaningful change) of HRQoL instruments is a continuing process. New contexts of use, including new technologies, may lead to further evidence on these properties being gathered to support the use of the HRQoL instrument.

As HRQoL instruments are required to collect data for both REA and cost-utility analyses, the differing needs cannot be considered in isolation. Moreover, the need for HTA bodies and payers to be seen as consistent across decisions and/or over time is not yet addressed.

#### 3.2.3.10. *Recommendations for Internal Position paper*

The guideline provides a well-considered and clearly presented review of HRQoL for REA, and includes recommendations for appropriate HRQoL instruments to consider and include in HTA submissions. Caveats are included for each method proposed, and it is noted that a gold standard may be impossible to achieve. While HTAs across the EU may decide to consider these recommendations, there will undoubtedly be variation in the extent of adherence due to variability in country requirements and preferences. Nevertheless, there are some missing recommendations, for example available types of analyses relevant for HRQoL (e.g., time to deterioration) and how these should be reported.

The guideline could benefit from separating COAs for use to assess HRQoL and estimation of utilities into two sections with a clear definition of data and methods. While connected, these areas are separate and more clarity could be achieved by presenting them as Part A (HRQoL: for comparative clinical effectiveness assessments) and Part B (Utilities: geared toward the cost-effectiveness analysis).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The following points are also worth addressing:

- Use of a systematic review to understand country-differences in practices for conducting REA
- The challenges presented by assessment of HRQoL in acute diseases, rare diseases, and novel therapies
- The need for evidence of thresholds for meaningful change or differences to aid interpretation of scores
- Considerations in the analysis/interpretation of data (e.g., use of norms for HRQoL and adjusting for sociodemographic factors such as age)
- Time points of administration of HRQoL instruments, participant burden, long-term data capture (e.g., post-progression) and exit interviews to establish treatment impact and to interpret meaningful COA scores
- Further detail on thresholds of meaningful change or differences, including approaches to their estimation and how to apply the thresholds
- Statistical approaches to use for this type of data and how to handle missing data
- Some further consideration of qualitative evidence for instrument selection and establishing the content validity of HRQoL instruments
- Qualitative evidence is imperative to ensure the instruments selected are appropriate and relevant for the patient population; in a more asymptomatic condition psychological effects such as fear of passing a genetic condition to future generations might a greater impact on HRQoL
- Reassessment of the discussion around the need for equivalence testing
- Discussion and recommendation on the use of iseCAT, item banks, and meta-analyses to establish HRQoL evidence in an indication
- Mapping should remain optional due to the practical issues of introducing new COAs, and the inclusion of a generic instrument (e.g., EQ-5D) in clinical trials should be considered
- Recommendations on statistical methods for mapping or preferences for mapping to domains / responses to create a health profit which can have a value set applied to it
- The value of data from mixed-methods approaches in the context of clinical trials
- The HRQoL field is rapidly evolving in terms of new instruments and evolving stakeholder opinions; for example, modern test theory was not previously endorsed by regulators for the development of COA instruments, but acceptance is growing for its use
- New FDA guidance is expected in 2020 on COAs and insights presented should be also taken into consideration for an updated EUnetHTA guideline

### 3.2.4. Clinical endpoints

The Endpoints Used for Relative Effectiveness Assessment: Clinical Endpoints guideline (2015) outlines the selection and assessment of clinical endpoints when completing a REA. Clinical endpoints are broadly categorised into three domains: mortality, morbidity and HRQoL.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

#### 3.2.4.1. *Scope & Purpose*

The stated objectives of the guideline are specifically addressed (i.e., to describe the common characteristics of clinical endpoints; describe issues relating to their measurement and presentation; briefly outline some of the problems arising when comparing or pooling clinical endpoint data from a number of studies; and provide a set of recommendations for the selection and the interpretation of clinical endpoints when completing an REA). It is important to reference COA endpoints in general, rather than just HRQoL, as this is just a subset of COAs.

Three problem statements are partially addressed albeit at a high level (i.e., which clinical endpoints are accepted for the assessment, how are (absolute, incremental, relative) differences between treatments assessed, and what is the role of absolute and relative differences?). The statement "*What clinical endpoints are accepted for assessment*" was not specifically addressed, as the guideline describes desirable properties of clinical endpoints, but without stating which should be used for the assessment. Additionally, the guideline does not elaborate on the time horizon of an assessment or the follow-up duration in clinical studies.

Whilst it is stated that the "*[..] EUnetHTA JA1 WP5 guidelines focus on [..] challenges encountered by HTA assessors [..]*", the audience/perspective and context of the remit of the guideline merit additional clarification. The guideline appears driven by regulatory considerations rather than HTA-related requirements, and therefore it is unclear whether the scope of the guidelines is fit for purpose for the stated audience or the decision problem(s) (e.g., recommendation [13] on hierarchy of end-points, and [14] on adjustments for multiple testing may not be relevant in an HTA framework). A clear definition of the audience and the context of the guideline would improve its usability among HTA bodies that focus on assessing effectiveness in clinical practice and, therefore, have a different perspective and approach on uncertainty to regulatory bodies (e.g., EMA).

#### 3.2.4.2. *Comprehensiveness*

A comprehensive list of evidence requirements for the selection and interpretation of clinical endpoints is given in the guideline scope. It focuses on one type of evidence only (clinical endpoints), but within this restriction allow for a variety of evidence types (morbidity, mortality, COAs). The guidelines provide an overview of decision criteria for assessment of the suitability of clinical endpoints; however, beyond a discussion of thresholds for meaningful change in scores on COAs, there is limited guidance on how particular outcomes are deemed patient-relevant. Additionally, comment on when the use of observer reported outcomes or clinician-reported outcomes (ClinROs) is acceptable or preferable would be useful to include in the guideline.

The guideline provides a thoughtful high-level overview of methods applicable for the choice and analysis of clinical endpoints. Nevertheless, granular information on methods is partially disease-specific and hence arguably out of scope for a general guidance. Given the HTA focus of the document, a more detailed



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

discussion of endpoint considerations for pragmatic use would be important, but may be less relevant in the specific context of a rapid REA.

There is a lack of discussion of the justification of endpoints that may be relevant to a specific subpopulation (but not the entire population) and also lack of guidance on endpoint assessment in unblinded studies, and on how to compute the MCID for endpoints.

#### 3.2.4.3. *Clarity and presentation*

The key recommendations are specific and unambiguous; nevertheless, some require additional elaboration, for example:

- ”All-cause mortality is the most unbiased endpoint”, while arguably cause-specific mortality may be more pertinent in certain cases, unless cross-over is allowed
- If a product only impacts a single dimension of mortality (e.g., cardiovascular risk) then overall mortality may arguably be biased due to competing risks
- Additional explanation and/or examples would be beneficial for the question of clinical relevance

The guideline provides high-level recommendations on methods for selection, measurement, presentation and pooling of endpoints with methods being clearly, although not extensively, described.

#### 3.2.4.4. *Applicability*

The guideline partially reflects practical HTA decision making. However, it does not provide extensive advice on situations where endpoint selection has been driven from a regulatory perspective and where endpoints are not directly relevant to HTA

The guideline does not consider differences across EU practice, but acknowledges that endpoints should be independent of jurisdictions (Recommendation 9, p. 6). Furthermore, the guideline does not discuss tools for applying endpoint selection. Tools supporting the implementation of the guidelines may be processes or protocols for regulatory-HTA parallel scientific advice, ensuring that important issues for clinical endpoints are addressed (Tafuri G, *et al*. 2016). Potential organisational barriers to implementation of the guidance are also not explicitly discussed.

#### 3.2.4.5. *Appropriate methods*

Most, but not all, methods are reflective of current practices by EU HTA agencies considering the differing approaches among countries. Adjustments for multiple endpoints are not universally relevant to HTA bodies (Leverkus F, *et al*. 2016) and qualitative with quantitative approaches to addressing the research problems are described. Examples include reproducibility and validity of endpoints (Section 2.1.5) and presentational aspects of endpoints (Section 2.2). In the context of reproducibility, discussion on intra-subject variability in self-reported end-outcomes may be warranted (e.g., for measurements of visual acuity) (Siderov J, *et al*. 1999).

Statements on the preference for OS, use of hierarchical endpoints, and COA endpoints as complementary endpoints only could be further discussed. Furthermore, potential regional biases in COA endpoints,



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

analysis of total versus sub-scale scores for COAs, the relationship between AEs and morbidity, the interpretation of Kaplan Meier curves where death is a censoring event (Schmoor C, *et al*. 2013), and challenges with subgroup analysis may merit elaboration.

#### 3.2.4.6. *State-of-the-art methods*

The guideline provides high-level recommendations for a complex problem where method development progresses at a measured pace. It presents a reasonable first attempt to define the landscape and set research objectives for endpoint selection in an HTA context. However, there are four recent developments that may be covered in more detail:

- Estimands may be explicitly discussed given the recent focus on the concept, although components of estimands are addressed in the guidelines; a detailed discussion should include case studies (e.g., in many oncologic trials if the comparison is between a new therapy and the SoC with treatment switching upon progression, the estimand must not allow the comparison to the SoC with switching (ITT analysis [Latimer *et al*., 2016])
- Evidence requirements for therapies that provide long-term clinical benefit and/or the potential for cure may be discussed in more detail. These therapies have specific requirements for surrogate markers (Scott M, *et al*. 2019; Unkel S, *et al*. 2018)
- Presentation of survival data may be extended to comprise different measures at different time-points, especially in oncology
- The rapid proliferation of health apps may provide novel methods to measure aspects of health (e.g., sleep) and would merit further discussion.

In terms of future research, given the potential economic impact of therapies that provide long-term clinical benefit and/or the potential for cure, guidance on relevant clinical and surrogate markers for these therapies may be important areas of research. Furthermore, improved opportunities for joint and timely advice from HTA bodies and regulatory agencies may decrease risk and improve outcomes for all stakeholders.

#### 3.2.4.7. *Uncertainty characterisation*

Methodological uncertainty is addressed throughout the published guideline and repeated recommendations are made to minimise uncertainty stemming from endpoint selection. Other sources of uncertainty have been identified and discussed, including the measurement of clinical endpoints and statistical issues including power considerations and multiple testing, as well as the impact of (non-)blinding (Sections 2.1.1 and Sections 2.2).

The focus of the guideline is on clinical trials and endpoint selection/evaluation. To the extent the audience does not have influence over the endpoint selection, the guideline does not elaborate on the appropriate course of action if the primary endpoint(s) are not patient-relevant. Given the focus on rapid clinical assessment, data from pragmatic clinical trials are likely to be unavailable; however, discussion of endpoint selection for pragmatic trials may still be warranted. Furthermore, there is no mention of extrapolation beyond the end of a trial period or imputation of missing data.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

#### 3.2.4.8. *Generalisability, transferability and feasibility*

Methods described in the guideline are transferable across EU jurisdictions, although there may be organisational barriers to their implementation. Given the document's scope, additional potentially HTA-related aspects not covered by the guideline include pragmatic clinical trials and a formal framework for validating surrogate endpoints post-market entry. The recommendations are general and may promote uptake of the proposed methods, but the high-level framework of the advice implies that important inter-country variation may remain even if the guideline methods are adopted.

