Review of methods, processes and topic selection for health technology evaluation programmes: conclusions and final update

Appendix: Further discussion and rationale for conclusions – methods

The methods review presents a broad package of improvements to our methods of health technology evaluation, addressing methodological issues across a wide range of topics. Following important feedback from the consultation on the proposals for change, we have incorporated a number of amendments, refinements and improvements to the draft methods. These include implementation of the proposed severity modifier, clarification of how we will consider uncertainty, updates to how we use a broad evidence base (including real-world evidence, health-related quality of life, and costs), additional information on non-reference-case analyses, and rebalancing the aligned manual to address issues relevant to medicines, medical technologies and diagnostics. Several key topics needed further detailed consideration at this stage of the review, to establish the conclusions presented in the Board paper and the final methods presented in the programme manual. This appendix provides additional details and discussions for those key topics. It explores issues raised by stakeholders, taking into account the full review of all consultation comments, and focusing on the key topics and areas of concern; additional issues that were raised but not shown in this document were considered but no further discussion was needed.

This appendix discusses:

* aligning the methods across health technology evaluation programmes
* severity of disease
* uncertainty
* health inequalities
* discounting
* understanding and improving the evidence base.

Aligning the methods across health technology evaluation programmes

An important objective of this stage of the methods review was to update and refine the draft manual. This aims to ensure that the manual is suitably aligned across NICE’s health technology evaluation programmes and that it provides relevant, implementable methods guidance that is appropriate for all evaluation contexts. During the consultation, stakeholders raised queries about the alignment of the methods, and the implications for medical technologies and diagnostics evaluations in particular.

The draft manual has been refined and rebalanced in several places, and updates relevant to medical technologies and diagnostics have been incorporated. These include terminology and application of methods in different evaluation contexts, ensuring the methods are clear and are appropriate to different types of technology.

Some stakeholders raised broader issues, such as the role of incremental cost-effectiveness ratios (ICERs) and cost-incurring technologies in the medical technologies evaluation programme. These issues are outside the scope of the methods review.

Severity of disease

At consultation we proposed a quantitative modifier that gives additional weight to health benefits in the most severe conditions, while the additional considerations for life-extending treatments at the end of life (the ‘end of life modifier’) would be removed. The new severity modifier was developed with an overall size similar to the end of life modifier (referred to as ‘opportunity cost neutral’), by matching the average quality-adjusted life year (QALY) weight per decision for the severity modifier to that under end of life. The proposed modifier used absolute and proportional QALY shortfall to measure severity and defined cut-offs at which additional QALY weights would apply, using 2 steps for different levels of severity. Two alternative options for the QALY weightings were presented.

The consultation revealed a broad range of views from across the stakeholder community. Consistent with the previous consultation, stakeholders broadly supported the proposal to introduce a severity modifier, in principle. They raised questions and concerns that fall under 3 headings:

* the overall size of the modifier
* the design of the modifier
* how it would be implemented and used in practice.

Size of the modifier

Several stakeholders commented on the overall size of the proposed modifier, and its effect on access to health technologies and health technology pricing. They raised issues including links to the policy environment (such as the Life Sciences Vision and EU exit), the effect on life sciences and innovations in general and on specific technologies (such as cancer and non-cancer medicines), and the effects on patients and patient access to health technologies. Some stakeholders proposed that we should adopt a greater modifier (that is, with lower cut-off levels and/or greater QALY weights). They argued that the proposed modifier would be a backwards step in the level of value applied to new medicines, would insufficiently value technologies for serious conditions and would make it harder for many patients (such as cancer patients) to access new treatments. Many stakeholders commented in particular on the role of the 2019 voluntary scheme for branded medicines pricing and access (‘2019 Voluntary Scheme’) in capping the medicines bill through the affordability mechanism. They argued that a larger modifier could be introduced without additional spending or cost impacts on the NHS. These comments have been explored in detail. Overall, we consider that the effects of a larger modifier on costs and NHS spending are important considerations and extend beyond the reach of the 2019 Voluntary Scheme, so must be taken into account.

