## Appendix 1 – Country profiles

### I. IQWiG

|  |
| --- |
| **IQWiG guidelines included:**   * Institute for Quality and Efficiency in Healthcare (IQWiG), “General Methods - Version 5.0,” Köln (Germany), 2017. [Online]. Available: https://www.iqwig.de/download/General-Methods\_Version-5-0.pdf.   **IQWiG assessments included:**  *Last 3 assessments of the ‘Benefit assessment according to §35a Social Code Book V’ process with available English extracts at the time of retrieval (11/04/2020), excluding addendums.*   * Institute for Quality and Efficiency in Healthcare (IQWiG), “Larotrectinib ( solid tumours with neurotrophic tyrosine receptor kinase [NTRK] gene fusion) – Benefit assessment according to § 35a - English Extract,” Köln (Germany), 2020. [Online]. Available: https://iqwig.de/download/A19-90\_Larotrectinib\_Extract-of-dossier-assessment\_V1-0.pdf. * Institute for Quality and Efficiency in Healthcare (IQWiG), “Ibrutinib (chronic lymphocytic leukaemia) – Benefit assessment according to § 35a - English Extract,” Köln (Germany), 2019. [Online]. Available: https://iqwig.de/download/A19-77\_Ibrutinib\_Extract-of-dossier-assessment\_V2-0.pdf. * Institute for Quality and Efficiency in Healthcare (IQWiG), “Ivacaftor (cystic fibrosis, 12 to < 24 months, with gating mutations) - Benefit assessment according to §35a Social Code Book V - English Extract,” Köln (Germany), 2019. [Online]. Available: https://iqwig.de/download/A19-69\_Ivacaftor\_Extract-of-dossier-assessment\_V1-0.pdf. |

IQWiG assesses all drugs entering the German market (excluding inpatient) (1) making a recommendation to the Federal Joint Committee (G-BA). The G-BA then issues a directive on the reimbursement of the drug (2).

According to IQWiG’s latest ‘General Methods’ (3), evidence-based medicine (EBM) guidelines for the assessment of the relative benefit and its certainty are followed, like for instance those of GRADE. IQWiG then proceeds with an ‘outcomes related assessment’ and, when possible, a ‘summarizing assessment’.

In the outcomes related assessment, it uses ‘certainty of results’ as an umbrella term under which lie the ‘quantitative certainty of results’ (statistical uncertainty; associated to study size) and ‘qualitative certainty of results’ (risk of bias; associated to study design) for the individual study and outcomes level (**IS**). It also uses the term ‘certainty of conclusions’ (for the outcomes level - **BE1**).

Quantitative certainty (**⇩ImP**) of results is used as a synonym to statistical uncertainty caused by random errors, being connected to sample size and expressed by statistical means of standard error and confidence intervals. Qualitative certainty of results (**⇩RoB**) on the other hand is connected to potential bias and study design and is assessed for statistically significant results. IQWiG judges the qualitative certainty of individual studies and outcomes based on a 3-level scale (low, moderate, high), assigning the two highest ranks to RCTs, with low or high risk of bias respectively, and the lowest one to non-randomized comparative studies.

The assessment of certainty of conclusions for a specific outcome (**BE1**) is based on a 3-level scale (hint, indication, proof) and is essentially the product of the qualitative certainty of the body of evidence available for a specific outcome (**⇩RoB**), the number of studies comprising it (**⇩ImP**) and the direction of results (**⇩InC**). The above that comes out of IQWiG’s guidelines, is also confirmed with a statement in the three assessments examined (4–6). However, the full assessment of these domains was not available in the English extract of the assessment reports taken into account. Furthermore, as stated in the same document, “a clear direction of risk of bias (…) can justify an increase in certainty” (**⇧PlC**). The above-mentioned approach is in-line with the outcome centric logic of GRADE, although it differs in its reporting and the wording used. The assessments examined provide us with a similar picture, where the certainty and magnitude of added benefit is reported on a per outcome basis.

In the summary assessment IQWiG weighs the benefits vs. the harms and draws a final conclusion on the certainty and extent of added benefit across all outcomes (**BEA, AB**). The summary is presented in the assessments (English extract) in the ‘Overall conclusion on added benefit’ (4–6).

A chapter entitled ‘Special aspects of the benefit assessment’ of the guidelines includes provisions regarding two more domains that affect conclusions on added benefit, the first being publication bias (**⇩PuB**). There are three possible scenaria affecting the inference of conclusions on added benefit, depending on the judgment surrounding the presence of publication bias as: (i) improbable (no limitations to conclusions), (ii) possible (conclusions drawn with restriction) or (iii) probable (no conclusion on the certainty and extent of added benefit can be drawn). The second is the ‘dramatic effect’ (**⇧LaE**), whose expected presence in the primary research on a specific question, would allow IQWiG to include studies of lower qualitative certainty in the evidence base. Explicit reference to ‘dramatic effect’, or the lack thereof, is made in two out of three assessments examined (4,5). Study duration, patient-reported outcomes and benefits and harms in small populations are domains that are also referred to, there is however no clear link to the quality of evidence ranking.

Four domains regarding the applicability/transferability/generalizability (**⇩InD**) of results have also been identified in the guidelines document. The first two regard the populations’ perspective. The first is the applicability to the German healthcare setting which should be assessed in a way independent to that of study design and qualitative certainty. However, its effect on certainty of conclusions is not explicitly explained. The second is subgroup analysis which is at times ‘a matter of scientific necessity’, and in this sense should not be valued negatively if included in the hypothesis a priori. The third regards indirect comparisons and IQWiG’s clear preference towards direct comparative studies. As a result, conclusions based on indirect comparisons would result in lower overall certainty. Finally, surrogates are also assessed (usually if validated beforehand), however there is no standard of operation towards them.

The scale used by IQWiG to express certainty seems to be similar to the one used by GRADE with the bottom two ranks (low and very low) being merged into one category (low). This is allowed by GRADE as well, provided that a reasoning is also stated (7), however a reasoning was not found in IQWiG’s latest guidelines.

Publication Bias

Plausible Confounding

Applicability in the German Health Setting

Subgroup analysis

NMA

Surrogate outcomes

**LaE**

**IE**

**BE1**

**BEA/AB**

**PlC**

**Figure 2**

Visual representation of certainty considerations operationalization for the Relative Effectiveness Processes of IQWiG. This should be viewed as a supplement aiding the conceptualization and allocation of the domains in the certainty assessment process.

**ICER**

### II. ICER

|  |
| --- |
| **ICER guidelines included:**   * D. A. Ollendorf and S. D. Pearson, “ICER Evidence Rating Matrix: A User’s Guide,” 2020. [Online]. Available: http://icer-review.org/wp-content/uploads/2020/01/ICER\_EBM\_Matrix\_User\_Guide\_013120.pdf.   **ICER assessments included:**  *Last three assessments that included a ‘final report’ by the time of retrieval (12/04/2020)*   * Institute for Clinical and Economic Review (ICER), “Acute Treatments for Migraine - Final Evidence Report,” 2020. [Online]. Available: https://icer-review.org/wp-content/uploads/2019/06/ICER\_Acute-Migraine\_Final-Evidence-Report\_022520.pdf. * Institute for Clinical and Economic Review (ICER), “Additive Therapies for Cardiovascular Disease: Effectiveness and Value - Final Evidence Report,” 2019. [Online]. Available: https://icer-review.org/wp-content/uploads/2019/02/ICER\_CVD\_Final\_Evidence\_Report\_101719.pdf. * Institute for Clinical and Economic Review (ICER), “Duchenne Muscular Dystrophy : Effectiveness and Value Director of Health Economics - Final Evidence Report,” 2019. [Online]. Available: https://icer-review.org/wp-content/uploads/2018/12/ICER\_DMD-Final-Report\_081519.pdf. |

ICER uses the guidelines of three good practices guidelines committees (GRADE, USPSTF and AHRQ) for its evidence to recommendation framework.

They do not endorse a specific scale of one of the good practices guidelines in order to assess the ‘strength of evidence’ (8). However, the term ‘strength of evidence’ itself derives from USPSTF guidelines (9) or AHRQ guidelines (synonym to quality of evidence in initial GRADE and certainty of evidence in later updates of GRADE and USPSTF guidelines (10)). From the assessments included in this study(11–13) there is also indication that they are using the three level (low, fair or moderate, good) scale of USPSTF in assessments for rating individual studies (**IS**). The gradings are cited in the available appendices of the assessment reports. The gradings of individual studies concern risk of bias (**⇩RoB**). Publication bias is stated to be assessed as part of the ‘quality assessment’ in all three assessments (**⇩PuB**).