#### 3.2.4.9. *Conclusions*

Overall, the guideline promotes consistency and predictability of the HTA process, and given the lack of detail, variations in HTA agencies' positions on clinical endpoints may remain even if it is adopted. For example, recommendations on standardised methods to determine the thresholds for the clinical importance of COAs may increase the consistency and predictability of the HTA process; overlap with the recommendations presented within the HRQoL/utilities guideline (section 3.2.3) should be taken into consideration. Further guidance on analysis of time-to-event data would promote consistency and predictability while further details are needed in the following methodology areas:

- Evidence requirements for therapies that provide long-term clinical benefit and/or the potential for cure
- Discussion of estimands and sensitivity analysis in RCTs
- Determination of whether specific requirements may be applicable for health apps
- Endpoint selection for pragmatic trials
- Guidance on the determination of thresholds for meaningful change or differences in endpoints

#### 3.2.4.10. *Recommendations for Internal Position paper*

Based on the EUnetHTA position that OS should be the gold standard for assessing clinical benefit, further guidance should be provided on the level of evidence and analytical methods that would be acceptable. The broad remit of the guideline calls for specifications to be made on a case-by-case basis, and thus the importance of early communication and alignment among manufacturers, regulatory agencies and HTA bodies should be underlined.

Furthermore, in certain areas, standardisation of methods may be feasible and improve consistency and predictability of decision-making among HTA bodies. The separate guideline on surrogate markers (see critique in section 3.2.6) suggests that EUnetHTA also considers surrogate markers to be an important topic. However, neither the guideline on clinical endpoints nor the guideline on surrogate markers provides detailed recommendations for the validation of surrogate markers. Further clear guidance is needed on methods to adjust OS if cross-over from the control to the experimental arm occurs, and how to estimate OS in trials where non-proportional hazards are observed.

Whilst the guidelines cover important aspects of endpoint definition and measurement, the updated ICH guideline arguably merits more attention with regard to the estimands (Scott M, *et al*. 2018). Further advice,



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

for example, on the definition of long-term assessment in terms of observation time would be valuable, and ED/scientific advice should be the stage to define endpoint acceptance by HTA bodies in Europe.

### 3.2.5. Composite endpoints

The Endpoints Used for Relative Effectiveness Assessment: Composite Endpoints guideline (2015) presents a set of recommendations and aspects to be considered for the assessment and interpretation of results of composite endpoints.

#### 3.2.5.1. *Scope & Purpose*

The overall objectives of the guideline are specifically described and addressed (i.e., advantages and disadvantages of the use of composite endpoints as opposed to single endpoints; and guidance for assessors with regard to construction, reporting and interpretation of composite endpoint results). The problem statements covered by the guidelines are specifically described (i.e., advantages and limitations of composite endpoints for REA, and methodological pitfalls related to the use, interpretation and assessment of composite endpoints). Additionally, the stated audience is clear: assessors are clearly stated in the guideline objectives.

Considering that the design of composite endpoints may be beyond the control of the assessors and thus the advice provided is limited, the guidance may not be fully fit-for-purpose for its stated audience.

The scope of the guidelines is fit for purpose for the stated decision problem. However, an HTA specific focus would be valuable in the context of REA, especially with more detailed guidance on the interpretation of composite endpoints and handling of missing data.

#### 3.2.5.2. *Comprehensiveness*

The guideline provides a comprehensive list of evidence requirements for the construction, reporting and interpretation of composite endpoints. It discusses a variety of endpoints that may be included in the composite (e.g., clinical events, PROs, ClinROs, caregiver-reported outcomes and laboratory values), in Section 2.1.

The guideline provides comprehensive and clear decision criteria (Recommendation Section), and a comprehensive list of methods for analysis of composite endpoints (Section 2.4). No important evidence requirements or methods are omitted, although the recommendation that heterogeneous endpoints (mix of subjective and objective endpoints) should be avoided is debatable. For example, composite endpoints comprising HRQoL and clinical disease severity scores are recommended for clinical decision-making in dermatology (Mrowietz U, *et al*. 2011). Furthermore, formal weighting of endpoints is mentioned but not expanded upon, while the role of Multi-Criteria Decision Analysis would be beneficial to elaborate on.

#### 3.2.5.3. *Clarity and presentation*

The key recommendations are specific and unambiguous while the methods are clearly presented (Section 2.4) in the guideline.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

#### 3.2.5.4. *Applicability*

The guideline partially reflects practical HTA decision-making, but does not provide extensive advice on situations where composite endpoints have been constructed for explanatory studies. These are supporting regulatory requirements, rather than HTA and therefore do not fulfil the requirements for composite endpoints in HTA (REA). The guideline acknowledges that the weighting of components may differ between countries (Summary Section, paragraph 8); however, there is no further elaboration on this key issue. Furthermore, no support tools are provided and organisational barriers to implementation are not discussed.

#### 3.2.5.5. *Appropriate methods*

The proposed methods are largely reflective of current practices by HTA agencies in the EU. As previously noted, the recommendation to avoid heterogeneous composite endpoints may be contested (see section 3.2.5.2).

Qualitative and quantitative approaches to addressing the research problem are described and clear recommendations are provided. However, the relevance of the *a priori* definition of the composite endpoint needs to be stressed. The suitability of composite endpoints for HTA should be further discussed as composite endpoints are often defined for regulatory purposes; non-binary endpoints and the role of repeated measurement analysis of continuous outcome (endpoint) variables, as well as which operationalisation to use, may also merit more discussion in the guideline.

Composite endpoints that differ between trials for the same conditions also make indirect comparison difficult; it would be useful to have a set of acceptable composites, or a set of endpoints to be used alone or in combination for a given condition in order to allow meta-analyses and indirect comparisons. Further guidance would be especially beneficial on the validation of new composite endpoints.

#### 3.2.5.6. *State-of-the-art methods*

The guideline provides recommendations that cover most aspects already included in the published literature. Nascent areas that may merit more discussion include the weighting of individual endpoints in the composite and recurrent events analysis.

In terms of future research, a number of avenues are suggested in the literature: establishing common criteria for the composition of composite endpoints and weighting of their components have been advocated (Armstrong PW, *et al*. 2017). This key issue is especially important in HTA, where there may be specific requirements on endpoints; for example, for links between risk factors and endpoints, if all risk factors are equally important for all endpoints within the composite then it is more powerful to work with the composite. However, if some risk factors (e.g., blood pressure) are more important for a given endpoint (e.g., stroke) within a composite (e.g., major adverse cardiovascular events) then it could be advantageous, from an HTA perspective, to conduct formal competing risks analysis than to analyse the composite endpoint.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

New insights into the weighting of endpoints to produce refined composites may advance the field (Ahmad Y, *et al*. 2015) and statistical methods for recurrent events analysis, within the context of composite endpoints in trials, deserve further attention and guidance (Anker SD, *et al*. 2016).

#### 3.2.5.7. *Uncertainty characterisation*

Methodological uncertainty is addressed in the guideline and recommendations are made to minimise uncertainty stemming from construction (Recommendations 5,6,7), reporting (Recommendations 9-12), and meta-analysis (Section 2.4).

The other main source of uncertainty discussed is missing data, which is specifically pertinent to composite endpoints, where a subset of the components comprising the composite may be missing (Section 2.4).

The guideline does not appear biased in terms of data assessment. However, there is limited guidance on practical decision-making for assessors that have no input on the choice of composite endpoint (e.g., discussion on the relevant course of action if one component of the composite endpoint is not clinically relevant) (Leverkus F, *et al*. 2016).

#### 3.2.5.8. *Generalisability, transferability and feasibility*

The outlined methods are transferable across EU jurisdictions, although there are barriers to their implementation given that HTA agencies may take different views on specific components of composite endpoints and/or their relative importance. Given the guideline scope in the context of JCA, relevant HTA-related aspects appear to be covered and included recommendations are informative and instructive, with the aim of promoting uptake of the proposed methods.

#### 3.2.5.9. *Conclusions*

The recommendations listed in the guideline are clear and interpretable and hence promote consistency and predictability of the HTA-process. Overall, the key gaps that should be elaborated in an upcoming version are:

- Discussion of weighting of components of the composite
- Statistical methods for recurrent events of composite endpoints
- Further discussion of competing risks

#### 3.2.5.10. *Recommendations for Internal Position paper*

The usefulness of the guideline would improve if it focused on interpretation and presentation for HTA assessors, which is its target audience, instead of taking a regulatory perspective.

Further clarification would be beneficial on the methods to validate a new composite endpoint for new technologies. Currently, the guideline provides a thoughtful and clear overview of issues important in construction, reporting and interpretation of the results of composite endpoints. Establishment of common criteria for the composition of endpoints and weighting of components would advance the field, especially if the guideline recommends which criteria are available.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

Discrepancy in the construction of composite endpoints for regulatory and HTA purposes underline the need for early consolidated regulatory and HTA parallel advice. Further research into the weighting of components of composite clinical endpoints, and improved understanding of competing risks and recurrent events have the potential to improve the acceptance of composite endpoints in HTA and allow for more streamlined clinical trials.

Composite endpoint selection should be informed by their value to patients; currently the HTA guideline focuses on statistical aspects and less on patientneeds.

### 3.2.6. Surrogate endpoints

The EUnetHTA Endpoints Used for Relative Effectiveness Assessment: Surrogate Endpoints guideline (2015) outlines when it is appropriate to use surrogate endpoints for REA, based on the published literature, and incorporates insights from recent reimbursement decisions in its recommendations.

#### 3.2.6.1. *Scope & Purpose*

The overall objectives of the guideline are specifically described (i.e., provision of a list of validated endpoints, description of process for validation, use of surrogate markers for diagnostic or screening purposes, and detailed information on statistical methods relevant to surrogate end-points). Two of the topics explicitly excluded from the objectives (i.e., process for validation; and statistical methods) could reasonably be expected to fall within the remit of a guideline for surrogate endpoints. Critique of the guideline indicates that its overall objectives are addressed, albeit with limited detail.

The problem statement covered by the guideline is specifically described, when considering the scope of the guideline. The problem statements include questions on: (1) acceptability of results on surrogate endpoints for REA (addressed in Section 2); (2) which type of surrogate endpoints are used in an REA – (addressed in Summary paragraph 3 and Section 1.1.1; brief examples of surrogates are also provided in Section 2, page 11); and what is the role of absolute differences (e.g., two months' survival gain) and relative differences (e.g., hazard ratios).