It is important to emphasise that NICE has critical responsibilities both in supporting innovative health technologies and in ensuring effective use of NHS resources. In the context of the proposed severity modifier, alongside its advantages in supporting health benefits and innovation for the most severe diseases, we must recognise the effects of healthcare displacement and opportunity cost in the NHS. A larger modifier would displace a greater amount of technologies, services, care and therefore health, for patients elsewhere in the NHS. The importance and impact of resource constraints and opportunity costs are perhaps acutely visible as the NHS responds to the COVID-19 pandemic, when the substantial pressures facing the NHS are shown in sharp relief. There is evidence that society values highly health benefits in severe diseases, but there is not enough evidence to clearly define the magnitude of that societal value, so we could not establish whether additional displacement would be justified. We have therefore agreed, as a starting point, a pragmatic approach to the modifier with an overall magnitude similar to that applied under the end-of-life criteria, to value health technologies for the most severe conditions while maintaining a similar level of health displacement.

The effect of the 2019 Voluntary Scheme in the context of displacement and opportunity cost has been explored in detail previously. Importantly, while the 2019 Voluntary Scheme is a valuable agreement with important benefits for the health system, it does not remove the need for robust, evidence-based health technology evaluations, and it does not eliminate the potential displacement effects and opportunity costs of introducing new health technologies. Because of intellectual property protection, the potential future costs incurred from spending on new health technologies extend well beyond the life of the 2019 Voluntary Scheme agreement. Although the 2019 Voluntary Scheme affordability mechanism (capping branded medicines sales at an agreed level, with sales above this level being paid back by scheme members) works based on a projected level of growth, the NHS has a statutory duty to operate within its overall financial envelope each year and it cannot defer payments to future years to compensate. Moreover, not all technologies that could receive a severity modifier are covered by the 2019 Voluntary Scheme (including medical technologies that are evaluated as technology appraisals, and medicines that are not covered by the scheme). Therefore, displacement and opportunity costs remain an important consideration in introducing the severity modifier.

Having considered the consultation responses in detail, and taking into account NICE’s responsibilities to innovation, patients and the NHS, we conclude that the proposed size of the severity modifier, based on opportunity cost neutrality compared with the end of life modifier, remains an appropriate, pragmatic starting point on which to base the modifier. We consider that this approach provides an appropriate balance between the importance of valuing severity and the displacement impact on the NHS. There is no case to change from the principles described in the consultation.

In that context, there remains a critical need for further research on societal value for severity, and ongoing scrutiny and monitoring of the modifier in practice; see section 42.

Design of the modifier

The consultation revealed broad support for the overall design of the modifier based on proportional and absolute QALY shortfall and corresponding cut-offs and QALY weights. The key concerns raised included the number of steps to reflect the spectrum of severity and the values chosen for the cut-offs and weights.

Number of steps

Stakeholders commented that the severity of diseases is a spectrum, and suggested that a sliding scale or greater number of steps should be used to better reflect the variation in severity. This was explored in detail in the previous stage of the methods review, and no new information was identified in the consultation responses. Importantly, any modifier must strike a balance between a highly precise reflection of severity and the practicalities for implementing the modifier in real evaluations. In practice, there will inevitably be uncertainties in the shortfall estimates and often a need for broader consideration of the totality of the evidence (including, for example, uncaptured effects) on which the committee must deliberate, meaning that a highly granular definition of severity levels risks needing spuriously precise estimates. In addition, there is a need for rapid, predictable and consistent decision making and efficient commercial negotiations to support timely access decisions. Although there may be theoretical appeal to a sliding scale or greater degree of granularity, we consider that it would not be practical. We conclude that the 2-step modifier strikes an appropriate balance in reflecting the range of severe conditions while remaining practical to implement.

Cut-offs and weights

Several stakeholders made proposals to change the cut-offs for absolute and proportional shortfall, and/or to apply higher QALY weights. It is perhaps helpful to reiterate that the severity modifier is intended to reflect an exceptional circumstance in which an additional weight is applied to health benefits in the most severe diseases. The current cut-offs (quite rightly) capture a greater number and broader range of topics than the end of life consideration, and to broaden further would contradict the principle of exceptionality. Moreover, in the context of opportunity cost neutrality (as confirmed in sections 8 to 11), any changes to the cut-offs or weights would need compensatory changes elsewhere – for example, applying a greater maximum weight than 1.7 might mean reducing the lower weight or restricting the number of topics receiving weights, whereas expanding the severity ranges would need a compensatory reduction in weights. Such approaches might be achievable but would be difficult to justify and would conflict with other consultation comments.