The focus of ICER guidelines is mostly on the domains (sources) of uncertainty. The domains identified by ICER are a mix of what the three good practices guidelines propose, and these include risk of bias (**⇩RoB**), applicability or generalizability (**⇩InD**), consistency (**⇩InC**), directness (**⇩InD**) and precision (**⇩ImP**). Other domains included are the existence of a dose response relationship (**⇧DoR**), biologic plausibility of the treatment effect and publication bias (**⇩PuB**).

The effect of these domains on the final certainty rating is intentionally not predefined by ICER, meaning they use a purely judgemental approach to account for these factors. This unconstrained approach is in contrast to that of GRADE, according to which there is a pre-set amount that each domain can contribute to the final ranking (**±**1, **±**2, -3). As a result, different factors may play a larger or smaller role, depending on the assessment and needs to be accompanied by high levels of transparency. For example, a large number of studies with high risk of bias but with high consistency of results could even lead to high levels of certainty on the added benefit (8). A clear distinction between certainty and risk of bias is also made, the latter being one of the factors that affect certainty. The important role of judgment is further highlighted by the fact that algorithms and mathematical equations are suggested to be supportive to decision making rather than to be determinant of the conclusions themselves.

The outcome-centred approach of GRADE is not evident in ICER’s guidelines, meaning they judge how ‘the body of evidence stacks up to each domain of uncertainty’ instead of viewing it as a characteristic parameter of each outcome. In the assessments examined there is a discussion on a per outcome basis for the presence and magnitude of clinical benefits and harms but not for the level of certainty that accompanies them.

When making conclusions on overall net benefit, ICER assigns a ‘joint rating’ (a letter) that corresponds to a predefined space on the ICER matrix and gives information both for the magnitude of added benefit as well as the certainty surrounding it (**BEA/AB**). The ICER matrix is comprised of two axes: an x-axis of ascending ‘comparative net health benefit’ and a y-axis of ascending ‘level of certainty in the evidence’. Apart from serving as a visual representation of possible results, it also serves as a conceptual tool for assessing relative clinical effectiveness. On a final note, although the magnitude of the ‘comparative net health benefit’ is presented separately and on a different axis than certainty, it may play a role on the increase of certainty if a large effect is present (**⇧LaE**), as was pointed out by ICER representative in the process of the verification of our results referring us to their assessment of Zolgensma.

### III. HAS

|  |
| --- |
| **HAS guidelines included:**   * Haute Autorité de Santé (HAS), “Pricing and Reimbursement of drugs and HTA policies in France,” 2014. [Online]. Available: http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-03/pricing\_reimbursement\_of\_drugs\_and\_hta\_policies\_in\_france.pdf. * Haute Autorité de Santé (HAS), “Medicinal products assessment. Transparency Committee Doctrine. Principles of medicinal products assessment and appraisal for reimbursement purposes.,” Saint-Denis (France), 2018. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine\_de\_la\_commission\_de\_la\_transparence\_-\_version\_anglaise.pdf. * Haute Autorité de Santé (HAS), “Niveau de preuve et gradation des recommandations de bonne pratique,” Saint-Denis (France), 2013. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat\_des\_lieux\_niveau\_preuve\_gradation.pdf.   **HAS assessments included:**  *Last three assessments that at the time of retrieval (12/04/2020) included an English translated Transparency Committee Opinion Summary. The assessments selected were for drugs being assessed for the first time (nouveau médicament).*   * Haute Autorité de Santé (HAS), “Transparency Committee Opinion Summary - Slenyto (melatonin) (English translation),” Saint-Denis (France). [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-11/slenyto\_summary\_ct17549.pdf. * Haute Autorité de Santé (HAS), “Commision de la Transparence - Transcription - Slenyto (melatonin),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-08/slenyto\_12062019\_transcription\_ct17549.pdf. * Haute Autorité de Santé (HAS), “Commision de la Transparence - Avis - Slenyto (melatonin),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/evamed/CT-17549\_SLENYTO\_PIC\_INS\_Avis2\_CT17549.pdf. * Haute Autorité de Santé (HAS), “Transparency Committee Opinion Summary - Caddera (calcium chloride dihydrate) (English translation),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-11/caddera\_summary\_ct17578.pdf. * Haute Autorité de Santé (HAS), “Commission de la Transparence - Avis - Caddera (calcium chloride dihydrate),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/evamed/CT-17578\_CADDERA\_PIC\_INS\_Avis1\_CT17578.pdf. * Haute Autorité de Santé (HAS), “Commission de la Transparence - Avis - Pifeltro (doravirine),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-05/pifeltro\_synthese\_ct17495.pdf. * Haute Autorité de Santé (HAS), “Transparency Committee Opinion Summary - Pifeltro (doravirine) (English translation),” Saint-Denis (France), 2019. [Online]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2019-08/pifeltro\_summary\_ct17495.pdf. |

The relative effectiveness assessments of HAS usually reach two critical conclusions: SMR (Service Médical Rendu, *EN:* Clinical Benefit) and ASMR (Amélioration du Service Médical Rendu, *EN:* Clinical Added Value). SMR is the answer to the questions: ‘Should the drug be reimbursed? Is the drug clinically interesting?’ while ASMR replies to the question: ‘Does the drug improve patients clinical situation, as compared to existing therapies?’ (14).

The two primary ranks of SMR are (i) sufficient (with subcategories of low, moderate and substantial or important) and (ii) insufficient. The decision on whether SMR is sufficient or not, essentially determines whether a drug will enter the positive reimbursement list. Each subcategory of a sufficient SMR corresponds to a percentage of reimbursement. ASMR is ranked on a five-level scale (from low to high: V, IV, III, II, I) and is related to the price that a new drug will be assigned. The lowest ASMR ranking (V) implies a maximum price of the existing comparators or higher if cost savings are induced, while the rest indicate a possibly higher price.

In the latest document (15) of the Transparency Committee (TC) of HAS, responsible for making recommendations on drug reimbursement to public authorities, there are detailed definitions for both these scales of (added) benefit. In both definitions there is reference to uncertainty as one of the factors affecting the ranking, especially in the lower ranks. This is also verified by three recent assessments taken into account (16–22), where uncertainties are being taken into account to assign an SMR and ASMR rating and are referred to in a narrative way in the summaries that accompany the rankings.

In a document examining various evidence assessment tools, HAS (23) presents its own grading system, which is based on a document (24) of the National Agency for Accreditation and Health Evaluation (ANAES / *FR:* Agence Nationale d'Accréditation et d'Evaluation en Santé) that was later grouped into HAS. Other evidence grading systems are also presented (e.g. GRADE). Although in the introduction of the document it is stated that another document with the final selection of HAS will follow, possibly updating their evidence assessment choices, we could not identify such a document in the website of HAS. There are also more recent references (25) to the same HAS system since the publication of that document, so we may assume that it is still in use. Furthermore, it is remarkable that no reported reference to the HAS evidence assessment framework or any other framework was identified in the three assessments examined.

In the same document, HAS defines ‘quality of research evidence’ (QoRE / *FR*: *niveau de preuve* or *qualité de la demonstration*). The English term ‘level of proof’ is also used, possibly since it is the literal translation from the French term. It refers to single studies (**IS**) and is based on a three-level scale (low, intermediate, strong). Based on the latest TC doctrine (15), QoRE is indicated to be affected by the following domains: statistical significance, sufficient power, whether the statistical analysis is adapted for the objectives or not (**⇩ImP**), directness in terms of comparators, population and endpoints (**⇩InD**), design of the study and risk of bias (**⇩RoB**). Consequently, both SMR and ASMR ranking (**AB**) are affected by QoRE, with an ‘insufficient’ SMR or an ASMR ‘V’ indicating possible issues in QoRE, however the possible effect is not clearly defined. Furthermore, there is reference to the ‘effect size’ - i..e. magnitude of effect – which could also be linked to the concept of large effect (**⇧LaE**), since a large effect would imply higher SMR and ASMR final ratings. It is generally not clear however how the effect size balances against the level of proof when assigning these ratings. That, together with the adverse effects as well as the duration of a study are taken into account in order to assign a ranking for SMR and ASMR and are presented as distinct concepts to that of QoRE[[1]](#footnote-1).

A QoRE for the whole body of evidence (*FR*: *ensemble d’etudes*) for a given research question is also provided by HAS (**BEA**). It is based on a four-level scale (from low to high: 4, 3, 2, 1) and it is affected by the relevance of the data regarding the research question, the aforementioned QoRE for the available studies and the consistency of the results (**⇩InC**). An outcome centered approach has not been identified as well as a fixed approach for inferring rankings on the HAS scale based on the above domains.