The stated audience is clear: the HTA assessor is stated as the audience in the document. However, the perspective and context of their remit could be described more clearly. The guideline could be improved by a clear definition of the context related to its use, as it focuses on issues pertinent to suitability and choice of surrogate markers for RCT (e.g., there is little guidance on the handling of surrogate markers if they have been deemed sufficient for market authorisation but their patient relevance is uncertain).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

#### 3.2.6.2. *Comprehensiveness*

The guideline provides a comprehensive, but not specific, list of evidence requirements (Section 2), and focuses on one type of evidence only (surrogate markers for clinical endpoints). Within this restriction it allows use of both biomarkers and intermediate endpoints as surrogates.[1]

The guideline provides an overview of decision criteria for the suitability of surrogate markers for an REA, although the decision criteria are not quantified (e.g., a threshold for adequate surrogacy is not defined).

Given the scope of the guideline, it provides a comprehensive overview of methods applicable for assessing surrogate markers focusing on validity (Section 2). In terms of omitted requirements, alternative requirements and methods for surrogate markers may be required for therapies that provide long-term clinical benefit and/or the potential for cure. Missing topics in the guideline include recommendations on the process for validation along with statistical methods and clarity on the use of surrogate endpoints for technologies with different mechanisms of action. It should also be noted that recommendations on linked evidence and linking surrogate outcomes to final outcomes are insufficient.

#### 3.2.6.3. *Clarity and presentation*

The guideline supports the use of surrogate markers in certain defined situations (see stated objectives in section 3.2.6.1) and describes important considerations for their assessment. The recommendations may be considered specific and unambiguous, although guidance on quantitative thresholds are not provided. The guideline authors note this gap and state that "*The use of surrogate endpoints in the assessment of (added) clinical benefit of a health technology is controversial*", and "*… differences in [HTA] may be related to … different interpretation of data coming from the same primary studies*" [Page 14, paragraph 3]. Whilst methods for derivation and validation of surrogate markers are discussed clearly, details are lacking since they are explicitly beyond its scope.

#### 3.2.6.4. *Applicability*

The guideline addresses practical HTA decision-making given that surrogate markers, their validity and use is an important and frequently discussed topic in REAs. It acknowledges differences in HTA practice among EU countries, but only partially addresses them in the proposed methods. In the light of intra-European variation, it is advised that a surrogate, which is a close and necessary step in the development of a clinical

[1] Important definitions and considerations for biomarkers in the guideline are:

a. An intermediate endpoint is a clinical endpoint such as measure of a function or of a symptom but is not the ultimate endpoint of the disease. (Page 7, Section 3)
b. A biomarker is defined as a characteristic that is objectively (reliably and accurately) measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention. The biomarker must lie on the pathophysiologic causal pathway of the disease; it must be correlated with a clinical endpoint to be useful in detecting disease and assessing prognosis, and validated through a validation or qualification process. (Page 7, Section 2)
c. For the purpose of REA, both biomarkers and intermediate endpoints will be considered, if used, as surrogate endpoints to substitute for a clinically meaningful (final) endpoint. (Page 7, Final paragraph).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

outcome, should be preferred to earlier endpoints.[2] For example, acceptance of PFS for HTA in Europe varies between agencies and, accordingly, the guideline's recommendations on this topic are limited.

The guideline is not supported with tools for applications and, additionally, potential organisational barriers are mentioned but not fully addressed.

#### 3.2.6.5. *Appropriate methods*

The guideline provides high-level information, with limited detail on methodology as a detailed review of statistical methods is explicitly out of scope. The proposed approach is reflective of currently used methods by HTA agencies in Europe.

The guideline mentions specific qualitative and quantitative methodologies applicable to the choice and evaluation of surrogate markers, although detailed quantitative recommendations are not provided (e.g., on adequacy of surrogacy). Clarity on when, how and why surrogate markers may be sufficient for licensing approval but not for HTA would be beneficial for the target audience.

#### 3.2.6.6. *State-of-the-art methods*

The guideline provides high-level recommendations and explicitly does not aim to provide specific methodological detail (in section 1.3). Considering the guideline's scope, it provides state-of-the-art advice to some extent, albeit with important parts missing (e.g., statistical methods). In terms of research needs, the guideline clearly states that "*Further methodological research on the use of surrogate outcomes is needed to inform future REA approaches for the handling of surrogates*". Specifically, efforts to reconcile the needs of regulatory and HTA agencies would be valuable.

Research is needed on how to optimally assess the link between a surrogate marker and final endpoint after market entry, and assess the adequacy thresholds of surrogate markers. Prior to the establishment of thresholds, the accepted methodology to interpret and reproduce surrogacy evaluation should be agreed since the current meta-analytic approach is impractical and requires a high number of studies measuring both the surrogate and clinical outcome of interest. Further development of the guideline may also add value, including explicitly stating and providing examples and/or case studies of best practice.

#### 3.2.6.7. *Uncertainty characterisation*

Methodological uncertainty is a key consideration for surrogate markers and is thoroughly addressed in a section on the validation of surrogate markers.

The use of surrogate markers may limit the number of patients exposed to the treatment during trials, and therefore, increase the uncertainty of its safety (i.e., surrogate markers facilitate smaller trials and all else being equal, fewer patients would have be exposed to a medication approved based on a surrogate end-point than based on the corresponding final clinical endpoint). The guideline advocates availability of large

[2] In the context of the EUnetHTA guideline, they refer to a biomarker (CIN3 vs CIN1) rather than an intermediate clinical endpoints; hence, surrogate marker is likely the preferred term.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

safety databases for treatments that have shown effectiveness only on a surrogate endpoint (Summary section, paragraph 7).

The guideline does not appear biased in terms of assessment of surrogate markers and refrains from making specific quantifiable recommendations. However, there is little guidance on practical decision-making for assessors that have no input on the choice of surrogate marker.

#### 3.2.6.8. *Generalisability, transferability and feasibility*

The described methods are transferable across EU jurisdictions, although there may be organisational barriers to implementation. Given the guideline scope, HTA-related aspects not covered include organisational barriers to implementation, pragmatic clinical trials, and framework for validating surrogate endpoints post approval.

The recommendations are relatively general and can therefore be expected to promote uptake of the proposed methods. However, the high-level framework of the advice implies that important inter-country variation is likely to remain even if the guidelines are implemented.

Given the variation on the required level of validation among HTA bodies, generalisability for firm advice may require the most restrictive path to ensure that all requirements are fulfilled; this would negatively impact patient access to new technologies.

#### 3.2.6.9. *Conclusions*

Overall, the guideline aims toward consistency and predictability in the use of surrogate markers in well-defined situations, although with noted limitations. It also provides advice on implications from the use of surrogate markers for re-assessment and data collection after market entry. Because of the lack of specific quantifiable decision rules, variation in the acceptance of surrogate markers across HTA bodies in Europe will likely remain even if the guidelines are implemented.

The guideline would greatly benefit if the topics below are explicitly addressed in an upcoming version:

- Provide recommendations for the reporting of surrogate endpoints
- Clarify under what circumstances the use of a surrogate endpoint would be acceptable for the JCA and the relevant threshold for surrogacy
- State surrogate endpoint requirements for therapies that provide long-term clinical benefit and/or the potential for cure
- Provide advice on validation and thresholds to assume adequate surrogacy, and guidance on the use of validated surrogates by drug class, disease type and disease stage

#### 3.2.6.10. *Recommendations for Internal Position paper*

The guideline provides a thoughtful overview of issues important in selection and evaluation of surrogate markers for REA. Given its broad remit, specifications have to be made on a case-by-case basis. Therefore, the guideline should emphasise the importance of early communication and alignment among



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

manufacturers, regulatory agencies and HTA bodies to address variability in requirements among HTA bodies.

Guidance will be needed for a framework for the evaluation of surrogate endpoints within the HTA setting, bearing in mind that a prescriptive approach may decrease inter-jurisdiction variability (e.g., on the levels of evidence needed for surrogate markers in an HTA context). Recommendations on reporting surrogate markers are needed in light of limitations observed in the peer-reviewed literature (La Cour JL, *et al*. 2010); requirements for clinical trials reporting surrogate markers should be standardised (e.g., a clear rationale for the choice of the relevant surrogate should be provided along with references to biological plausibility and evidence of validation).

Detailed advice on the adequacy of surrogate markers, the validation process and statistical methods is needed. The validity of surrogates is missing, especially with a view on handling the current scientific perspective and criteria between two options: (1) first in class and curative therapies, and (2) therapies in an established disease area. Upon the introduction of a new technology (e.g., CAR-T), guidance should be provided on how to re-validate existing surrogate endpoints.

Adequacy of surrogate markers for therapies that provide long-term clinical benefit, including early disease markers and/or the potential for cure, are likely to become important in the coming years and may, therefore, merit specific discussion in the guideline.

In the light of apparent differences in the acceptance of surrogate markers among HTA bodies, provision of examples or case studies of good practice may be a good first step towards more detailed advice. Furthermore, ED/scientific advice with regulators and HTA bodies should be ongoing in order for agreement to be reached on the validation process and acceptability of uncertainty.

Overall, surrogate endpoints should be informed by their value to patients (e.g., how PFS is valued in terms of no tumour growth or the absence of metastasis); thus, the gap between patient relevance and surrogate endpoints needs to be bridged without necessarily focusing on COAs.

### 3.2.7. Internal validity of NRSs

The Internal validity of NRS on interventions guideline (2015) is intended to provide recommendations on the assessment of the internal validity of NRSs used for the evaluation of effects of interventions.

#### 3.2.7.1. *Scope & Purpose*

The overall objective of the guideline is well defined; the aim is to recommend tools or checklists that are suitable for assessing RoB in NRS evidence. The guideline aims to answer the key questions that define the guideline scope: “how to classify NRS evidence according to study design, and how to best assess RoB of specific NRS types?”. According to the guideline, NRSs are study designs that lack randomised allocation



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

of interventions. NRSs or observational studies can be grouped into cohort, case-control and non-comparative studies.

The guideline does not address generaliseability in terms of the external validity of NRS studies, and relevant instruments for assessing the external RoB were not included in the scope of this guideline. EUnetHTA has published a methodological guideline (EUnetHTA, 2015) to provide guidance on the applicability and external validity assessment of both RCTs and NRS in REAs.

Furthermore, the guideline does not consider "Quality of reporting" instruments. RoB tools that included sections on quality of reporting, such as CONSORT, PRISMA or STROBE, were excluded.

The guideline does not explicitly provide guidance on the validity of RWE or registry studies. EUnetHTA is currently developing a methodological tool in registry studies (the Registry Evaluation and Quality Standards Tool (REQueST); EUnetHTA 2019a), which is built upon the results of the PARENT JAs and aims to support HTA organisations in guiding and evaluating registries for effective usage in HTA. The tool and its Vision Paper are open for public consultation at the time of writing this report (July 2019).