Stakeholders raised 2 specific points that warrant further consideration: the dataset used to generate the cut-offs and weights within the opportunity cost neutrality principle, and concerns about equalities and fairness.

#### Dataset

Some stakeholders noted limitations in the dataset used to calculate the average QALY weights on which the opportunity cost neutral options were based. Key concerns included:

* The data covered a sample of appraisals from 2011 to 2019, and not the whole time period in which the end of life criteria were active.
* The analyses were based on past appraisals, and not future topics to which the new modifier will apply.
* The analyses considered the average weight per decision, and did not account for the relative size (for example, population size or budget impact) of each.

We acknowledge stakeholders comments and agree that the data does carry some uncertainties. In particular, the uncertainty arising from using retrospective data for a future modifier is acknowledged. However, it is not possible to predict QALY shortfalls for future evaluations without introducing even greater uncertainty. Accounting for the relative sizes of appraisals may improve the analysis, although this would need substantial additional data and analysis. Nevertheless, in the context of this modifier as a pragmatic solution in the absence of full, quantitative evidence on the extent of societal value for severity, we consider that the analysis and approach provides a sufficient basis from which to establish a modifier. We will monitor the use of the modifier in practice, which may allow us to generate better evidence on QALY shortfalls and weightings in future evaluations (see section 42).

In addition, the uncertainties associated with using retrospective data means that the most relevant date range for past appraisals is not clear-cut. There have been several changes in health technologies, evaluations and the policy landscape over the relevant time period (for example, the previous methods review in 2013, the introduction of the cancer drugs fund, and iterations of the voluntary schemes and government policies). While we would ideally seek to use appraisals that most closely represent future evaluations, this is not possible to predict with any degree of certainty.

Nevertheless, for completeness, an additional analysis was done to expand the dataset to include a wider date range. This included appraisals for which absolute and proportional shortfall could be calculated in the time available, between January 2009 and April 2011 and February 2020 and March 2021. It is acknowledged that this sample has limitations and additional uncertainties, so this analysis is considered exploratory.

A total of 99 additional decisions were included, bringing the total sample from 364 to 463 decisions. Overall, the expanded dataset did not change the conclusions from the data: although there was a small change in mean QALY weights both under end of life and severity, the proposed modifier (both options 1 and 2) remained opportunity cost neutral (table 2). Analysis of mean QALY weights over time showed notable variation over time but no clear trend, perhaps reflecting the natural variation in topics being evaluated (figure 2). We conclude that the additional analysis is reassuring, and provides no signal that a change in QALY weights or cut-offs is needed.

Figure 2. Average QALY weight per decision over time, from January 2009 to March 2021, under end of life or 2 options for a severity modifier



Table 2. Mean QALY weight under end of life or severity (option 1 or 2) for the original and expanded range datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean weight: End of life | Mean weight: Severity modifier option 1 | Difference: option 1 versus end of life | Mean weight: Severity modifier option 2 | Difference: option 2 versus end of life |
| Original dataset: 2011 to 2019 (n=364) | 1.125 | 1.119 | -0.006-0.56% | 1.117 | -0.008-0.67% |
| Expanded dataset: 2009 to 2021 (n=463) | 1.122 | 1.116 | -0.006-0.56% | 1.114 | -0.008-0.71% |

#### Equalities

It is crucial to consider whether the proposed severity modifier creates any issues relating to equality and fairness, within NICE’s legal and ethical duties to eliminate discrimination and promote equality. Some stakeholders flagged equalities concerns with the design of the modifier, which we considered alongside other potential equality issues. This document summarises our considerations; a full discussion of these issues is provided separately, in the equality impact document.

In general, the principle of the severity modifier is accepted – there is evidence that society values highly health benefits in severe diseases, and it is legitimate that NICE values benefits in line with this societal value. Applying a modifier as a QALY weight, using QALY shortfall to set boundaries, is a pragmatic and objective way to achieve the above aim. Absolute and proportional shortfall are used by other health technology assessment bodies, and are appropriate ways to measure severity as the amount of health lost because of the condition. Using both together mitigates the limitations of each.