### IV. NICE

|  |
| --- |
| **NICE guidelines included:**   * National Institute for Health and Care (NICE), “Guide to the methods of technology appraisal 2013 (PMG9),” 2013. [Online]. Available: https://www.nice.org.uk/process/pmg9/. * National Institute for Health and Care Excellence (NICE), “Single technology appraisal: User guide for company evidence submission template (PMG24),” 2018. [Online]. Available: https://www.nice.org.uk/process/pmg24/. * National Institute for Health and Care Excellence (NICE), “Developing NICE guidelines: the manual (PMG20),” 2014. [Online]. Available: http://www.nice.org.uk/article/pmg20. * NICE, Methods for the development of NICE public health guidance: Appendix H, no. February. 2018. * National Institute for Health and Care Excellence (NICE), “Types of technology appraisal recommendation.” https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/types-of-recommendation (accessed Mar. 30, 2020).   **NICE assessments included:**  *Last three technology appraisal guidance documents at the time of retrieval (13/04/2020):*   * National Institute for Health and Care Excellence (NICE), “Lenalidomide with rituximab for previously treated follicular lymphoma,” 2020. [Online]. Available: https://www.nice.org.uk/guidance/ta627. * National Institute for Health and Care Excellence (NICE), “Patiromer for treating hyperkalemia,” 2019. [Online]. Available: https://www.nice.org.uk/guidance/ta623. * National Institute for Health and Care Excellence (NICE), “Sotagliflozin with insulin for treating type 1 diabetes Technology appraisal guidance,” 2020. [Online]. Available: https://www.nice.org.uk/guidance/ta622. |

NICE views relative effectiveness assessment in parallel to cost-effectiveness, without addressing them in an isolated manner (26). Special weight is attributed to the cost-effectiveness part. This is highlighted by the fact that NICE explicitly adopts a cost-effectiveness threshold for concluding on recommendations (27).

Systematic reviews are used in the reference case in order to identify and assess health effects. All studies included in those should be critically appraised[[2]](#footnote-2), with a clear preference for inclusion of studies of ‘the best available quality’[[3]](#footnote-3) and of RCTs directly comparing the intervention with relevant comparators. Guidelines for conducting systematic reviews of the University of York are cited (27,28) and should be followed by the ‘independent academic groups’ that conduct the analyses.

According to these guidelines, quality assessment aims to establish the proximity of the results of a study to the truth. The domains taken into account are: study design with respect to the research objective, risk of bias (**⇩RoB**), choice of outcome measure, statistical issues (**⇩ImP**), quality of reporting and publication bias (**⇩PuB**), quality of the intervention and generalizability (**⇩InD**). Furthermore, NICE explicitly refers to the following domains in their guidelines: relevance (**⇩InD**), indirect comparisons (**⇩InD**), study design and risk of bias (**⇩RoB**), transparency and reproducibility[[4]](#footnote-4) as well as heterogeneity and inconsistencies (**⇩InC**)[[5]](#footnote-5). All of the above lead to the ‘robustness of the synthesis’ assessment in the end of the synthesis process, which includes the methodological quality of the included studies as well as the credibility of the product of the synthesis process.

Use of validity assessment to conclude on whether the evidence included are weak, moderate or good. This is followed by a critical reflection on the synthesis process (the methodology of the synthesis, the evidence used, the assumptions made, discrepancies and uncertainties identified, expected changes in technology or evidence, aspects that may influence implementation and effectiveness in real settings). Drawing conclusions on its generalisability and robustness. Finally, checking with authors of primary studies is also advised, especially when the number of studies included is small.

NICE has also published a list of minimum ‘quality assessment’ criteria for the ‘relevant clinical effectiveness evidence’ submitted, that include aspects of risk of bias, generalizability, publication bias (29). Indeed, the ‘Committee Papers’ (ID1374, ID877 and ID1376) of all three assessments examined (30–32), include a ‘quality assessment’ chapter in the ‘Company Evidence Submission’ (Document B) that include the above-mentioned minimum criteria. These minimum criteria are not explicitly referred to in the final assessment reports themselves (‘technology appraisal guidance’). From the assessments included we can presume that uncertainty for the body of evidence is expressed in a narrative way in the committee discussion as well as the summaries accompanying the recommendation.

|  |  |
| --- | --- |
| **Table 1**  Quality assessment criteria as described by NICE in their ‘User guide for company evidence submission template’[[6]](#footnote-6) | |
| Was the randomisation method adequate? |  |
| Was the allocation adequately concealed? |  |
| Were the groups similar at the outset of the study in terms of prognostic factors, for example severity of disease? |  |
| Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blind to treatment allocation, what might be the likely impact on the risk of bias (for each outcome)? |  |
| Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for? |  |
| Is there any evidence to suggest that the authors measured more outcomes than they reported? |  |
| Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data? | - |
| Also consider whether the authors of the study publication declared any conflicts of interest. | - |
| Consider how closely the trials reflects routine clinical practice in England. |  |

In order to quantify decision uncertainty, models (that combine clinical with cost effectiveness) are likely to be used. Models are also used in cases of indirectness, differences in population, outcomes, or comparators regarding the evidence available. In most technology appraisals, models are preferred in order to project health benefit and cost effectiveness in the long-term. As a result, uncertainty associated with models and their input data is described extensively in the guidelines of NICE.

If there is high uncertainty about the clinical and/or cost effectiveness of a new intervention, NICE may make an ‘only in research’ recommendation in favour of the generation of additional evidence (33). As shown in one of the assessments included in the study (31), even an assessment that leads to a ‘Recommended’ outcome may include ‘Recommendations for research’ where more valuable clinical evidence could be generated.

It is worth noting that NICE employs a different approach for their ‘guidelines’ processes[[7]](#footnote-7) (34). First, they classify studies based on their study design and complete ‘methodology checklists’ assessing qualitative aspects of studies based on their design (different checklists for systematic reviews and meta-analyses, RCTs, cohort studies, case control studies (35)). The domains identified in the checklists are risk of bias (**⇩RoB**), selection bias (**⇩InC**), performance bias, attrition bias. Therefore, data from non-randomised studies may be required to supplement RCT data. Observational data are associated with higher uncertainty, as is explicitly stated.

Evidence are summarized in GRADE evidence profiles. These tables include the certainty of evidence rating labels implemented by GRADE[[8]](#footnote-8). In the absence of such tables, estimates of clinical effectiveness and the confidence in them are presented in narrative ‘evidence statements’ with predefined structure and included terminology[[9]](#footnote-9). Since NICE endorses GRADE for this process, they also make clear that the quality of a study may vary depending on the outcome considered. As they state however (34), they do not use an ordinal scale (or ‘labels’) for rating quality of evidence across all outcomes but instead express it by using the appropriate wording in their recommendations. Another deviation, apart from not using the GRADE scale, is that of the inclusion of a review of cost effectiveness studies, to produce ‘economic evidence profiles’.

### V. ZIN

|  |
| --- |
| **ZIN guidelines included:**   * Zorginstituut Nederland (ZIN), “Package advice in practice,” Diemen (The Netherlands), 2017. [Online]. Available: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2018/09/05/package-advice-in-practice---deliberations-for-arriving-at-a-fair-package/Package+advice+in+practice+-+Deliberations+for+arriving+at+a+fair+package.pdf. * M. Offringa, W. J. J. Assendelft, and R. J. P. M. Scholten, “Kritisch beoordelen van een artikel,” in Inleiding in evidence-based medicine, Bohn Stafleu van Loghum, onderdeel van Springer Uitgeverij, 2008, pp. 54–142. * Zorginstituut Nederland (ZIN), “Assessment of ‘established medical science and medical practice’ (English summary),” Diemen (The Netherlands), 2015. [Online]. Available: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2015/01/19/assessment-of-established-medical-science-and-medical-practice/Assessment+of+established+medical+science+and+medical+practice.pdf. * Zorginstituut Nederland (ZIN), “Beoordeling stand van de wetenschap en praktijk,” Diemen, 2015. [Online]. Available: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/01/15/beoordeling-stand-van-de-wetenschap-en-praktijk/Beoordeling+stand+van+de+wetenschap+en+praktijk.pdf.ZIN, “Criteria voor beoordeling therapeutische waarde,” 2014. * Zorginstituut Nederland (ZIN), “Criteria voor beoordeling therapeutische waarde,” Diemen (The Netherlands), 2014. [Online]. Available: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/publicatie/2014/07/01/criteria-voor-beoordeling-therapeutische-waarde/Criteria+voor+beoordeling+therapeutische+waarde.pdf. * J. de Boer and P. Pasman, “Procedure beoordeling extramurale geneesmiddelen,” Diemen (The Netherlands), 2018. [Online]. Available: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2016/09/09/procedure-beoordeling-extramurale-geneesmiddelen/Procedure+beoordeling+extramurale+geneesmiddelen+%28GVS%29+versie+november+2018.pdf.   **ZIN assessments included:**  *Last three relative effectiveness assessments that at the time of retrieval (13/04/2020) included a part of them translated in English.*   * Zorginstituut Nederland (ZIN), Farmacotherapeutisch rapport voretigene neparvovec (Luxturna®) bij de behandeling van visusverlies door erfelijke retinale dystrofie met bi-allelische RPE65-mutaties. Diemen (The Netherlands), 2019. * Zorginstituut Nederland (ZIN), “Package advice voretigene neparvovec (Luxturna®),” Diemen (The Netherlands), 2020. [Online]. Available: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2020/02/17/voretigene-neparvovec-luxturna/Voretigene+neparvovec+%28Luxturna®%29+package+advice.pdf. * Zorginstituut Nederland (ZIN), “Farmacotherapeutisch rapport emicizumab (Hemlibra®) als routineprofylaxe van bloedingen bij patiënten met ernstige hemofilie A zonder remmers tegen factor VIII,” Diemen (The Netherlands), 2019. [Online]. Available: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/adviezen/2020/02/17/pakketadvies-emicizumab-hemlibra/Pakketadvies+emicizumab+%28Hemlibra®%29.pdf. * Zorginstituut Nederland (ZIN), “Package advice emicizumab (Hemlibra ®),” Diemen (The Netherlands), 2020. [Online]. Available: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2020/02/17/emicizumab-hemlibra/Emicizumab+%28Hemlibra®%29+package+advice.pdf. * Zorginstituut Nederland (ZIN), “Relative effectiveness report metreleptin (Myalepta ®) for treatment of lipodystrophy,” Diemen (The Netherlands), 2019. [Online]. Available: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2019/10/21/metreleptin-myalepta-for-treatment-of-lipodystrophy/Metreleptin+%28Myalepta®%29+for+treatment+of+lipodystrophy.pdf. |