The guideline is framed by the following assumption by EUnetHTA: "It was assumed that HTA reports are more likely to include NRS as the sole rather than an additional source of information on effectiveness and safety" (p. 9). This assumption cannot be seen as providing a realistic starting point in terms of evidence submitted to HTAs given that in many cases NRSs are complementary evidence presented as part of the full evidence package. Indeed, the problem statement of the guideline is that the inclusion of NRSs in a systematic review conducted as part of a HTA may be useful in specific circumstances, but leads to several challenges in terms of internal validity assessment.

#### 3.2.7.2. *Comprehensiveness*

The EUnetHTA guideline provides a comprehensive overview of currently available RoB tools selected based on pre-defined criteria and identified in literature reviews. The guideline does not, however, deep dive into the details of how these apply to different study designs and which RoB tools are the most commonly used by HTA agencies. Importantly, regulators are also developing frameworks for RWE programs (ElZarrad M.K. 2019).

Many of the challenges related to RWE are exacerbated by a lack of universally accepted standards or principles for the design, conduct, analysis and/or reporting of RWE. The guideline pertains to NRSs; however, there seems to be a demand for more precise methodological guidance on the internal validity of RWE studies submitted to HTA agencies. This need may be partly addressed by the development of REQueST (EUnetHTA 2019a). However, REQueST only covers quality (and the potential relevance) of data registries; it does not cover other types of RWD or RWE studies, the validity of which depends on other factors besides data quality.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

Further guidance on preference studies would be beneficial and in RWE-based evidence generation, guidance is needed on assessing the internal validity of innovative study designs such as nested trials within cohorts.

Selection bias is a key potential challenge for observational RWE, and there are techniques available to identify, describe and/or adjust to it (e.g., by adjusting for covariates, matching, or using instrumental variables). The definition of selection bias vs. confounding in subsection 2.1 is blurred and contradicts the distinction between these two sources of bias as correctly stated in section 1.1. Time-dependent confounding is not addressed in the guideline, although it plays an increasing role in causal inference from observational studies.

#### 3.2.7.3. *Clarity and presentation*

The EUnetHTA recommendations are clear and specific regarding recommended tools. The guideline also reviews other tools and explains why these are not recommended. The guideline should, however, provide: (a) clarity on the definition of biases and include up-to-date advances made in causal inference, and (b) examples of which study design may prove to be problematic to assess using the recommended ACROBAT / ROBINS-I tool.

#### 3.2.7.4. *Applicability*

Guidance on the role of NRSs as complementary evidence would be beneficial. In terms of practical decision-making, the assumption by EUnetHTA that NRSs are used as the sole source of evidence does not fully reflect reality. Companies may include evidence from NRSs as supportive and complementary evidence of a treatment's benefit and demonstrate external validity. Many sources of evidence are now used in combination and NRSs are sources of evidence not only in cases where RCT data are lacking.

In terms of organisational barriers, cost or time implications, these are not considered very relevant for the guideline as the RoB tools are already used by HTA agencies and present no major resource burden.

#### 3.2.7.5. *Appropriate methods*

The EUnetHTA guideline does not elaborate on which of the RoB tools are currently used or widely accepted by HTA agencies, but does provide an overview of currently available RoB tools selected based on pre-defined criteria and identified in literature reviews. It would be beneficial to have examples of different observational study designs and present case studies of RoB Tools used by the majority of HTA agencies, especially in RWE studies.

#### 3.2.7.6. *State-of-the-art methods*

The ROBINS-I (previously ACROBAT-NRSI - A Cochrane RoB Assessment Tool) developed by the NRS Methods Group of the Cochrane Collaboration, is recommended by EUnetHTA as the current best tool for assessing RoB of NRSs on interventions. The preference for the ROBINS-I / ACROBAT-NRSI tool mainly stems from its advantages, such as the availability of more detailed instructions and documentation guides,



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

compared with other tools (e.g., RoB Assessment tool for NRS, RoBANS). In practice, there is no consensus between HTA groups on the preferred NRS appraisal tool.

The guideline notes that although many quality assessment tools exist and have been used for appraising NRSs, most omit key quality domains (Deeks *et al*., 2003).

EUnetHTA notes that the ISPOR Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform Healthcare Decision Making covers all key domains of bias, but the questionnaire contains several items on the quality of reporting, such as reporting both absolute and relative effect measures. Therefore, this RoB tool is not recommended for general use. This recommendation is considered unclear and merits further explanation.

In addition to those RoB tools reviewed by EUnetHTA, a Checklist for Retrospective database studies in the form of 27 questions is available (Motheral *et al*., 2003).

Future research would need to focus on gaining agreement on the definitions of NRS designs to allow for the use of an appropriate RoB tool; the role of RoB tools in re-assessments also needs to be clarified.

Additionally, the role of retrospective observational studies has been under-emphasised and published work from the ISPOR Task Force on retrospective database analysis should be considered in the next version of the guideline (Berger *et al*., 2009; Cox *et al*., 2009; Johson *et al*., 2009).

#### 3.2.7.7. *Uncertainty characterisation*

The EUnetHTA guideline proposes key domains of RoB which are addressed by the recommended ACROBAT / ROBINS-I tool. All in all, five types of bias are addressed by the recommended tool: selection bias (including bias due to confounding), performance bias, detection bias, attrition bias and reporting bias. Many of the existing tools for RoB assessment fail to address all relevant domains or they confuse the assessment of internal validity with that of the quality of reporting, according to EUnetHTA. Definitions of, and methods to address the different biases, were not fully or correctlyelucidated.

An area of uncertainty is correct study design classification, for example, mislabelling studies. Moreover, the EUnetHTA guideline mentioned that including NRSs on top of RCTs in the REA would result in increased uncertainty.

#### 3.2.7.8. *Generalisability, transferability and feasibility*

The recommended RoB tools seem to be transferable across EU jurisdictions as they are not country-specific. As noted earlier, there should be more detailed guidance on RWE evidence bias assessment and innovative study designs which may include flexible sample sizes, enrichment or the use of historical information, multiple-arm and multiple-domain trials and novel endpoints.

#### 3.2.7.9. *Conclusions*

The main methodological gaps identified in the guideline relate to the assessment of RWE, uncertainty and acceptability of NRSs from the HTA assessor's perspective. In RWE-based evidence generation, guidance



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

is needed on assessing the internal validity of innovative study designs. An area of uncertainty and need for further guidance is correct study design classification. Furthermore, clarifying under what circumstances NRSs are acceptable, and identifying study subjects and acceptable outcomes and analytical methods would be beneficial.

Overall, the Internal Validity of NRS guideline provides clear recommendations for bias assessment of NRSs. However, it does not give detailed guidance on the different study designs that may be challenging to review using the recommended ROBINS-I tool, nor does it provide recommendations for how to deal with the RoB assessment results. Moreover, the guideline does not elaborate on the use of single-arm trials for ethical purposes, conditional approval for targeted therapies or stronger markers of clinical benefit.

NRSs can be submitted as complementary evidence of treatment benefit, although EUnetHTA assumes NRSs to typically be the sole source of evidence in an HTA submission. This assumption may not fully reflect reality.

In terms of predictability, the guideline does promote the ROBINS-I tool in a clear manner. The guideline also provides clear reasoning for not recommending other RoB tools used by HTA agencies.

#### 3.2.7.10. *Recommendations for Internal Position paper*

A key recommendation for the future improvement of the internal validity of NRS guideline is for it to provide more detailed guidance on the validity and acceptability of RWE studies, single-arm trials and cross-sectional studies. The validity and feasibility of the ROBINS-I tool should also be established.

Furthermore, it is recommended that agreement be attained on the definitions of NRS designs to allow for the use of an appropriate RoB tool.

The scope of the guideline should also reconsider the role of NRSs as a source of complementary evidence in support of an HTA. Currently, the EUnetHTA assumes that HTA reports are more likely to include NRSs as the sole, rather than an additional, source of information on effectiveness and safety and only in cases where higher-level evidence is not available.

Clearer guidance is needed on understanding under what circumstances NRS evidence would be acceptable, the most appropriate methods for identifying study subjects as well as acceptable outcomes and analytical methods.

Single-arm trials, a very important source of information for many orphan drugs and targeted therapies with high ORR effects, should be considered. With regard to historical comparisons in cases where only single-arm trials are available, it should be made more explicit how internal validity is assessed in single-arm trials and cross-sectional studies.

Pragmatic trials should be considered as RCTs and are therefore not included in the scope of the NRSs. However, pragmatic trials may have post-randomisation confounding and selection bias, which are not



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

sufficiently addressed in this guideline. Recent literature should be considered for adoption (Hernán *et al*., 2017; Latimer *et al*., 2016).

The guideline does not provide recommendations for how to deal with observational evidence or the use of appropriate methods to enhance the role of observational evidence (e.g., G-methods for time-dependent confounding).

Finally, a broader consideration of aspects related to external validity are important to include in the guideline, while recognising that this may have been out of scope for the original guideline. A holistic consideration of the external and internal validity of the whole evidence base is preferred and recommended.

### 3.2.8. Internal validity of RCTs

The Internal validity of RCTs (2015) guideline focuses on the assessment of the RoB of RCTs, the most relevant trials for REAs. A separate EUnetHTA guideline deals with the problem of assessing applicability and external validity.

#### 3.2.8.1. *Scope & Purpose*

The overall objective of the guideline is well defined; the aim is to provide recommendations for the assessment of the internal validity of RCTs. The problem statement is presented clearly: “To what extent can it be assessed whether the data from a study (e.g., an RCT) or a collection of studies (e.g., a meta-analysis within an REA) are likely to reflect the “truth” by considering methodological quality criteria?”. It was essential to understand to what extent data can be assessed to reflect the RoB to allow conclusions to be made about the certainty of results for the decision-making process.

The guideline does not address an important type of bias (i.e.,post-randomisation confounding) since there was no published literature on the topic when the guideline was published. This is relevant when interpreting estimands in RCTs with treatment switching and pragmatic trials controlling for compliance (Hernán *et al*., 2017).

To meet the overall objective of providing recommendations for the internal validity assessment, the authors explored the likely biases that may be found in SLRs/NMAs, and considered possible implications (e.g., blinding).

#### 3.2.8.2. *Comprehensiveness*

The guideline provides examples RoB tools used in clinical trials with a focus on the Cochrane tool for the assessment of bias. No extensive literature search was conducted by EUnetHTA. The guideline describes the Cochrane RoB tool, as it is the generally accepted standard and used by several HTA agencies active in EUnetHTA. Furthermore, the guideline refers the HTA assessors to the Cochrane Handbook for details and support for judgment (chapter 8, in particular chapter 8.5 and table 8.5.d).