QALY shortfall depends on the age of the population. Equality concerns relating to age have been raised for absolute and proportional QALY shortfall in the past (for example, in discussions around Value Based Assessment). It may be expected that there will be some variation in the application of the modifier by age. This is because it is reasonable to expect that different severe diseases will arise differently at different ages because of the biology of the conditions. Nevertheless, it is appropriate to consider any concerns relating to fairness by age. In the consultation, stakeholders argued that the shortfall cut-offs are “imbalanced”, such that the proportional shortfall cut-offs are more stringent, and that there are limited scenarios in which older people could qualify for the modifier.

Concerns about imbalance are perhaps not truly informative to equality discussions. The validity and meaning of balance are value judgements (that is, it is debatable whether absolute and proportional shortfall should be balanced, and if so what such balance would look like; each measures a different aspect of severe diseases). Stakeholder concerns about the limited scenarios in which older populations could qualify for the modifier are not supported by how often such scenarios arise. Reviewing past appraisals by age shows that the probability that younger and older groups reach the proposed proportional shortfall cut-offs is similar (table 3), and there is no visible age profile to the modifier (figure 3). This contrasts with the end of life modifier, which was more frequently applied in older groups (figure 4). Although a higher proportion of younger people benefit from the severity modifier, the number of topics is lower, and this variation is likely to reflect the severity of childhood diseases evaluated and the number of topics. Alternative cut-off options (including possible ones that meet the principle of opportunity cost neutrality, and a further stakeholder-suggested one) do not substantially change the age profile of the modifier application (figure 5). We conclude that the proposals do not create inappropriate unfairness on the basis of age, and may promote equality. We also conclude that there is no signal that a change in modifier design or shortfall cut-offs is needed.

Another potential equality consideration is that of disability. The proposed modifier presents a broader, fairer and more evidence-based way to reflect societal value than end of life. The effects of some conditions that are disabilities under the Equality Act will meet the definition of severity in the modifier; others will not. Similarly, some end of life conditions would be defined as disabilities but would get a lower or no severity modifier. This is not inappropriate, as the Equality Act and the severity modifier serve different purposes. There may be circumstances in which the nature of the disability affects application of the modifier (for example, the appropriateness of EQ‑5D, availability of evidence). Existing arrangements for addressing such issues (for example, alternative measures of health-related quality of life [HR-QoL], deliberation and reasonable adjustments by committee) are appropriate and will be considered throughout the evaluations. For completeness, we will include disabilities in examples and case studies when exploring the implementation. We conclude there are no additional equality concerns for disability associated with the proposed modifier.

Considering other protected characteristics, and exploring in detail issues for medical technologies, diagnostics, highly specialised technologies, other complexities and implementation challenges, and our duties under the Public Sector Equality Duty identified no further equality concerns.

Having explored equalities in detail, we are reassured that the proposals do not create inappropriate unfairness or equality concerns, and there is no signal that a change in modifier design or shortfall cut-offs is needed.

Table 3. Percentage of appraisals in under or over 65s that would meet the proposed proportional shortfall cut-offs (appraisals from January 2011 to November 2019 for which shortfall data are available; n=364 decisions)

|  |  |  |  |
| --- | --- | --- | --- |
|  | n | PS≥0.85 | PS≥0.95 |
| Appraisals of <65s | 298 | 26% | 7% |
| Appraisals of ≥65s | 66 | 39% | 6% |

Figure 3. Percentage of past topics that would have had a severity weighting, by age band (appraisals from January 2011 to November 2019 for which shortfall data are available; n=364 decisions)



AS, absolute shortfall; PS, proportional shortfall

Figure 4. Percentage of past topics that received the end of life modifier, by age band (appraisals from January 2011 to November 2019 for which shortfall data are available; n=364 decisions)



Figure 5. Percentage of past topics that would have had a severity weighting, by age band, for 2 alternative cut-off scenarios. (A) Proportional shortfall 0.8, 0.95; absolute shortfall 14, 18. (B) Stakeholder proposed option for exploration only: Proportional shortfall 0.75, 0.9; absolute shortfall 12, 18. (Appraisals from January 2011 to November 2019 for which shortfall data are available; n=364 decisions)

A



B



#### Choice of option 1 or option 2

Given that there is no case to depart from the principle of opportunity cost neutrality and no case to reconsider the weights and cut-offs proposed in the consultation, we need to establish a preferred weighting profile from the 2 options presented for consultation. Reviewing the consultation comments in detail, we conclude that there is no clear mandate from the consultation to choose between either of the options. Many respondents did not express a preference. Of those of who did, many added heavy caveats. The written responses provided similarly persuasive arguments for each option. In the quantitative analysis, there was, overall, a marginal preference for option 1 (mostly driven by industry respondents), but this is not a sufficient basis to make a choice.