ZIN assessments cover four criteria (36): necessity, effectiveness, cost-effectiveness and feasibility. For the effectiveness step, which is of interest to this study, the Scientific Advisory Board (*NL:* *Wetenschappelijke Adviesraad* – WAR) reviews scientific reports and makes a clinical assessment regarding the value of the effect (*NL:* *waardering van het effect*), the size of the effect (*NL:* *groote van het effect*) and the probability of the effect (*NL:* *waarschijnlijkheid van het effect*).

Regarding the assessment of individual studies (**IS**), ZIN refers to the necessity of the assessment of their methodological aspects[[10]](#footnote-10). Furthermore, there is reference to further EBM methodology on how to proceed with the critical assessment of one study (37). In these there is reference to various aspects of risk of bias (**⇩RoB**), e.g. blinding and follow-up, which also include publication bias (**⇩PuB**). In the above reference there are checklists cited that assist the assessment of the validity, applicability and importance of the results.

Moving on to the body of evidence (**BE1**, **BEA**), ZIN began implementing the GRADE approach in their relative effectiveness assessment practices in 2015 and the ‘probability of the effect’ assessment is based on it (38). The term ‘quality of evidence’ is also used. The domains of GRADE are also all cited, and the GRADE approach is stated to be followed ‘whenever possible’ (39) when assessing certainty in a body of evidence. The above statement is also verified by two out of the three REAs that have been examined (40–44), where full GRADE evidence profiles (45) are also included in the appendices of the reports – although omitting the **⇩PuB** domain.

In its guidelines[[11]](#footnote-11), ZIN also makes a connection between quality of evidence and conclusions (**AB**) of the assessment depending on the direction of effect (positive/negative). High quality of evidence will in principle be decisive, medium or low quality will either call for further assessment or will be insufficient to take a position and very low quality will be decisive leading to a negative opinion. In a different guidelines document, it is once again verified that a conclusion on ‘therapeutic value’ (*NL:* therapeutische waarde) that is based on less data is of more limited value (46). However, further elaboration on the terms positive or negative effect, positive or negative opinion or that of limited value was not identified.

It should be noted that based on their ‘therapeutic value’, drugs are classified into three categories of less, equal or added value compared to other treatment options (46). ‘Therapeutic value’ is the product of weighing the following criteria (47):

1. Beneficial effects,
2. Adverse effects,
3. Experience:
   1. Limited: <3 years/<100,000 prescriptions/<20,000 patient years,
   2. Sufficient - >100,000 prescriptions/>20,000 patient years,
   3. Extensive - >10 years,
4. Applicability compared to the standard or usual treatment in terms of the population group included:
   1. less broadly used,
   2. equally broadly, or
   3. more broadly used, and
5. Ease of use:

e.g. dosing frequency, time of administration, administration form, taste, packaging, etc.

Viewing the above criteria, it is worth highlighting that the population applicability domain (iv) (one of the four aspects considered in the **⇩InD** domain as defined by GRADE) is considered in the conclusion on the ‘therapeutic value’ itself, rather than just on the certainty of evidence assessment that should accompany it.

The final conclusion is based on the balance between advantages and disadvantages, the quality of evidence, arguments on the ‘appropriate evidence approach’ and input from professional groups and patients. According to the guidelines (42), when determining the final assessment/conclusion quality of evidence is should be expressed in the form of arguments. The fact that there is no further reference to a pre-specified way of expressing the final conclusion and the fact that specific reference to transparency of reporting is made in the guidelines[[12]](#footnote-12) leaves us to presume that the way of communicating the final conclusion is narrative, probably combining the above scale for the ‘therapeutic value’ with certainty arguments.

The above is verified by the REA reports examined. An overall certainty of evidence rating based on the GRADE scale was not identified. Instead, in the Final Assessment part of the report there is a narrative approach in expressing the level of certainty (e.g. ‘*we are* *confident that (the new intervention) is at least as effective (as the existing comparators)* (48))’.

### VI. CADTH

|  |
| --- |
| **CADTH guidelines included:**   * A. Bai, V. Shukla, G. Bak, and G. Wells, Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2012. * Canadian Agency for Drugs and Technologies in Health (CADTH), “Therapeutic Review Framework and Process,” 2019. [Online]. Available: https://cadth.ca/. * Canadian Agency for Drugs and Technologies in Health (CADTH), “Procedures for the CADTH Common Drug Review and Interim Plasma Protein Product Review,” 2020. [Online]. Available: https://cadth.ca/.   **CADTH assessments included:**  *Last three relative effectiveness assessments of the Common Drug Review (CDR) process that at the time of retrieval (13/04/2020) included a Clinical Review Report and a Drug Expert Committee Recommendation.*   * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review - TAFAMIDIS (Vyndaqel) Clinical Review Report,” 2020. * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Tafamidis Meglumine (Vyndaqel – Pfizer Canada ULC),” 2020. [Online]. Available: https://cadth.ca/sites/default/files/cdr/complete/SR0625 Vyndaqel - CDEC Final Recommendation February 20%2C 2020 for posting.pdf. * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Vortioxetine (Trintellix — Lundbeck Canada Inc.),” 2020. [Online]. Available: https://www.cadth.ca/sites/default/files/cdr/complete/SR0611 Trintellix - Final CDEC Recommendation February 14%2C 2020\_For posting.pdf. * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review - Clinical Review Report Vortioxetine Hydrobromide (Trintellix),” 2020. [Online]. Available: https://cadth.ca/sites/default/files/cdr/clinical/sr0611-trintellix-clinical-review-report.pdf. * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Upadacitinib (Rinvoq — AbbVie),” 2020. [Online]. Available: https://www.cadth.ca/sites/default/files/cdr/complete/SR0614 Rinvoq - CDEC Final Recommendation February 6%2C 2020\_for posting.pdf. * Canadian Agency for Drugs and Technologies in Health (CADTH), “CADTH Common Drug Review - Clinical Review Report Upadacitinib (Rinvoq),” 2020. [Online]. Available: https://www.cadth.ca/sites/default/files/cdr/clinical/sr0614-rinvoq-clinical-review-report.pdf. |

A report published by CADTH on quality assessment of evidence (49) defines the terms ‘quality’ and ‘strength of a body of evidence’. Quality refers to individual studies and is used as a synonym to internal validity (related to design, conduct, analysis / selection, measurement and confounding biases) (**⇩RoB**) following the guidelines of the US Agency for Healthcare Research and Quality (AHRQ) published in 2002 (50). AMSTAR and SIGN 50 are the ‘quality assessment instruments’ used to assess internal validity, depending on the study design. The ‘strength of a body of evidence’ refers to the size of the estimated risk and its accompanying confidence intervals, again following the definitions of the AHRQ. It is comprised of three major domains: (1) quality, that has the abovementioned definition (**⇩RoB**), (2) quantity, meaning the number of studies evaluating the question, sample size, magnitude of treatment effect (**⇩ImP**, **⇩InD**) and (3) consistency of findings (**⇩InC**).