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The guideline proposes a standardised RoB tool; however, it is unclear whether this should be used in parallel to the Cochrane framework, or whether this has been developed because the Cochrane Collaboration has not provided such a sheet. The guideline also summarises RoB tools for SLRs if primary data are not collected. It should be noted that guidance on RCTs with treatment switching at progression was missing, as was guidance onissues around and the potential use of pragmatic trials.

#### 3.2.8.3. *Clarity and presentation*

The EUnetHTA recommendations for assessing the internal validity in RCTs are clear and specific in terms of which tool is recommended. It can be assumed that the recommendation will be updated to the newer version of the Cochrane handbook, which is expected to be published in 2019. The guideline does not review any other RoB tools, but suggests HTA agencies should not use the Jadad scale.

The guideline gives four strategies how to deal with RoB, although these are still open to interpretation and could be considered non-specific. Examples are missing in terms of how the results should be presented. This could be addressed by presenting a summary table along with text explaining the assessment.

#### 3.2.8.4. *Applicability*

EUnetHTA assumes that REAs are mainly performed on the basis of primary studies, but acknowledges that if a systematic review is the basis, the underlying methodology should be assessed in alignment with guidance provided on comparator selection and ITCs.

The guideline sees the Cochrane tool as the gold standard for RoB assessment for RCTs. It does mention a difference with the IQWiG guideline, but does not specify any other assessment bodies or differences across the EU.

In terms of applicability of the guideline, an overview of other tools that are available for RoB assessments and the reasoning for not recommending these for REAs, wsd missing.

Consideration of the organisational barriers, cost or time implications are not considered relevant as the Cochrane RoB tool is already used by many HTA agencies and present no major resource burden.

#### 3.2.8.5. *Appropriate methods*

The EUnetHTA guideline states the Cochrane tool to be “generally accepted standard or gold standard, as it is used by a number of HTA agencies active in EUnetHTA” (p. 5). However, it was not compared with other RoB tools (other published checklists or even HTA tools such as the NICE abbreviated tool).

While the recommendation that the Cochrane tool should be used may be appropriate, it has not been assessed or confirmed. This was an assumption at the start of the guideline and thus making it a conclusion may not reflect the view of HTA assessors in practice.

The guideline focusses on qualitative approaches and states that scales should not be used to assess internal validity. HTA agencies that currently make use of the Jadad scale should change the scale,



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

according to EUnetHTA. Furthermore, guidance on the number of assessors for scoring purposes was missing.

In the appendix, a proposed standard assessment template is provided to be used to assess the RoB by HTA agencies. However, it is not clear whether this is done in parallel with the Cochrane RoB assessment, or in addition to it.

#### 3.2.8.6. *State-of-the-art methods*

It is stated in the guideline that, "The Cochrane RoB tool can be regarded as state of the art" (p. 17.) while EUnetHTA acknowledges that other instruments exist. Indeed, other tools are needed to assess the quality of systematic reviews (Shea BJ, *et a*l., 2007). Two are formally validated: the Oxman and Guyatt index and the AMSTAR instrument. EUnetHTA does not provide justification as to why the Cochrane tool should continue to be the gold standard.

As noted in section 3.2.8.1, the guideline needs to be updated with recent methodological developments (e.g., description of causal inference methods (G-methods) for different estimands of RCTs).

#### 3.2.8.7. *Uncertainty characterisation*

The guideline proposes four strategies on how to deal with studies with a high or unclear RoB and recommends HTA assessors to specify in advance which strategy to use within each REA.

Uncertainty was identified around statistical pooling, as HTA agencies allow different p-values. EUnetHTA recommends specifying in advance in the protocol for a HTA how to address heterogeneity. No clear guidance is provided on how to deal with different perceptions or approaches across the EU.

Regarding the assessment for systematic review, EUnetHTA mentions that no threshold is available in the overall assessment of the AMSTAR instrument, thus it should be defined beforehand. However, no guidance is provided for how HTA assessors should set the threshold, and therefore the outcome will stay uncertain.

#### 3.2.8.8. *Generalisability, transferability and feasibility*

The recommended RoB tools seem to be transferable across EU jurisdictions as they are not country-specific. The guideline provides RoB tool recommendations for submissions based on both the primary studies and systematic reviews. Further guidance on when bias should be considered, "low", "high" or "unclear" would minimise subjectivity. The proposed standardised RoB assessment does ask clear questions with explanations for the definitions (Annex 2, p. 20).

Few details are provided on how the Oxman and Guyatt index and the AMSTAR instrument tools should be used, as this was beyond the scope of this guideline.

#### 3.2.8.9. *Conclusions*

The main methodological gaps identified in the Internal Validity of RCTs guideline pertain to four areas: comparative evidence on the appropriateness of the Cochrane RoB tool over other existing tools; guidance



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

on how the Oxman and Guyatt index, and the AMSTAR instrument tools should be used; guidance on when bias is seen as “low”, “high” or “unclear” to avoid subjectivity; and guidance on the internal validity of RCTs with treatment switching at progression.

While the guideline provides clear definitions for the types of bias in RCT studies and the risk groups, it does not provide clear reasoning for not recommending other RoB tools that are available for assessing RCTs. Moreover, it does not provide clear guidance on assigning a level (low, high or unclear risk) for each type of bias and how to avoid such biases using appropriate analytical methods.

In terms of predictability, the guideline promotes the Cochrane Handbook in a clear manner, which can be interpreted as promoting predictability. The guideline promotes two tools for systematic reviews; however, only a high-level explanation of the tools is provided.

EUnetHTA states that is necessary to have appropriate training (e.g., in causal inference and estimand use) and clear, consistent decision rules to ensure acceptable reproducibility. However, no recommendations/initiatives are made regarding who should have responsibility and ownership of these quality measures.

#### 3.2.8.10. *Recommendations for Internal Position paper*

The key recommendations addressing the identified gaps in methodology are:

- To ensure predictability and consistency, the proposed standardised RoB assessment should become mandatory for all HTA agencies to use
- Any updated guidance should consider the new version of the Cochrane Handbook (v.6) which will be published later in 2019

To promote consistency in assessments, it is critical that the proposed standardised RoB assessment become mandatory for all EU HTA agencies, with an expectation for EU HTA bodies to use the same tool and have the same rating systems for RoB.

Based on the critical review of the guideline, it would be beneficial for HTA assessors to specify in advance how to deal with studies which have a high or unclear RoB. Furthermore, clear guidance should be in place to allow uniform assessments in line with the warning against multiple rating scales (e.g., Jadad scale). It would also be beneficial to have examples of different ways to deal with bias across Europe.

Future research may be needed to further understand differences in internal validity outside of RCTs and to ensure alignment across HTA bodies.

### 3.2.9. Safety

The EUnetHTA Endpoints used in Relative Effectiveness Assessment: Safety guideline (2015) provides a framework for the evaluation of relative safety performed by HTA assessors when conducting REA of pharmaceuticals.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

#### 3.2.9.1. *Scope & Purpose*

The aim of the guideline, as stated in the "Summary section" is to provide a "framework for the evaluation of relative safety performed by HTA assessors in the context of REA". The described scope focuses on pharmaceutical interventions and medical devices and covers six key issues: objectives of HTA assessors; terminology; identification of adverse reactions; sources of information; evaluation of sources of information; and synthesis and report of results compared to other interventions.

Each of these topics is addressed; however, there is limited guidance on how to synthesise and compare data on safety endpoints. Furthermore, the introduction does not state whether only AEs that are at least possibly related to the study drug should be assessed or whether all treatment-emergent AEs are to be considered. This is a tremendous difference that needs to be addressed.

The problem statement, while broadly described in the context of the importance of weighing both the benefits and harms of alternative interventions, is not explicitly stated [Section 1.2]. Nevertheless, the scope of the guideline is appropriate for the stated audience (HTA assessors) and assumed problem statement.

#### 3.2.9.2. *Comprehensiveness*

The guideline provides a detailed discussion of the pros and cons of a spectrum of potential data sources (e.g., regulatory, RCTs, observational studies, and case reports) to consider for identifying adverse effects relevant to the REA. This includes information on assessing RoB and study quality when considering data sources (Recommendation 5). Using existing checklists from the Cochrane Collaboration, STROBE, etc. was recommended, but there was no specific criteria for inclusion or exclusion (i.e., no maximum required RoB score). However, given these decisions are likely to be disease- and/or treatment-specific, this information is arguably be out of scope for a general guidance.

It is not clear why safety is not recommended to be assessed in the same way as efficacy endpoints. Relative effect measures such as OR, RR and (if appropriate) HR are mentioned, but not to the same extent as in the guideline on endpoints. It remains unclear, why efficacy endpoints are assessed only when an appropriate evidence level is reached while safety data can be extracted from any source.

Specific recommendations on how, when and what information to summarize from individual studies are provided (Sections 2.5.1 and 2.5.2 and also in the EUnetHTA submission templates) but there is limited guidance on how to synthesise this information (quantitatively or qualitatively). There is also no recommendation for how frequently safety assessments should be conducted.

It is recommended that HTA assessors should "describe whether there is a clinically significant difference in adverse reactions between the interventions" but there is no guidance on how to evaluate clinical significance. A clear statement on the risk-benefit assessment is needed if safety is regarded in the REA.

#### 3.2.9.3. *Clarity and presentation*

The key recommendations (p. 6-8) are generally clearly laid out, specific and unambiguous. The exception would be Recommendation 9, which broadly recommends conducting a balanced assessment of the



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

relative safety and relative benefits of the intervention. No actual guidance is provided on this point, although this is a much larger question than the stated scope of the guideline. On the whole, recommendations are clearly presented and high-level. However, there could be a clearer separation between initial and repeat safety assessments.

Section 2.5.4 states that "some frameworks reporting both benefits and adverse effects were proposed" with reference to citation 53, which is missing from the bibliography.

#### 3.2.9.4. *Applicability*

The guideline for the most part reflects practical HTA decision-making. Recommendation 3 states that HTA assessors use precise terminology from the MedDRA dictionary to describe adverse reactions, however HTA assessors do not collect the data or play a role in describing data in the studies they evaluate.

The guideline references specific tools and checklists to apply in the evaluation of relative safety assessment (e.g., STROBE, Cochrane Collaboration, NICE guidance). There is no specific mention of differences across EU practice, but a summary of the key methods for reporting safety data by institution is included (Annexe 3, Table 6). Potential organisational barriers are also not explicitly discussed.

It is stated that the reporting of safety is variable even in RCTs. However, given variability is higher in publications than in study reports, variability should be differentiated by source of reporting rather than by study type.