We therefore select an option based on the relative merits of each. On balance, we note that option 1 has the benefits of retaining the maximum QALY weighting used for end of life, and ensures some degree of consistency between end of life and the most severe conditions. This is consistent with a pragmatic approach in which the severity modifier is developed using the end of life considerations as a benchmark. We therefore adopt option 1; that is, weights of 1.2 and 1.7 according to the QALY shortfall (table 4).

Table 4. Final severity modifier cut-offs and QALY weights

|  |  |  |
| --- | --- | --- |
| QALY weight | Proportional QALY shortfall | Absolute QALY shortfall |
| 1 | <0.85 | <12 |
| x1.2  | 0.85 to 0.95 | 12 to 18 |
| x1.7 | ≥0.95 | ≥18 |

Implementing the modifier

Stakeholders raised a range of questions about how the proposed modifier would work and be implemented in practice. Detailed discussion of how the methods will be implemented, and the support and activities in progress, will be published in a separate document. These include worked examples, external engagement activities, and training and resources. Monitoring the use of the modifier in practice (see section 42) will allow us to identify and respond to any issues or concerns in implementing the modifier as they arise, and to identify any additional needs for effective and efficient implementation. Key issues are explored below.

Sources of data: population EQ-5D norms and survival

To calculate absolute and proportional QALY shortfall, it is necessary to know the average HR-QoL, by age, in the general population. To do so consistently, it is desirable to specify a consistent preferred source for population HR-QoL data.

The Decision Support Unit (DSU) compared sources of population HR-QoL using NICE’s preferred measure (EQ‑5D) (available on the [DSU website](http://nicedsu.org.uk/)). It compared 3 existing published sources (the Measuring and Valuing Health Study [MVH] and the Health Survey for England [HSE] from 2003 and 2010) and presented 2 additional data sources (HSE from 2014 and the ‘EEPRU dataset’ from Hernández Alava and colleagues that informs our preferred EQ-5D-5L mapping function). In addition, Schneider et al. (2021) have presented a further data source [in a preprint paper](https://www.medrxiv.org/content/10.1101/2021.12.13.21267671v1), using EQ‑5D‑5L data from HSE in 2017 and 2018 (mapped to EQ-5D-3L using the tool by van Hout et al. 2012). They also present a [tool](https://r4scharr.shinyapps.io/shortfall/) for calculating shortfall automatically, which could be a helpful resource for stakeholders.

Importantly, the DSU comparison highlights that absolute and proportional QALY shortfalls are not sensitive to the source of population EQ-5D data (for the 5 sources it compared). This is reassuring, because it suggests that selecting a preferred source is desirable but not essential. It also suggests that there is no need to reconsider the average QALY weights or corresponding shortfall cut-offs and QALY weights on the basis of population EQ-5D norm data source.

The ‘EEPRU dataset’ is immediately appealing because it was collected recently and has a very large sample size. However, the challenges and uncertainties associated with it being collected during a COVID-19 lockdown highlighted by the DSU are important and persuasive. Until further work has been done, we accept the DSU’s proposal that it would be prudent not to use these data for population EQ-5D norms. The 2 most recent HSE data sources are, then, potentially valid options that warrant further consideration. HSE is a large, high-quality, long-running national survey that is designed to cover a representative sample of the population, so is likely to provide a robust data source for population health estimates. Each of the 2 sources has pros and cons. For example, the HSE data from 2017 and 18 is more recent and a larger sample, but it uses EQ-5D-5L mapped to EQ-5D-3L rather than directly observed EQ-5D-3L, and does not use NICE’s preferred methods for mapping from 5L to 3L. Neither has been peer reviewed, and it has not been possible to scrutinise either in detail as part of the methods review. Given that QALY shortfall is not sensitive to the choice of dataset, it is sufficient at this stage to specify in the manual that population EQ-5D norms be based on a recent and robust source but not to mandate a specific source. By that same token, it is not possible to establish at this stage whether the calculation tool presented by Schneider et al. (2021) can be adopted in NICE evaluations.