According to the ‘Quality Assessment Tools’ (QAT) project carried out in 2012 by CADTH (51) in order to make a choice for an ‘evidence grading system’, GRADE was selected as the QAT of choice. However, after consulting with a CADTH representative it was verified to us that the above report does not describe the current way of practice of the organization. Instead, evaluations are based on an (undisclosed) annotated document. Regarding the Clinical Review of the Common Drug Review (CDR) or the Therapeutic Review processes, CADTH states that ‘authors appraise, analyze and interpret the clinical data to generate a reproducible, transparent and rigorous review of the available clinical evidence’ in its health technology management program guidelines (52) or that ‘CADTH summarizes and critically appraises the relevant studies in the clinical report’ and documents ‘strengths and limitations with respect to both internal validity (…) and external validity’ in its drug reimbursement recommendation guidelines (53–58).

Indeed, based on three recent assessments of the CDR process (or ‘product line’) included in this study (59), internal and external validity are discussed in sections named ‘Critical Appraisal’ of the recommendation reports that follow the presentation of results of pivotal studies, indirect treatment comparisons and other relevant studies taken into account – as indicated in the ‘Clinical Review Report’ template (7). In these chapters, internal and external validity are being discussed in a narrative way. In the same assessments however, there is no reported reference to any ‘evidence assessment framework’ to conduct the abovementioned uncertainty or ‘strength of evidence’ assessments.

Finally, CADTH also provides domains that ‘contribute to the uncertainty of clinical benefit’ in its ‘reimburse with conditions’ recommendation of the CDR process. These include the number of clinical studies, small sample sizes (**⇩ImP**), the absence of comparator groups (**⇩InD**), trial designs (**⇩RoB**), study durations or follow-up (**⇩RoB**), the inability to distinguish disease severity in heterogeneous manifested rare diseases (**⇩InC**), limited to surrogate end points (**⇩InD**), insufficient evidence on meaningful clinical end points (**⇩InD**) and greater uncertainty in statistical analyses (**⇩ImP**). In this sense, uncertainty may also be implied in the final recommendation.

### VII. EUnetHTA

|  |
| --- |
| **EUnetHTA guidelines included:**   * European Network for Health Technology Assessment (EUnetHTA), “EUnetHTA Joint Action 2, Work Package 8. HTA Core Model ® version 3.0 (Pdf),” 2016. [Online]. Available: www.htacoremodel.info/BrowseModel.aspx. * European Network for Health Technology Assessment (EUnetHTA), “Levels of evidence: Internal validity of randomised controlled trials,” 2015. [Online]. Available: https://www.eunethta.eu/wp-content/uploads/2018/01/16\_WP7-SG3-GL-int\_val\_RCTs\_amend2015.pdf. * European Network for Health Technology Assessment (EUnetHTA), “Internal validity of non-randomised studies (NRS) on interventions,” 2015. [Online]. Available: https://www.eunethta.eu/wp-content/uploads/2018/01/Internal-validity-of-non-randomised-studies-NRS-on-interventions\_Guideline\_Final-Jul-2015\_0.pdf. * European Network for Health Technology Assessment (EUnetHTA), “Levels of evidence: Applicability of evidence for the context of a relative effectiveness assessment,” 2015. [Online]. Available: https://www.eunethta.eu/wp-content/uploads/2018/01/Levels-of-Evidence-Applicability-of-evidence-for-the-context-of-a-relative-effectiveness-assessment\_Amended-JA1-Guideline\_Final-Nov-2015.pdf. * European Network for Health Technology Assessment (EUnetHTA), “Comparators & Comparisons: Direct and indirect comparisons,” 2015. [Online]. Available: https://eunethta.eu/wp-content/uploads/2018/01/Comparators-Comparisons-Direct-and-indirect-comparisons\_Amended-JA1-Guideline\_Final-Nov-2015.pdf.   **EUnetHTA assessments included:**  *Last three relative effectiveness assessments at the time of retrieval (13/04/2020) that included a final assessment report:*   * Autoridade Nacional do Medicamento e Produtos de Saude - Portugal and National Centre for Pharmacoeconomics - Ireland, “Relative effectiveness assessment of pharmaceutical technologies. Siponimod for the treatment of adult patients with secondary progressive multiple sclerosis (SPMS) with active disease evidenced by relapses or imaging features of inflammatory activity. Jo,” Diemen, (The Netherlands), 2020. * Ministry of Health of the Republic of Croatia (MIZ Croatia) - Croatia, Agency for Health Technology Assessment and Tariff System (AOTMiT) - Poland, and Dental and Pharmaceutical Benefits Agency (TLV), “Ustekinumab for the treatment of adult patients with moderately to severely active ulcerative colitis (UC) who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or have medical contraindications(...),” EUnetHTA Project ID: PTJA07. 2019. * Finnish Medicines Agency (FIMEA) - Finland, Spanish Agency of Medicine and Sanitary Products (AEMPS) - Spain, and Association of Austrian Social Insurance Institutions (DVSV) - Austria, “Relative Effectiveness Assessment of pharmaceutical technologies. Brolucizumab for the treatment of adults with neovascular (wet) age-related macular degeneration (AMD). Joint Assessment.,” EUnetHTA Report No.: PTJA09. 2020. Available from: https //www.eunethta.eu, Diemen (The Netherlands). |

The HTA Core Model guidelines of EUnetHTA (61) describe nine domains that should be taken into account for health technology assessment, the fourth one being the clinical effectiveness (abbreviated as EFF). The goal of this domain is to assess the ‘net benefit’ (i.e. benefits – harms) as well as the ‘certainty of the evidence’. The reference of the term ‘certainty of the evidence’, which is of interest to us in this study, is stated to derive from the 2007 USPSTF guidelines (10) (which has also been maintained in their latest updates of the guidelines (62)) and can be viewed as a synonym to GRADE’s quality (or certainty) of evidence (63–65), having the same ‘umbrella term’ essence to it. The term ‘certainty of evidence’ is not further elaborated in the document.

Although the USPSTF term is used above, later in the same guidelines there is a general urge to the GRADE approach for grading quality of evidence and strength of recommendations. However, this urge is expressed in a non-binding manner. This is further verified by looking into the three assessments included in this study (66–68): while all three assessments make reference to the GRADE approach (either to state that they are following it or not) only one states to follow it to assess certainty of evidence.

In the same guidelines document, there is a summary of the possible sources of uncertainty. The EFF domain[[13]](#footnote-13) refers to errors and biases as well as appraisal tools on a principles level. In the same section there is the explicit acknowledgement of the difference between ‘quality of evidence’ and ‘validity’ (or ‘risk of bias’), with the first one having the essence that GRADE has assigned to it. The term ‘(methodological) quality of individual study’ is also used as a synonym to internal validity whereas the term ‘quality of the body of evidence’ is used in the sense that GRADE meant it. The sources of bias in a systematic review can be found in the whole evidence base, on the individual study level or for individual endpoints in the single study level.

EUnetHTA has also published a series of three guidelines papers under the title ‘levels of evidence’, giving guidance for the assessment of both the internal and external validity of individual studies (69). In these guidelines there is further elaboration on internal and external validity, recommendations on how to deal with them, including suggestion for specific quality checklists and assessment tools, like those provided by Cochrane. The separated way that these guidelines are presented (rather than being part of an overall uncertainty assessment approach) is also reflected in the assessments included in the present study. Based on the assessment reports reviewed, risk of bias and external validity are being discussed on separate chapters, following EUnetHTA guidelines and an overall verdict on uncertainty is not reached. Instead, there is a narrative way of incorporating reflections on available knowledge in the conclusions part.

When viewing the three ‘Levels of Evidence’ guidelines in combination with the EFF domain in the HTA core model guidelines, we can conclude that there is consideration of all the domains contributing to uncertainty, as set by GRADE. The guidelines of EUnetHTA are based on various guidelines from different HTA agencies and good practice guidelines, including IQWiG, GRADE and AHRQ.

In the ‘Internal Validity for RCTs’ guideline, EUnetHTA makes reference to the process of IQWiG for going from ‘uncertainty with regard to study results’ to ‘conclusions on the evidence base’. When mentioning this process, there is reference to the risk of bias domain (**⇩RoB**) as well as the direction of the estimates of effect (**⇩InC**). In the ‘Internal Validity of NRS on interventions’ there is discussion of all 8 uncertainty domains of GRADE, with primary focus to the possibility of a large effect (**⇧LaE**) as well as the other two upgrading quality of evidence. The viewpoint here is that HTA authors should take into account possible counteraction of the higher RoB of NRS due to these domains. The ‘Applicability of evidence for the context of a relative effectiveness assessment’ guideline, follows the methodological choices proposed in the AHRQ guidelines (70). The Applicability (**⇩InD**) domain here is one covering most of the aspects related to the Indirectness domain of GRADE (3 out of 4), apart from the effect of indirect comparisons. Indirect comparisons, which are another aspect included in the indirectness domain as defined by GRADE, are described in a separate document (60). Finally, applicability for EUnetHTA should also be viewed in the context of its transnational role. As a result, some elements of applicability may need to be judged on a national level.