#### 3.2.9.5. *Appropriate methods*

The developers conducted a literature search of existing guidelines used by HTA agencies and summarised the key methods for reporting safety data by institution Annexe 3, Table 6). While the proposed approach generally aligns with methods used by HTA agencies in the EU, showing summary tables of adverse reactions is not sufficient to assess safety.

Relative effect measures should be reported for treatment emergent AEs, although it is unclear how relativity is assessed (i.e., whether the subjective investigator assessment is sufficient and appropriate). Specific recommendations for how and what information to summarise from individual studies is provided but there is limited guidance on qualitative and quantitative synthesis, how to determine "clinical relevance" of adverse reactions, and how to weight the relative safety against the relative benefits. It could also be beneficial to suggest comparing the number needed to treat with the number needed to harm for both trial and real-life patient populations. Overall, the guideline under-reports the importance of deriving absolute measures (e.g., risk differences) from safety studies, as benefits and harms must be weighted against each other (trade-off assessment).

#### 3.2.9.6. *State-of-the-art methods*

The guideline covers main methods and available guidance for accurately collecting and presenting AEs. All necessary elements to be included when synthesising data from various sources is provided (e.g., use



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

of claims data should be undertaken in a systematic way). However, it is stipulated that the collection and assessment (or re-assessment) of safety data are within the remit of regulators rather than HTA bodies.

Clinically significant difference in adverse reactions between interventions is not meaningful alone; it must be seen in the context of a benefit comparison (i.e., the guideline does not provide sufficient advice on how to integrate benefits and harms).

Literature published after the development of this guideline includes causal inference methods (i.e., pharmacoepidemiologic studies must focus on the causal effect of interventions on side effects). Thus, historical results and any pitfalls should be mentioned in the next version of the guideline (e.g., Murray EJ and Hernán MA., 2018).

Although no immediate future research is required, it would be pertinent to consider including search strategies that could capture safety data within a systematic process. Finally, the role of using claims databases in a systematic way should be part of the state-of-the-art methods.

#### 3.2.9.7. *Uncertainty characterisation*

The guideline adequately describes the limitations and caution to be undertaken when comparing safety data from various sources. Inconsistent definition of safety terms, and differences in study designs and populations are well described as sources of uncertainty in AE rates and comparison. Further uncertainty may arise from under-reporting in pharmacovigilance databases. Moreover, reporting of AE severity may invite uncertainty as patients and physicians do not perceive AEs in the same way. Additional discussion should focus on uncertainty arising from confounding and time-dependent confounding (see also section 3.2.9.6 on causal inference).

#### 3.2.9.8. *Generalisability, transferability and feasibility*

The guideline supports different types of evidence submission including pharmaceutical drugs and medical devices. The methods and evidence provided are very comprehensive and cover multiple HTAs and geographic jurisdictions. However, given the lack of an EU joint process for HTA, there is little clarity on how safety endpoints should be assessed in relation to current treatments. It is also not specified whether reports on PT and SoC level should be available for all AEs, or just those of special interest. The guideline should also summarise information across the EU on differences (e.g., legal) among countries that prescribe how HTA bodies should use safety data.

#### 3.2.9.9. *Conclusions*

The guideline indicates the consistency of terminology of safety data reporting across the EU within the regulatory space but recognises that safety data assessment is not within the purview of an HTA. A summary of potential sources of safety data is presented, but the guideline does not specify that safety assessment (re-assessment) is within the remit of regulatory as opposed to HTA assessors. Furthermore, in comparative safety assessments all interventions (pharmaceutical, medical devices and non-drug therapies) should be considered.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

The approach presented in this guideline is not sufficient to assess the safety as counterpart to the efficacy for a consistent risk-benefit assessment in a REA. The guideline provides a summary of how safety data may be reported but does not state how the heterogeneity of data can be addressed when considering comparative safety data.

#### 3.2.9.10. *Recommendations for Internal Position paper*

Whilst the guidelines lay out the sources, consistency, limitations of sources and reporting of AEs, additional considerations could include:

- Hierarchy of safety evidence
- Systematic process and search strategies to identify all relevant safety data
- Process or methodology of the comparison of safety data between technology and comparators, in particular where observational or single-arm studies are being utilized
- Consistency in safety reporting (i.e., who determines the severity of the AE [patient or physician])

Additionally, guidance on how to account for different follow-up periods and methods to quantify AEs should be considered in an updated version of the document.

Looking into the future and with the increasing use of e-Health applications, guidance would be needed on how to assess safety with connected devices alongside medications, and how to use tools through which patients directly report the severity of their experienced AEs.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# 4. Discussion

In order to establish a harmonised HTA environment in Europe, variation in HTA practices among MS and how this impacts decision-making for new medicines needs to mapped. Several publications address these from different angles (Allen et al. 2017; Kristensen FB 2017); however, the fundamental challenges related to objective assessment and effective collaboration among MSs can only be addressed by a closer look at the accepted and used methods across jurisdictions. The outcomes of this study focus on improving the understanding of the EFPIA HTA WG on methodologies to be considered when approaching JCAs in Europe and the potential for consensus among stakeholders. Due to the multiple perspectives on methodological standards, this is the first step toward assessing the clinical HTA landscape and aligning on scientific standards for inputs (e.g., study design and selection of endpoints, patient populations) and analytical quality with state-of-the-art methods.

The online survey results indicated that manufacturers consistently experience challenges with streamlined evidence generation, and thus seek HTA scientific advice for the selection of patient population(s), endpoints and comparators. The use of additional evidence (e.g., COAs and RWE) to support clinical trial data was not as frequently recurring, but is a reported concern for survey respondents.

The proposed recommendations from manufacturers target primarily the consultation process, and include suggestions for face-to-face meetings with HTA bodies in order to reach early consensus on the evidence development planning.

A limitation of the survey is the potential for recall bias. Feedback on ED conducted in the past, particularly from team-members who may no longer work for the responding company, may not be sufficiently captured (i.e., gaps in organisational memory). Additionally one representative was appointed by each company to collect and summarize the organisation-wide ED/scientific advice activity across several TAs, assets and jurisdictions over the period of one month. Because of this restriction, it is likely that important insights and information may have not been properly communicated, especially from country affiliates/local operating companies.

The EUnetHTA guidelines were developed under the remit of the EUnetHTA JA1, and the selected documents for critical review were published in 2015. Since their publication year, new treatment modalities, statistical methods and evidence development approaches have been introduced which are not thoroughly accounted for in the documents. Improvements in guidelines should focus on the effective incorporation of relative effectiveness data and ensuring the availability of these data throughout the product's life cycle. Uncertainty at the time of launch as to the product's REA could be addressed by consideration of post-marketing data – a key area for JA to focus on.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

Certain guidelines provide detailed guidance on the respective methodologies. However, many of the recommendations remain too high level to be applied in practical decision-making. It is acknowledged that there are inevitable divergences in certain areas of methodology (e.g., choice of comparator), but alignment between the regulatory and HTA requirements for clinical development plans and alignment on accepted levels of evidence are necessary. For example, in some cases, the choice of comparators must satisfy different expectations from the national regulatory agencies or EMA and HTA agencies. There is potential for ED to improve the alignment in procedures such as the Parallel Consultation, which aims to provide consolidated feedback from both EMA and HTA bodies on the proposed trial design.

Overall EUnetHTA has summarised current practices at the HTA agencies' level but not developed state-of-the-art methods. This is to be expected, as methodology development may be outside the remit and resources of the majority of HTA bodies in EU; there is no clear system forcing them to evolve from the current national HTA frameworks that satisfy national decision-makers. The EU Regulation is in the right direction to set a clear framework that would force the development of a centralised European process for clinical HTA.

To ensure that the guidelines include internationally recognised good practice (state-of-the-art) methods and are suitable for new innovative technologies, the future update of the guidelines must address potential issues that manufacturers may face in meeting the EC's centralised process including SoC selection, long-term extrapolation from key pivotal studies, variability on surrogate endpoints assessment, and acceptance of RWE. Furthermore, it is considered of high importance that ED is conducted in a transparent and collaborative manner to provide consolidated guidance for manufacturers in regard to clinical trial development, with minimal divergence between the regulatory advice versus HTA advice (e.g., in comparator and endpoint selection).

The EFPIA HTA WG is expected to disseminate a deep dive critique of existing EUnetHTA guidelines along with recommendations for their future content development via different channels (e.g., publications, issue panels in major conferences), in order to inform the health policy debate across Europe along these dimensions:

- Advocate for state-of-the-art "clinical" HTA, balancing the concepts of scientific rigour and flexibility as well as predictability versus evolution of science
- Account for the levels of uncertainty that may be faced by HTA agencies and payers
- Support timely patient access and national decision-making based on the best available evidence



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# References


Ahmad Y, Nijjer S, Cook C, El-Harasis M, Graby J, Petraco R, et al. A new method of applying randomised control study data to the individual patient: A novel quantitative patient-centred approach to interpreting composite end points. *Int J Cardiol*. 2015; 195:216-224.

Allen N, Liberti L, Walker SR, Salek S. A Comparison of Reimbursement Recommendations by European HTA Agencies: Is There Opportunity for Further Alignment? *Front Pharmaco*l. 2017;30(8):384

Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *The European journal of health economics : HEPAC : health economics in prevention and care.* 2018; 19(1):123-52

Anker SD, Schroeder S, Atar D, Bax JJ, Ceconi C, Cowie MR, et al. Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. *European journal of heart failure*. 2016;18(5):482-9

Armstrong PW, Westerhout CM. Composite End Points in Clinical Research: A Time for Reappraisal. *Circulation*. 2017;135(23):2299-307

Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. *Value Health*. 2009; 12(8): 1044-52

Charles River Associates. A comparative analysis of the role and impact of Health Technology Assessment: 2013. Final report. Available from: https://www.efpia.eu/media/25706/a-comparative-analysis-of-the-role-and-impact-of-health-technology-assessment-2013.pdf [Accessed January 3 2019]

Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report--Part II.

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health technology assessment* (Winchester, England). 2003;7(27):iii-x, 1-173.