It may be interesting in the implementation phase to explore further the different sources of population EQ-5D and tools available. Longer-term, it may be valuable for further research to consider population HR-QoL, including the effects of the COVID-19 pandemic as they evolve, and to consider whether and, if so how, future data sources might inform calculation of QALY shortfall in NICE evaluations.

Similarly, absolute and proportional QALY shortfall calculations also need data on the population survival, and it may be desirable to specify a preferred source of data for this. High-quality, up-to-date data on population survival is routinely available from the Office for National Statistics. It is appropriate to use the most up-to-date Office for National Statistics survival data for England in absolute and proportional QALY shortfall calculations. However, COVID-19 has affected population survival statistics from 2020 (both because of COVID‑19 itself and knock-on effects on health services), which may affect shortfall calculations. This may not be representative of future population survival. It is therefore prudent to use the most recent pre-COVID-19 population survival data. If survival data for the technology and/or comparator(s) has been collected during the pandemic, it is likely that the committee would need to consider such effects throughout the evaluation, and this should include consideration of the shortfall calculations. Further research both on post-COVID-19 population survival and on the effect on shortfall estimates may be relevant longer term.

Aligning the use of modifiers across programmes

Some stakeholders expressed concerns that the severity modifier was not proposed to apply outside technology appraisals.

* Highly Specialised Technologies Programme: The consultation paper explained that severity was already ‘implicitly captured’ in this programme. This is to say that, by definition (because of the topic routing criteria), conditions considered in highly specialised technology evaluations are severe. So, the severity modifier is already captured for such topics through the programme’s value framework (including the level at which technologies are considered to provide value for money, of £100,000 per QALY gained). Therefore, it is appropriate that no additional severity modifier is applied.
* Medical Technologies Evaluation Programme: For medical technologies considered through a cost-comparison evaluation (in this programme), a quantitative QALY weight cannot (meaningfully) apply. That is, to be recommended through this route, the technology must provide cost savings, so formally placing additional value on health benefits because the condition is severe would make no difference to whether the technology is recommended. Therefore, the quantitative QALY weight severity modifier cannot apply in medical technology evaluation cost-comparison evaluations. We consider that it remains appropriate to consider severity in committee deliberations. We are working with the committee and stakeholders to explore how this can best be implemented in practice. It is perhaps helpful to emphasise that evaluations of medical technologies and devices that are done as technology appraisals can include the severity modifier as described.
* Diagnostics Assessment Programme: Stakeholders raised concerns about the potential inconsistency generated by not applying the severity modifier to diagnostics evaluations. However, on review, we remain of the view that the severity of the disease measured through QALY shortfall at the time the technology is used is unlikely to reflect societal value for a diagnostic in the same way as for a medicine or other health technology. This is because of how diagnostics sit within a pathway of care, and the effects of population mix (people with and without the condition) and disease trajectories (how many people go on to develop a particular symptom, complication or stage of disease, and when). Therefore, we propose that the severity modifier will not normally apply. This wording leaves some flexibility and hence the potential that it could be applied in some unique circumstances. It is not possible to list or describe all such circumstances upfront, so the wording is retained, and it would be for the committee to deliberate if it considered it appropriate to make an exception.

Further complexities and implementation challenges

Stakeholders raised a range of questions about how the severity modifier should operate in various complex and challenging circumstances. These included:

* how severity should be considered for multiple subgroups, or when there are multiple comparators
* how uncertainty in absolute and proportional shortfall will be considered
* what should be done when EQ-5D is not appropriate for the condition
* how to calculate shortfall for children.

The methods manual sets the fundamental principles on which the severity modifier operates. It is not possible (or indeed desirable) to specifically describe all possible circumstances and scenarios. Rather, it is appropriate that the committee considers each evaluation individually, taking into account the specific circumstances, evidence and context for each topic. Take, for example, subgroups. In common with all evaluations and past practice with end of life, the approach a committee should take for the population as a whole, and the subgroups within that population, must be considered individually for that evaluation. This should include considering factors such as the nature of the condition, population and subgroups, the evidence, current practice, equalities and fairness, and so on. For the severity modifier, there may be some evaluations in which there is a subgroup with a greater severity than the population as a whole, such that the modifier would only apply to that subgroup. In that case, it may be appropriate for a committee to consider that subgroup separately with the modifier applied. But it would need to take into account all relevant subgroup considerations (as defined in the methods) to do so. We are continuing to explore the operation of the severity modifier in complex circumstances during the implementation phase for the methods, for example, using worked examples and other activities.