Coming back to the HTA Core Model guidelines, where ‘certainty of the evidence’ is stated to be one of the two goals of the EFF domain, after having also taken into account the consideration the ‘Levels of Evidence’ guidelines series, we can say that the principles that rule EUnetHTA processes are clear, although not appearing to belong to a holistic approach toward uncertainty.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.**  EUnetHTA guidelines on domains affecting certainty of evidence. | | | | | | | | | | | | |
| Levels of Evidence guidelines series | | | | | | | | | | EFF domain in  HTA core model guideline (68) | | |
| Internal Validity | | | | | | | | Applicability  (with a main focus on RCTs) (66) | | There is explicit reference to the concepts of internal validity, external validity and precision. They state GRADE could be used in assessments and mention the 8 domains taken into account. | | |
| IV for RCTs (67) | | | IV for NRTs (27) | | | | |
| For going from certainty in individual studies to body of evidence, they refer to IQWiG’s process of certainty. | | | The 8 domains of GRADE are referred to with a special focus in the d omains that grade upwards, specifically **⇧LaE** | | | | | The guidelines of AHRQ are followed on assessing applicability. | |
|  | | | As general scope of the guideline | | In the context of GRADE | | |  | | Domains referred to outside the context of GRADE | | |
|  |  |  |  |  | **LaE** | **DoR** | **PlC** |  |  |  |  |  |

## Appendix 2 – About GRADE

### A summary of GRADE

In the core of the GRADE framework lies the ‘outcome centric’ (45) approach towards ‘certainty of evidence’, meaning that certainty is not interpreted as a standalone value that solely depends on the studies themselves, but as one referring to the body of evidence available for each outcome. As a result, for GRADE the quality of a body of evidence varies across outcomes, as does the body of evidence itself as different studies might be included every time depending on the outcome. In a setting where recommendations are being made, like HTA, a method to extract an overall rating across all outcomes is also provided.

It should be clear that under the ‘certainty (or quality) of evidence’ term defined by GRADE, lie all the domains (5+3) that affect the grading of a body of evidence, either downwards (5) or upwards (3) on a four-level scale (very low, low, moderate, high). The domains that rate down the ranking are risk of bias, imprecision, inconsistency, indirectness and publication bias whereas the ones that cause rating it up are large magnitude of effect, a dose response gradient and influence of residual plausible confounding(71).

Rating up mainly concerns observational studies and non-randomised experimental or interventional designs which start from the grade of ‘low’ (2/4). RCTs start from the grade of ‘high’. Deviations from the above rule exist in some applications (72), where ‘extremely serious’ (73) concerns may exist and non-randomised studies start from the ‘high’ level (cf. Table 3, RoB).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3.**  The GRADE domains for rating down the QoE and their meaning | | | | | |
| Color coding | Domain set by GRADE | Explanation by GRADE | Visual Representation (72) | Effect on certainty of evidence | Magnitude of effect |
|  | Risk of bias  (study limitations) | Also referred to as problems with (internal) validity by others. Regards study design or conduct. (72,74) | A close up of a logo  Description automatically generated | Rating down  ⇩ | -1, -2 or -3  levels |
|  | Imprecision  (random error) | Regards how narrow the CIs are (sample size & number of events) and whether ‘they cross a clinical decision threshold that dictates recommending versus not recommending the intervention’. (75) | A close up of a logo  Description automatically generated | Rating down  ⇩ | -1 or -2  levels |
|  | Inconsistency | Regards differences in the magnitude of effect (heterogeneity) and direction of the effect. (76) | A close up of a logo  Description automatically generated | Rating down  ⇩ | -1 or -2  levels |
|  | Indirectness | Regards the applicability (different populations or interventions), surrogate outcomes and indirect comparisons. (77) | A close up of a logo  Description automatically generated | Rating down  ⇩ | -1 or -2  levels |
|  | Publication Bias | Regards cases when studies are left out leading to a downwards (positive studies left out) or an upwards (negative studies left out) bias in the estimate of effect. (78) | A picture containing flower, game  Description automatically generated | Rating down  ⇩ | -1 or -2  levels |
| **LaE** | Large effect | Refers to cases of ‘extreme confidence in effectiveness’. (79) | A picture containing drawing  Description automatically generated | Rating up  ⇧ | +1 or +2 levels |
| **DoR** | Dose response gradient | A long-recognized important factor for establishing a cause-effect relationship. (79) | A close up of a logo  Description automatically generated | Rating up  ⇧ | +1 level |
| **PlC** | Plausible confounding | Also ‘residual confounding or biases’. Regards cases when the final effect might be larger than the data suggest (79) | A picture containing drawing, clock  Description automatically generated | Rating up  ⇧ | +1 level |

Certainty, quality or strength of evidence should not be confused with ‘strength of recommendation’, which is ranked on a two-level scale as either weak or strong. Although moderate or high ‘quality of evidence’ ratings are associated with strong recommendations, GRADE gives examples indicating that this is not always the case (80). ‘Certainty of net benefit’ was also recently introduced (in a non-guideline paper) for clinical guidelines preparation (81).

## References (Appendix 1 & 2)

1.         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. European Journal of Health Economics. 2018;19(1):123–52.

2.         G-BA. The benefit assessment of medicinal products in accordance with the German Social Code, Book Five (SGB V), section 35a [Internet]. [cited 2019 Oct 1]. Available from: https://www.g-ba.de/english/benefitassessment/

3.         Institute for Quality and Efficiency in Healthcare (IQWiG). General Methods - Version 5.0 [Internet]. Köln (Germany); 2017. Available from: https://www.iqwig.de/download/General-Methods\_Version-5-0.pdf

4.         Institute for Quality and Efficiency in Healthcare (IQWiG). Larotrectinib ( solid tumours with neurotrophic tyrosine receptor kinase [NTRK] gene fusion) – Benefit assessment according to § 35a - English Extract [Internet]. Köln (Germany); 2020. Available from: https://iqwig.de/download/A19-90\_Larotrectinib\_Extract-of-dossier-assessment\_V1-0.pdf

5.         Institute for Quality and Efficiency in Healthcare (IQWiG). Ibrutinib (chronic lymphocytic leukaemia) – Benefit assessment according to § 35a - English Extract [Internet]. Köln (Germany); 2019. Available from: https://iqwig.de/download/A19-77\_Ibrutinib\_Extract-of-dossier-assessment\_V2-0.pdf

6.         Institute for Quality and Efficiency in Healthcare (IQWiG). Ivacaftor (cystic fibrosis, 12 to < 24 months, with gating mutations) - Benefit assessment according to §35a Social Code Book V - English Extract [Internet]. Köln (Germany); 2019. Available from: https://iqwig.de/download/A19-69\_Ivacaftor\_Extract-of-dossier-assessment\_V1-0.pdf

7.         Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook [Internet]. 2013. p. 1–103. Available from: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.2uab3znt2cji%5Cnhttp://tech.cochrane.org/gradepro

8.         Ollendorf DA, Pearson SD. ICER Evidence Rating Matrix: A User’s Guide [Internet]. 2020. Available from: http://icer-review.org/wp-content/uploads/2020/01/ICER\_EBM\_Matrix\_User\_Guide\_013120.pdf

9.         U.S. Preventive Services Task Force. Procedure Manual. 2015;(December). Available from: https://www.uspreventiveservicestaskforce.org/Home/GetFile/6/319/procedure-manual2017\_update/pdf

10.        Krist AH, Wolff TA, Jonas DE, Harris RP, LeFevre ML, Kemper AR, et al. Update on the Methods of the U.S. Preventive Services Task Force: Methods for Understanding Certainty and Net Benefit When Making Recommendations. American Journal of Preventive Medicine [Internet]. 2018;54(1):S11–8. Available from: http://dx.doi.org/10.1016/j.amepre.2017.09.011

11.        Institute for Clinical and Economic Review (ICER). Acute Treatments for Migraine - Final Evidence Report [Internet]. 2020. Available from: https://icer-review.org/wp-content/uploads/2019/06/ICER\_Acute-Migraine\_Final-Evidence-Report\_022520.pdf

12.        Institute for Clinical and Economic Review (ICER). Additive Therapies for Cardiovascular Disease: Effectiveness and Value - Final Evidence Report [Internet]. 2019. Available from: https://icer-review.org/wp-content/uploads/2019/02/ICER\_CVD\_Final\_Evidence\_Report\_101719.pdf