Drummond MF, Schwartz JS, Jönsson B, et al. *International Journal of Technology Assessment in Health Care.* 2008, 24 (3): 244–258

EFPIA. Position on proposal for a Regulation of the European Parliament and the Council on health technology assessment and amending Directive 2011/24/EU. Available from: https://www.efpia.eu/media/361850/efpia-position-paper-on-the-commission-proposal-for-a-regulation-on-hta.pdf [Accessed January 3 2019]

El Zarrad MK. 2019. Framework for FDA's Real-World Evidence Program. Available from: https://www.actiac.org/system/files/ElZarrad%20MK%20PPT%20to%20ACT-IAC%20-FIN.pdf [Accessed January 3 2019]

EUnetHTA 2015. Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Endpoints used for Relative Effectiveness Assessment: Composite Endpoints v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Endpoints used for Relative Effectiveness Assessment: Surrogate Endpoints v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]




© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

EUnetHTA 2015. Endpoints used for Relative Effectiveness Assessment: Safety v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Endpoints used for Relative Effectiveness Assessment Health: related quality of life and utility measures. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Comparators & Comparisons: Criteria for the choice of the most appropriate comparator(s) v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Comparators & Comparisons: Comparators & Comparisons: Direct and indirect comparisons v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Internal validity of randomised controlled trials v2.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2015. Internal validity of non-randomised studies (NRS) on interventions v1.0. Available from: https://www.eunethta.eu/methodology-guidelines/ [Accessed January 3 2019]

EUnetHTA 2019. An Analysis of HTA and Reimbursement Procedures in EUnetHTA Partner Countries: Final Report2018. Available from: https://www.eunethta.eu/wp-content/uploads/2018/02/WP7-Activity-1-Report.pdf [Accessed January 3 2019]

EUnetHTA 2019a. Public Consultation for the Registry Evaluation and Quality Standards Tool, REQueST. Available from: https://www.eunethta.eu/request-tool-and-its-vision-paper-are-now-available-for-public-consultation/ [Accessed January 3 2019]

EUnetHTA 2015. Levels of Evidence – Applicability of evidence for the context of a relative effectivenss assessment Amended JA1 Guideline. Available from: https://www.eunethta.eu/levels-of-evidence-applicability-of-evidence-for-the-context-of-a-relative-effectiveness-assessment-amended-ja1-guideline-final-nov-2015/ [Accessed January 3 2019]

EUnetHTA HTA Adaptation Toolkit & Glossary. Work Package 5: Adapting existing HTAs from one country into other settings. https://www.eunethta.eu/wp-content/uploads/2011/01/EUnetHTA_adptation_toolkit_2011_version_5.pdf [Accessed January 3 2019]

Faria R, Hernandez Alava M, Manca A, Wailoo AJ. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data.2015. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/TSD17-DSU-Observational-data-FINAL.pdf [Accessed January 3 2019]

FDA. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims: Guidance for Industry. 2009. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims [Accessed January 3 2019]

Gsteiger S, Windisch R, Bryden P, Wiksten A. Non-Proportional Hazards in Network Meta-Analysis: Efficient Strategies for Model Building and Analysis. *Value in Health*. 2017; 20(9): A402

Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med*. 2017; 377(14):1391-1398

Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016; 183(8):758-64

iCOM. Health Technology Assessment in the European Union: State of Art and Future Scenarios 2017. Available from: http://www.astrid-online.it/static/upload/7787/7787e169a7f0afc63221153a6636c63f.pdf [Accessed January 3 2019]



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

ISPOR. Good Practices for Outcomes Research. https://www.ispor.org/heor-resources/good-practices-for-outcomes-research [Accessed January 3 2019]

Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health*. 2009; 12(8):1062-73

Kleijnen S, Lipska I, Leonardo Alves T, et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. *Annals of Oncology*. 2016; 27, 1768-1775

Kleijnen S, Toenders W, de Groot F, et al. European collaboration on relative effectiveness assessments: What is needed to be successful? *Health Policy*. 2015; 119:569-576

Kristensen FB. Mapping of HTA methodologies in EU and Norway 2017. Available from: https://ec.europa.eu/health/sites/health/files/technology_assessment/docs/2018_mapping_methodologies_en.pdf [Accessed January 3 2019]

Kristensen FB, Husereau D, Huic M, Drummond M, Berger ML, Bond K, et al. Identifying the Need for Good Practices in Health Technology Assessment: Summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA. *Value in Health*. 2019; 22(1):13-20

Kuehne F, Jahn B, Conrads-Frank A, Bundo M, Arvandi M, Endel F, Popper N, Endel G, Urach C, Gyimesi M, Murray E, Danaei G, GazianoTA, Pandya A, Siebert U. Guidance for a causal comparative effectiveness analysis using “big real world” evidence: when to start statin treatment. *J Comp Eff Res* 2019 [in press].

La Cour Jeppe Lercher, Brok Jesper, and Gøtzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. *BMJ*. 2010;341:c3653

Latimer NR, Henshall C, Siebert U, Bell H. Treatment Switching: Statistical and decision-making challenges and approaches. *Int J Technol Assess Health Care*. 2016; 32(3):160-6

Leverkus F, Chuang-Stein C. Implementation of AMNOG: An industry perspective. *Biometrical journal Biometrische Zeitschrift*. 2016; 58(1):76-88

Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. *Am J Epidemiol.* 2019; pii: kwz100. doi: 10.1093/aje/kwz100

Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, et al. A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. *Value in Health*. 2003;6(2):90-7

Mrowietz U, Kragballe K, Reich K, Spuls P, Griffiths CE, Nast A, et al. Definition of treatment goals for moderate to severe psoriasis: a European consensus. Archives of dermatological research. 2011; 303(1):1-10

Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health and quality of life outcomes*. 2015; 13:167

Murray EJ and Hernán MA. Improved adherence adjustment in the Coronary Drug Project. *Trials*. 2018; 19:158

NICE. Erenumab for preventing migraine [ID1188] 2019. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10302/documents [Accessed January 3 2019]

Pavlovic M, Teljeur C, Wieseler B, Klemp M, Cleemput I, Neyt M. Endpoints for relative effectiveness assessment (REA) of pharmaceuticals. *International journal of technology assessment in health care*. 2014; 30(5):508-13



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2018;38(2):200-11

Schmoor C, Schumacher M, Finke J, Beyersmann J. Competing risks and multistate models. *Clinical Cancer Research*. 2013; 19.1:12-21

Schnell-Inderst P, Iglesias CP, Arvandi M, Ciani O, Matteucci Gothe R, Peters J, Blom AW, Taylor RS, Sculpher M, Siebert U. A bias-adjusted evidence synthesis of RCT and observational data: the case of total hip replacement. *Health Economics* 2017; 26(Suppl. 1):46–69

Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: I. Guidelines for guidelines. *Health Research Policy and Systems.* 2006; 4:13

Scott M, Alsop J. Estimands—What Do They Mean for Health Technology Assessment? 2018. Available from: https://www.ispor.org/docs/default-source/publications/value-outcomes-spotlight/september-october-2018/ispor-vos-october-2018-heor-article-martin-scott.pdf?sfvrsn=656ca2b6_2. [Accessed January 3 2019]

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*. 2007; 7:10.

Siderov J, Tiu AL. Variability of measurements of visual acuity in a large eye clinic. *Acta ophthalmologica Scandinavica*. 1999; 77(6):673-6

Tafuri G, Pagnini M, Moseley J, Massari M, Petavy F, Behring A, et al. How aligned are the perspectives of EU regulators and HTA bodies? A comparative analysis of regulatory-HTA parallel scientific advice. *British Journal of Clinical Pharmacology*. 2016; 82(4):965-73

The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care.* 2003; 12:18–23

TLV. Aimovig (erenumab) 2018. Available from: https://tlv.se/download/18.2c1b1aa6167c5e758eb293ec/1545400450432/bes181218_underlag_aimovig.pdf [Accessed January 3 2019]

Unkel S, et al. On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical statistics*. 2018; 18:166-183.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# Appendix A: Online survey questionnaire

## EFPIA Members to Complete Online Survey Questionnaire

**[EMAIL INVITATION TEXT]**

**Research Project Title**: Manufacturers' perspectives focussing on methodological aspects of HTA scientific advice / ED in Europe, that relate to relative effectiveness assessment

**Background**: The European Federation of Pharmaceutical Industries and Associations (EFPIA) has sponsored a project that aims to critique the existing methods guidance for HTA, with a focus on EUnetHTA methods and early scientific advice. This will enable the Health Technology Assessment Working Group (HTA WG) to engage in the political debate and examine state-of-the-art HTA methodology in the context of the proposed EC's Regulation on JCA in HTA.
We are contacting you because you have been nominated by your company, based on your expertise in this area, and we would greatly appreciate your participation by completing a web-based survey. The survey should take approximately 30 minutes of your time.

**Purpose of the online survey**: To report key evidence requirements for which EFPIA company-members may seek ED / HTA scientific advice with the aim to identify evidence generation challenges that must be dealt with by utilising state-of-the-art HTA methods. Clarity on accepted and novel HTA methods would address divergence of opinions (e.g., between HTA bodies and regulators or among HTA bodies), mitigate uncertainty for HTA bodies and close evidence gaps for companies.

Key topics for which HTA advice is sought include relative effectiveness assessments and challenges with endpoint selection, comparator selection, evidence synthesis, health-related quality of life (HRQoL) and utilities, internal validity of clinical studies, and personalized medicine.

**Definitions**:

- **HTA ED or scientific advice**: This dialogue centres on evidence generation plans for new pharmaceuticals and aims to allow manufacturers to obtain feedback on their development plans from at least one HTA body, and in some instances, with participation of regulators (e.g., parallel consultation between EMA and EUnetHTA). The objective is to help generate optimal and strong evidence that satisfies the needs of HTA bodies (https://www.eunethta.eu/services/early-dialogues/)

**Disclosure of your Personal Information**: The results gathered in this online survey will be reported to the EFPIA in aggregate and/or in a de-identified form only. The EFPIA HTA WG may elect to publish or present the de-identified results of the Research Project in public forums; however, your name or other identifying information (e.g., company name) will not be used or disclosed. To minimize the potential conflict of interest, data collection and analysis will be managed by investigators of the research team not associated with the EFPIA HTA WG. Your privacy will be protected. This Research Project is subject to



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

applicable laws governing the protection of Personal Information and your information will be treated confidentially.

**Study team**: This Research Project is being conducted by a research team working for ICON plc under the direction of Manpreet Sidhu and Earlene Biggs. Address: ICON Clinical Research UK Ltd, 80 Wood Lane, 5th Floor Translation and Innovation Hub, London W12 0BZ, United Kingdom. For questions please email the Project Manager, Ioannis Katsoulis (ioannis.katsoulis@iconplc.com).

**[SURVEY WELCOME]**

Welcome to our survey.

The purpose of this online survey is to identify the evidence gaps (e.g., between company and HTA bodies, between HTA bodies, or between HTA bodies and regulators) that arise during scientific advice sessions and that may need to be addressed using specific methodologies.

This survey will take approximately 30 minutes to complete.

The results gathered in this online survey will be reported to the EFPIA in aggregate and/or in a de-identified form only.

Please complete the survey and submit your responses by **March 4, 2019**.