Other complexities raised in consultation are addressed in the final methods manual. These include:

* Discounting: absolute and proportional shortfall should use discounting at the reference-case discount rate.
* Time horizon: QALY shortfall is defined as the amount of health lost over a person’s remaining lifetime. Therefore, a lifetime time horizon is likely to be needed to establish QALY shortfall and hence apply a severity modifier.
* Carers: QALY shortfall is defined as the amount of health lost by a person with a condition; other people, such as carers, should not be included. When relevant, the economic evaluation may include the health benefits of a technology for carers as described in the manual.

Future steps with the severity modifier

Because there is not enough evidence to define the magnitude of societal value for health benefits in severe diseases (see section 9), we identified a research priority to address this evidence gap. When available, such research would be highly valuable for reviewing, and if appropriate refining or changing, the severity modifier – potentially through a modular methods update. We will explore in future documents the key considerations for designing and completing this research, so that it provides the best possible evidence to inform future improvements to our methods. We note stakeholders’ interest in this topic, and will welcome their involvement and participation. We also need to actively monitor, scrutinise and learn from the use of the severity modifier in practice. This will allow us to gather evidence on measures of severity (absolute and proportional QALY shortfall) in real evaluations, and alignment between our modelled predictions and practice. Such monitoring will also allow us to identify any additional needs for effective implementation of the modifier in practice.

Uncertainty

The consultation paper proposed broad improvements and clarifications to the consideration of uncertainty in our evaluations. This included clarification of the additional flexibility in considering uncertainty when evidence generation is difficult.

There was broad support among stakeholders for the proposals.

Some questions were raised about how the uncertainty considerations would be implemented in decision making. These issues are explored in detail in the implementation programme, including through worked examples, training and resources. We acknowledge the importance of ensuring the methods for uncertainty are robustly and consistently used by committees and are transparently reported. Going forward, continuing monitoring and review of uncertainty considerations in evaluations may be valuable.

In particular, some stakeholders requested additional clarity or precise definitions for the circumstances in which the additional flexibility can apply. The unique circumstances of each evaluation mean that it is not possible or desirable for a precise, prescriptive definition to be made in the methods guidance; rather, this is a deliberative decision for committees in each evaluation. Moreover, this illustrates that the draft manual was not fully clear. The acceptance of greater uncertainty was not intended to be simply applied in an all-or-nothing fashion if a particular criterion is met. Rather, it is a deliberative decision for committees, to accept uncertainties based on an understanding of how those uncertainties have arisen and how they relate to challenges in generating evidence and the circumstances in which such challenges occur. Additional text has been added to the manual to address and clarify this issue: the committee will consider how the nature of the condition or technology(s) impacts the ability to generate high-quality evidence before applying greater flexibility. In this way, we ensure there remains appropriate and proportionate consideration of uncertainty, managing the risks to patients and the NHS while preventing inappropriate barriers to valuable innovations.

Stakeholders proposed that unmet need should be considered in the provisions for uncertainty. Unmet need is routinely considered as part of committees’ normal deliberations, and is explicitly captured in the severity of the condition. Therefore, additional consideration of unmet need through uncertainty is not needed.

Some questions were raised about the implementation of uncertainty considerations for diagnostics. The manual emphasises that uncertainties will be considered appropriately and proportionately, in the evaluation context (including the nature of the evidence and technology). Additional concerns were raised about the use of budget impact analyses to assess the scale of decision risk, when such analyses are not always available in an appropriate form for diagnostics and other medical technologies. The manual has been amended to allow flexibility in the evidence requirements to establish the scale of decision risk.