13.        Institute for Clinical and Economic Review (ICER). Duchenne Muscular Dystrophy : Effectiveness and Value Director of Health Economics - Final Evidence Report [Internet]. 2019. Available from: https://icer-review.org/wp-content/uploads/2018/12/ICER\_DMD-Final-Report\_081519.pdf

14.        Haute Autorité de Santé (HAS). Pricing and Reimbursement of drugs and HTA policies in France [Internet]. 2014. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-03/pricing\_reimbursement\_of\_drugs\_and\_hta\_policies\_in\_france.pdf

15.        Haute Autorité de Santé (HAS). Medicinal products assessment. Transparency Committee Doctrine. Principles of medicinal products assessment and appraisal for reimbursement purposes. [Internet]. Saint-Denis (France); 2018. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine\_de\_la\_commission\_de\_la\_transparence\_-\_version\_anglaise.pdf

16.        Haute Autorité de Santé (HAS). Transparency Committee Opinion Summary - Slenyto (melatonin) (English translation) [Internet]. Saint-Denis (France); Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-11/slenyto\_summary\_ct17549.pdf

17.        Haute Autorité de Santé (HAS). Commision de la Transparence - Transcription - Slenyto (melatonin) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-08/slenyto\_12062019\_transcription\_ct17549.pdf

18.        Haute Autorité de Santé (HAS). Commision de la Transparence - Avis - Slenyto (melatonin) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/evamed/CT-17549\_SLENYTO\_PIC\_INS\_Avis2\_CT17549.pdf

19.        Haute Autorité de Santé (HAS). Transparency Committee Opinion Summary - Caddera (calcium chloride dihydrate) (English translation) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-11/caddera\_summary\_ct17578.pdf

20.        Haute Autorité de Santé (HAS). Commission de la Transparence - Avis - Caddera (calcium chloride dihydrate) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/evamed/CT-17578\_CADDERA\_PIC\_INS\_Avis1\_CT17578.pdf

21.        Haute Autorité de Santé (HAS). Commission de la Transparence - Avis - Pifeltro (doravirine) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-05/pifeltro\_synthese\_ct17495.pdf

22.        Haute Autorité de Santé (HAS). Transparency Committee Opinion Summary - Pifeltro (doravirine) (English translation) [Internet]. Saint-Denis (France); 2019. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-08/pifeltro\_summary\_ct17495.pdf

23.        Haute Autorité de Santé (HAS). Niveau de preuve et gradation des recommandations de bonne pratique [Internet]. Saint-Denis (France); 2013. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat\_des\_lieux\_niveau\_preuve\_gradation.pdf

24.        Agence Nationale d’Accréditation et d’Evaluation en Santé (ANAES). Guide d’analyse de la littérature et gradation des recommandations [Internet]. Anaes. Paris; 2000. Available from: https://www.has-sante.fr/upload/docs/application/pdf/analiterat.pdf

25.        Autran B, Billette De Villemeur A, Buisson Y, Chavade D, Floret D, Hamers F, et al. Méthodologie de gradation des recommandations en pratique vaccinale. 2016.

26.        Mestre-Ferrandiz J, Puig-Peiró R, Towse A. Researching the Relative Efficacy and Relative Effectiveness of Medicines Across Europe: Analysis and implications for the future of HTA in Europe [Internet]. Vol. 44. 2010. Available from: https://www.ohe.org/system/files/private/publications/344 - Researching\_the\_Relative\_25October2010.pdf?download=1

27.        National Institute for Health and Care (NICE). Guide to the methods of technology appraisal 2013 (PMG9) [Internet]. 2013. Available from: https://www.nice.org.uk/process/pmg9/

28.        Centre for Reviews and Dissemination (CRD). CRD’s guidance for undertaking reviews in healthcare [Internet]. CRD, University of York; 2009. Available from: https://www.york.ac.uk/media/crd/Systematic\_Reviews.pdf

29.        National Institute for Health and Care Excellence (NICE). Single technology appraisal: User guide for company evidence submission template (PMG24) [Internet]. 2018. Available from: https://www.nice.org.uk/process/pmg24/

30.        National Institute for Health and Care Excellence (NICE). Lenalidomide with rituximab for previously treated follicular lymphoma [Internet]. 2020. Available from: https://www.nice.org.uk/guidance/ta627

31.        National Institute for Health and Care Excellence (NICE). Patiromer for treating hyperkalemia [Internet]. 2019. Available from: https://www.nice.org.uk/guidance/ta623

32.        National Institute for Health and Care Excellence (NICE). Sotagliflozin with insulin for treating type 1 diabetes Technology appraisal guidance [Internet]. 2020. Available from: https://www.nice.org.uk/guidance/ta622

33.        National Institute for Health and Care Excellence (NICE). Types of technology appraisal recommendation [Internet]. [cited 2020 Mar 30]. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/types-of-recommendation

34.        National Institute for Health and Care Excellence (NICE). Developing NICE guidelines: the manual (PMG20) [Internet]. 2014. Available from: http://www.nice.org.uk/article/pmg20

35.        NICE. Methods for the development of NICE public health guidance: Appendix H [Internet]. 2018. Available from: https://www.nice.org.uk/process/pmg20/resources/appendix-h-pdf-2549710190

36.        Zorginstituut Nederland (ZIN). Package advice in practice [Internet]. Diemen (The Netherlands); 2017. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2018/09/05/package-advice-in-practice---deliberations-for-arriving-at-a-fair-package/Package+advice+in+practice+-+Deliberations+for+arriving+at+a+fair+package.pdf

37.        Offringa M, Assendelft WJJ, Scholten RJPM. Kritisch beoordelen van een artikel. In: Inleiding in evidence-based medicine [Internet]. Bohn Stafleu van Loghum, onderdeel van Springer Uitgeverij; 2008. p. 54–142. Available from: 10.1007/978-90-313-6636-1\_4

38.        Zorginstituut Nederland (ZIN). Assessment of “established medical science and medical practice” (English summary) [Internet]. Diemen (The Netherlands); 2015. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2015/01/19/assessment-of-established-medical-science-and-medical-practice/Assessment+of+established+medical+science+and+medical+practice.pdf

39.        Zorginstituut Nederland (ZIN). Beoordeling stand van de wetenschap en praktijk [Internet]. Vol. januari. Diemen; 2015. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/01/15/beoordeling-stand-van-de-wetenschap-en-praktijk/Beoordeling+stand+van+de+wetenschap+en+praktijk.pdf

40.        Zorginstituut Nederland (ZIN). Farmacotherapeutisch rapport voretigene neparvovec (Luxturna®) bij de behandeling van visusverlies door erfelijke retinale dystrofie met bi-allelische RPE65-mutaties [Internet]. Diemen (The Netherlands); 2019. 6–58. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/adviezen/2020/02/17/pakketadvies-emicizumab-hemlibra/Pakketadvies+emicizumab+%28Hemlibra®%29.pdf

41.        Zorginstituut Nederland (ZIN). Package advice voretigene neparvovec (Luxturna®) [Internet]. Diemen (The Netherlands); 2020. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2020/02/17/voretigene-neparvovec-luxturna/Voretigene+neparvovec+%28Luxturna®%29+package+advice.pdf

42.        Zorginstituut Nederland (ZIN). Farmacotherapeutisch rapport emicizumab (Hemlibra®) als routineprofylaxe van bloedingen bij patiënten met ernstige hemofilie A zonder remmers tegen factor VIII [Internet]. Diemen (The Netherlands); 2019. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/adviezen/2020/02/17/pakketadvies-emicizumab-hemlibra/Pakketadvies+emicizumab+%28Hemlibra®%29.pdf

43.        Zorginstituut Nederland (ZIN). Package advice emicizumab (Hemlibra ®) [Internet]. Diemen (The Netherlands); 2020. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2020/02/17/emicizumab-hemlibra/Emicizumab+%28Hemlibra®%29+package+advice.pdf

44.        Zorginstituut Nederland (ZIN). Relative effectiveness report metreleptin (Myalepta ®) for treatment of lipodystrophy [Internet]. Diemen (The Netherlands); 2019. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2019/10/21/metreleptin-myalepta-for-treatment-of-lipodystrophy/Metreleptin+%28Myalepta®%29+for+treatment+of+lipodystrophy.pdf

45.        Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011;64(4):383–94.