**[QUESTIONAIRE]**

A. How many HTA scientific advice/early dialogue engagements have you personally or your company experienced within the past 5 years?
**[SELECT NUMBER 0-100]**

1. Personal experience
2. Company-wide experience

B. Which of the following types of scientific advice/early dialogues have you previously needed/requested? Select all that apply.
**[Select all that apply]**

1. EUnetHTA Consolidated Parallel Consultation with EMA
2. EUnetHTA Individual Parallel Consultation with EMA
3. EUnetHTA Multi-HTA Early Dialogue
4. Single country joint regulatory and HTA advice/early dialogue
5. Single country HTA advice/early dialogue



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

6. None of the above **[EXCLUSIVE]**

C. In which European HTA bodies have you sought single country HTA advice during the past 5 years? Select all that apply.
**[DISPLAY IF B = 4 OR 5]**
**[Select at least one EU country (tick box)]:**

- HVB (Austria)
- RIZIV-INAMI (Belgium )
- SUKL (Czech Republic)
- FIMEA, FinOHTA, FinCCHTA (Finland)
- HAS (France)
- G-BA (Germany)
- NCPE (Ireland)
- AIFA, RER (Italy)
- ZiN (Netherlands)
- NoMa (Norway)
- AOTMiT (Poland)
- INFARMED (Portugal)
- AEMPS (Spain)
- AETSA (Spain)
- TLV (Sweden)
- NICE (UK)
- SMC (UK)
- Other **[COMPULSORY TEXT]**

D. Which of the following therapeutic areas have you requested HTA scientific advice/early dialogue? Select all that apply.
**[Select at least one from 27 therapeutic areas]:**

- Bone/Osteoporosis
- Cardiovascular
- CNS
- Dermatology
- Diabetes
- Endocrinology
- Ear/Nose/Throat (ENT)
- Gastroenterology
- Haematology
- Immunology
- Infectious diseases
- Inflammation
- Internal diseases



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

- Metabolic diseases
- Nephrology
- Oncology
- Ophthalmology
- Orthopaedics
- Pulmonology
- Respiratory
- Rheumatology
- Surgery
- Urology
- Vaccines
- Other **[COMPULSORY FREE TEXT]**
- Cannot answer **[EXCLUSIVE]**

E. What were the three key evidence generation challenges that required clarity on HTA methods when developing: (a) the company positions for Parallel Consultations from an HTA perspective and (b) company positions in single-HTA scientific advice? Please provide context on the therapeutic area and style of evidence plan.
**[COMPULSORY FREE TEXT]**

F. What were the key issues that you sought advice for, among the following domains (i.e., not limited to HTA methods):
**[Select from list of options (tick box) and please provide full explanation in text box]:**

- Patient population (e.g., sub population with highest unmet need, populations with differing biomarkers)
  **[COMPULSORY FREE TEXT]**
- Comparator selection (e.g., relevant comparator in a treatment landscape with multiple options or only off-label treatments applied in clinical practice)
  **[COMPULSORY FREE TEXT]**
- Endpoints (e.g., surrogates for assessment of patient relevant benefit, appropriateness of surrogate measures as primary endpoints, use of composite primary or secondary endpoints)
  **[COMPULSORY FREE TEXT]**
- Study design (e.g., appropriateness of single arm study or historical cohort as control arm)
  **[COMPULSORY FREE TEXT]**
- HRQL/utilities (e.g., selection of PRO outcome measures, analysis/derivation of utilities)
  **[COMPULSORY FREE TEXT]**
- General evidence generation (e.g. the role of supportive real world evidence or requirements of data that are not feasible to produce)
  **[COMPULSORY FREE TEXT]**
- Analysis and controlling uncertainty (e.g., extrapolating clinical benefit, adjusting for patient cross-over)
  **[COMPULSORY FREE TEXT]**
- Other
  **[COMPULSORY FREE TEXT]**



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

G1. Did you identify any divergences in the advice received from the EMA and HTA bodies for your HTA scientific advice briefing package (i.e., parallel consultation of EUnetHTA with the EMA, https://www.eunethta.eu/services/early-dialogues/parallel-consultations/)?

1. Yes
2. No

**[DISPLAY IF B = 1 OR 2 (EMA-HTA parallel consultation)]**

G2. What were the key underlying evidence and methodological issues in the feedback received that were conflicting or divergent between the EMA and HTA bodies (e.g. use of RWE, MAIC to combine RCT with RWE)?

**[DISPLAY IF G1 = 1 (Yes)]**

**[COMPULSORY FREE TEXT]**

H1A. What is your level of satisfaction with the **clarity** of the additional information provided from the HTA scientific advice/early dialogue for the key methodological issues you selected?

**[SCALE 1-5, 1 = EXTREMELY UNSATISFIED, 5 = EXTREMELY SATISFIED]**

**[INCLUDE ONLY THE TOPICS SELECTED IN QUESTION F]**

H1B. What is your level of satisfaction with the **usefulness** of the additional information provided from the HTA scientific advice/early dialogue for the key methodological issues you selected?

**[SCALE 1-5, 1 = EXTREMELY UNSATISFIED, 5 = EXTREMELY SATISFIED]**

**[INCLUDE ONLY THE TOPICS SELECTED IN QUESTION F]**

H2. Did the HTA scientific advice/early dialogue help you to plan and/or develop the required evidence that you needed for the HTA submissions, provided that the clinical program was successful and advice was implemented?

**[LIST TOPICS SELECTED IN F]**

1. Yes
2. No
3. Confirmed proposed evidence plan was optimal
4. Cannot answer



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

H3. How did the HTA scientific advice/early dialogue on these key issues you identified earlier, listed below, help you to plan and/or develop the evidence for the HTA submissions? Where discussed as a need please provide feedback on recommendations for methodologies to handle post-licensing evidence generation (PLEG) data.

**[LIST TOPICS SELECTED IN F]**
**[DISPLAY IF H2 = 1 (Yes)]**
**[COMPULSORY FREE TEXT]**

H4. What other useful methodological topics not in the guidelines developed by HTA bodies, and not discussed in previous questions, did the HTA scientific advice/early dialogue report provide?
**[COMPULSORY FREE TEXT]**

I Was there a lack of consensus among HTA bodies on the approach for the company to take following the HTA scientific advice/early dialogue engagement for any topics?

1. Yes
2. No

**[DISPLAY RESPONSES FROM QUESTION F TO REMIND]**

J For areas with a lack of consensus and divergence of opinion among the HTA bodies' recommendations, what were your actions following the HTA scientific advice/early dialogue engagement?
**[DISPLAY IF I = 1 (Yes)]**
**[COMPULSORY FREE TEXT]**

K1. In **EUnetHTA Multi-HTA early dialogue** (https://www.eunethta.eu/services/early-dialogues/multi-hta/) and **consolidated parallel consultation** (https://www.eunethta.eu/services/early-dialogues/early-dialogues-faqs/), does the final documentation, clearly state where consensus has been reached amongst HTA bodies and where consensus has not been reached (i.e., divergence of opinion)?

1. Yes
2. No

K2. In which methodological topics was consensus not reached among the HTA bodies? Please provide details or examples of the information on consensus and non-consensus among the participating parties for the different early dialogue pathways (e.g., EUnetHTA Consolidated Parallel Consultation vs. Multi-HTA) in the final documentation.



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

**[DISPLAY IF K1 = 2 (No)]**

**[COMPULSORY FREE TEXT]**

L1. In Multi-HTA early dialogue, have you experienced a question being denied and national advice suggested instead?

1. Yes
2. No

L2. Please provide details on a denied question and the re-direction to national HTA bodies.

**[DISPLAY IF L1 = 1 (Yes)]**

**[COMPULSORY FREE TEXT]**

M. What recommendations do you have for future/additional methodological guidelines (https://www.eunethta.eu/methodology-guidelines/) related to preparation for EU Joint Clinical Assessments?

- Specifically of the recommendations suggested, which do you think are the most pertinent and achievable given the changes under consideration within a broader HTA framework (including ex-EU HTA bodies)?
- Do you have any other comments relating to methodological issues that were not covered by the previous questions (e.g., predictability/consistency of advice, state-of-the-art methods, digital endpoints)?

**[COMPULSORY FREE TEXT]**



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

# Appendix B: Summary questions in Critical Review Framework

**Table 11 Guiding questions per domain of the critical review framework**

| Domain | General guideline appraisal domains |
|---|---|
| **Scope & Purpose** | • Are the overall objective(s) of the guideline described and clear?<br>• Are the overall objective(s) of the guideline addressed?<br>• Are the problem statement(s) covered by the guideline specifically described?<br>• Is the audience/perspective clearly defined?<br>• Is the scope of the guideline fit-for-purpose for the stated audience?<br>• Is the scope of the guideline fit-for-purpose for the decision-problem? |
| **Comprehensi-veness** | • Is the guideline providing a comprehensive list of evidence requirements? Does the guideline restrict to only one type of evidence?<br>• Is the guideline providing a comprehensive list of decision criteria?<br>• Is the guideline providing a comprehensive list of methods?<br>• Are there any important evidence alternatives requirements/methods omitted? |
| **Clarity & Presentation** | • Are the key recommendations specific and unambiguous?<br>• Are the key methods clearly presented? |
| **Applicability** | • Does the guideline address practical HTA decision making?<br>• Does the guideline consider the differences across EU practice and how is this incorporated in the proposed methods?<br>• Is the guideline supported with tools for application (e.g., HTA Core Model)?<br>• Are potential organisational barriers in applying the recommendations discussed? |
| **Appropriate Methods** | • Are the proposed methods reflective of currently used methods by HTA agencies in the EU?<br>• Are there qualitative and/or quantitative approaches presented in the guideline to address the research problem? |
| **State-of-the-art Methods** | • What are the state-of-the-art methods in this area and how close are the EUnetHTA methods to these standards?<br>• What future research needs on the guideline topic would you prioritize based on importance, desirability of new research, feasibility, and impact? |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.

| | |
|---|---|
| **Uncertainty Characterisation** | • How is methodological uncertainty addressed?<br>• What other sources of uncertainty have been identified in the guideline?<br>• Is the guideline biased in terms of data assessment (e.g. types of trials) or background of the guideline author(s) (e.g. scientific expertise or affiliation with HTA agency)<br>• [Clinical and safety endpoints only] Do the proposed methods mitigate uncertainty regarding clinical evidence (e.g. due to limitations in study design, outcome measures, and the size and duration of clinical trials)? |
| **Generalisability / Transferability/ Feasibility** | • Are the proposed (analytical) methods and evidence transferable across the EU jurisdictions?<br>• What HTA related aspects or issues are not addressed in the guideline?<br>• How does the guideline support different types of evidence submissions?<br>• Are recommendations presented to ensure national uptake of the assessments with proposed methods? |
| **Summary questions** | • Is the guideline promoting consistency of evidence reviewed (and interpreted) across the EU HTA bodies?<br>• Is the guideline supporting predictability in the methods that are used in relative effectiveness evaluations? |



© ICON, 2019. ICON holds copyright to the concepts, statements and formats described in this document.
The unauthorized use of any portion of this document is prohibited.