46.        Zorginstituut Nederland (ZIN). Criteria voor beoordeling therapeutische waarde [Internet]. Diemen (The Netherlands); 2014. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/publicatie/2014/07/01/criteria-voor-beoordeling-therapeutische-waarde/Criteria+voor+beoordeling+therapeutische+waarde.pdf

47.        de Boer J, Pasman P. Procedure beoordeling extramurale geneesmiddelen [Internet]. Diemen (The Netherlands); 2018. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2016/09/09/procedure-beoordeling-extramurale-geneesmiddelen/Procedure+beoordeling+extramurale+geneesmiddelen+%28GVS%29+versie+november+2018.pdf

48.        Bai A, Shukla V, Bak G, Wells G. Quality Assessment Tools Project Report [Internet]. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012. Available from: https://www.cadth.ca/media/pdf/QAT\_final.pdf

49.        West S, King V, Carey T. Systems to rate the strength of scientific evidence. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. April 2002.;

50.        Berkman N, Lohr K, Ansari M, McDonagh M, Balk E, Whitlock E, et al. Grading the Strength of a Body of Evidence when Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Comparative Effectiveness Reviews [Internet]. AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2013. Available from: http://www.effectivehealthcare.ahrq.gov/reports/final.cfm

51.        Canadian Agency for Drugs and Technologies in Health (CADTH). Therapeutic Review Framework and Process [Internet]. 2019. Available from: https://cadth.ca/

52.        Canadian Agency for Drugs and Technologies in Health (CADTH). Procedures for the CADTH Common Drug Review and Interim Plasma Protein Product Review [Internet]. 2020. Available from: https://cadth.ca/

53.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review - TAFAMIDIS (Vyndaqel) Clinical Review Report. 2020.

54.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Tafamidis Meglumine (Vyndaqel – Pfizer Canada ULC) [Internet]. 2020. Available from: https://cadth.ca/sites/default/files/cdr/complete/SR0625 Vyndaqel - CDEC Final  Recommendation February 20%2C 2020 for posting.pdf

55.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Vortioxetine (Trintellix — Lundbeck Canada Inc.) [Internet]. 2020. Available from: https://www.cadth.ca/sites/default/files/cdr/complete/SR0611 Trintellix - Final CDEC Recommendation February 14%2C 2020\_For posting.pdf

56.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review - Clinical Review Report Vortioxetine Hydrobromide (Trintellix) [Internet]. 2020. Available from: https://cadth.ca/sites/default/files/cdr/clinical/sr0611-trintellix-clinical-review-report.pdf

57.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review- CADTH Canadian Drug Expert Committee Recommendation (Final): Upadacitinib (Rinvoq — AbbVie) [Internet]. 2020. Available from: https://www.cadth.ca/sites/default/files/cdr/complete/SR0614 Rinvoq - CDEC Final Recommendation February 6%2C 2020\_for posting.pdf

58.        Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Common Drug Review - Clinical Review Report Upadacitinib (Rinvoq) [Internet]. 2020. Available from: https://www.cadth.ca/sites/default/files/cdr/clinical/sr0614-rinvoq-clinical-review-report.pdf

59.        Canadian Agency for Drugs and Technologies in Health (CADTH). Clinical Review Report (Sample Template) [Internet]. CADTH. 2020. Available from: https://www.cadth.ca/sites/default/files/cdr/clinical/SR0430\_KalydecoR117H\_CL\_Report.pdf

60.        European Network for Health Technology Assessment (EUnetHTA). EUnetHTA Joint Action 2, Work Package 8. HTA Core Model ® version 3.0 (Pdf) [Internet]. 2016. Available from: www.htacoremodel.info/BrowseModel.aspx

61.        Sawaya G, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on the Methods of the U.S. Preventive Services Task Force: Estimating Certainty and Magnitude of Net Benefit. Annals of Internal Medicine. 2007;147(2):871–5.

62.        Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. Medical Care. 2010;48(6 SUPPL.):145–52.

63.        Autoridade Nacional do Medicamento e Produtos de Saude - Portugal, National Centre for Pharmacoeconomics - Ireland. Relative effectiveness assessment of pharmaceutical technologies. Siponimod for the treatment of adult patients with secondary progressive multiple sclerosis (SPMS) with active disease evidenced by relapses or imaging features of inflammatory activity. Jo. Diemen, (The Netherlands); 2020.

64.        Ministry of Health of the Republic of Croatia (MIZ Croatia) - Croatia, Agency for Health Technology Assessment and Tariff System (AOTMiT) - Poland, Dental and Pharmaceutical Benefits Agency (TLV). Ustekinumab for the treatment of adult patients with moderately to severely active ulcerative colitis (UC) who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or have medical contraindications(...). EUnetHTA Project ID: PTJA07. 2019.;

65.        Finnish Medicines Agency (FIMEA) - Finland, Spanish Agency of Medicine and Sanitary Products (AEMPS) - Spain, Association of Austrian Social Insurance Institutions (DVSV) - Austria. Relative Effectiveness Assessment of pharmaceutical technologies. Brolucizumab for the treatment of adults with neovascular (wet) age-related macular degeneration (AMD). Joint Assessment. Diemen (The Netherlands): EUnetHTA Report No.: PTJA09. 2020. Available from: https //www.eunethta.eu;

66.        European Network for Health Technology Assessment (EUnetHTA). Levels of evidence: Internal validity of randomised controlled trials [Internet]. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/16\_WP7-SG3-GL-int\_val\_RCTs\_amend2015.pdf

67.        European Network for Health Technology Assessment (EUnetHTA). Internal validity of non-randomised studies (NRS) on interventions [Internet]. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/Internal-validity-of-non-randomised-studies-NRS-on-interventions\_Guideline\_Final-Jul-2015\_0.pdf

68.        European Network for Health Technology Assessment (EUnetHTA). Levels of evidence: Applicability of evidence for the context of a relative effectiveness assessment [Internet]. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/Levels-of-Evidence-Applicability-of-evidence-for-the-context-of-a-relative-effectiveness-assessment\_Amended-JA1-Guideline\_Final-Nov-2015.pdf

69.        Atkins D, Chang S, Gartlehner G, Buckley D, Whitlock E, Berliner E, et al. Assessing the Applicability of Studies When Comparing Medical Interventions. AHRQ Publication No.11-EHCO19-EF. 2011;

70.        European Network for Health Technology Assessment (EUnetHTA). Comparators & Comparisons: Direct and indirect comparisons [Internet]. 2015. Available from: https://eunethta.eu/wp-content/uploads/2018/01/Comparators-Comparisons-Direct-and-indirect-comparisons\_Amended-JA1-Guideline\_Final-Nov-2015.pdf

71.        Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology. 2011;64(4):401–6.

72.        Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. Journal of Clinical Epidemiology. 2019;111(2019):105–14.

73.        Piggott T, Morgan RL, Cuello-Garcia CA, Santesso N, Mustafa RA, Meerpohl JJ, et al. Grading of Recommendations Assessment, Development, and Evaluations (GRADE) notes: extremely serious, GRADE’s terminology for rating down by three levels. Journal of Clinical Epidemiology. 2020;120:116–20.

74.        Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence - Study limitations (risk of bias). Journal of Clinical Epidemiology. 2011;64(4):407–15.

75.        Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence - Imprecision. Journal of Clinical Epidemiology. 2011;64(12):1283–93.

76.        Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence - Inconsistency. Journal of Clinical Epidemiology. 2011;64(12):1294–302.

77.        Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence - Indirectness. Journal of Clinical Epidemiology. 2011;64(12):1303–10.

78.        Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - Publication bias. Journal of Clinical Epidemiology. 2011;64(12):1277–82.

79.        Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology. 2011;64(12):1311–6.

80.        Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation - Determinants of a recommendation’s direction and strength. Journal of Clinical Epidemiology. 2013;66(7):726–35.

81.        Alper BS, Oettgen P, Kunnamo I, Iorio A, Ansari MT, Murad MH, et al. Defining certainty of net benefit: A GRADE concept paper. BMJ Open. 2019;9(6):1–8.

1. § 3.3 \ ▸Sufficient CB [↑](#footnote-ref-1)
2. §5.2.6 (27) [↑](#footnote-ref-2)
3. §5.2.1 (27) [↑](#footnote-ref-3)
4. §3.2 (27) [↑](#footnote-ref-4)
5. §5.2.10 (29) [↑](#footnote-ref-5)
6. § 2.5 (34) [↑](#footnote-ref-6)
7. NICE defines guidelines as ‘evidence-based recommendations for health and care in England’. This is a different process to that of ‘technology appraisal’. [↑](#footnote-ref-7)
8. §6.4 \ Evidence Statements (34) [↑](#footnote-ref-8)
9. §6.4 \ Structure and content of evidence statements (39) [↑](#footnote-ref-9)
10. §3.2.1 (39) [↑](#footnote-ref-10)
11. §3.3.2 (38) [↑](#footnote-ref-11)
12. § Fixed EBM Steps \ Step 3: Determining the final assessment/conclusion (48) [↑](#footnote-ref-12)
13. § ‘Methodology’ section \ ‘Tools for Critical Appraisals’, ‘Analyzing and Synthesizing Evidence’ & ‘Reporting and Interpreting’ (60). [↑](#footnote-ref-13)