As part of the refinement and alignment of the manual, questions have been raised about how the additional flexibility for uncertainty when evidence generation is difficult applies specifically in the context of managed access in technology appraisals and highly specialised technologies evaluations. To address this, it is important to emphasise the sequence of decision making for managed access:

* The committee first considers whether the technology should be recommended for routine commissioning. In this decision, the committee considers whether the uncertainties in the evidence are acceptable for a routine decision (including whether those uncertainties have arisen because of difficulties in generating evidence).
* If the committee decides that the technology cannot be recommended for routine commissioning, it may then consider a recommendation for managed access. If so, at this point the additional flexibility becomes moot, because the decision focuses on the potential to address uncertainties through evidence collection (as defined in the manual) and the plausible potential for cost effectiveness, and not on the overall acceptability of the uncertainty in a routine recommendation.
* After a period of managed access, the committee will reconsider the technology for routine commissioning. This may include consideration of whether evidence generation difficulties have contributed to uncertainties, although the committee will be aware that a period of managed access has been completed.

Stakeholders also commented on issue of visualisation tools for uncertainty to support decision making. Several stakeholders considered this to be a priority activity. We agree this could be a valuable resource. Although it does not need to be specified in the methods manual, it is appropriate to prioritise this as an implementation and ongoing development activity. We have developed further ideas and scheduled additional engagement events to develop and refine this approach, with a view to implementing (or at least piloting) uncertainty visualisation approaches.

Health inequalities

The consultation identified that there remains a case to include consideration of whether recommending a technology can reduce health inequalities. The existing methods, NICE’s statutory duties, the NICE Principles and routine deliberative decision making, combined, provide the flexibility to take into account relevant considerations for individual evaluations. This may include health inequalities if it arises. However, further work would be needed to implement a formalised modifier. Addressing health inequalities remains an important priority, and a broad and complex area that will need ongoing consideration across NICE, as described in our Principles and NICE Strategy.

Stakeholders generally supported the proposals presented and recognised the further work that is needed before a formalised modifier could be implemented. As a result, no additions to the methods manual have been made. Because of the importance of health inequalities across NICE (described in [The NICE strategy 2021 to 2026](https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026)) and the ongoing institute-wide work on this topic (including, for example, through the [NICE Listens programme](https://www.nice.org.uk/get-involved/nice-listens)), it may be appropriate to prioritise health inequalities in future modular updates.

Several stakeholders specifically raised rare diseases in the context of health inequalities, and argued that there was a case to reconsider adding a modifier for rare diseases. The first phase of the methods review emphasised that health benefits are of equal value (regardless of other characteristics of the technology and people having those benefits), except in exceptional circumstances. It also emphasised that such exceptional circumstances would need to be underpinned by a clear evidential or normative case. The evidence review at that stage found that there was no evidence that society values more highly health benefits in rare diseases. The information provided in the consultation responses did not provide usable or robust evidence of societal preferences for placing additional value on health benefits in rare diseases and displacing care elsewhere in the system, and does not provide sufficient evidence to support adding a modifier. As a result, no changes are needed. Other elements of the methods review are relevant to rare diseases, including the severity modifier, flexibility in uncertainty considerations and the emphasis on a comprehensive evidence base including real-world, qualitative, surrogate and expert evidence. These updates mitigate the barriers faced by rare disease technologies while maintaining an evidence-based, robust and proportionate evaluation approach.

There may be a case to explore whether future modular updates on topics relevant to rare diseases would be useful. More broadly, it may be interesting to explore whether rarity of disease is best considered as a component of health inequality, or is better considered separately. We note stakeholders’ comments that people with rare diseases have lower health outcomes than those with more common conditions. The rarity of a disease differs from several other determinants of health inequality, such as socioeconomic status, in that it is a specific facet of the disease in question. Moreover, the health outcomes associated with rare diseases are influenced by complex interactions between the severity of the diseases, current diagnostic and treatment options, clinical knowledge, research and development, and health service design and delivery. Many of these factors are captured in specific modifiers and decision-making considerations. So, this question remains complex and challenging, and beyond the scope of the current review. It seems appropriate to consider whether this topic can be explored and, if so how, in future methods considerations and broader health inequalities work.

Discounting

In the previous phase of the review, we maintained our view that there is an evidence-based case for changing the reference-case discount rate to 1.5% for costs and health effects. However, because of the wider policy and fiscal implications and interdependencies that are beyond the reach of this review, no change to the reference-case discount rate was proposed.

The consultation highlighted substantial disagreement with the proposals, particularly from the pharmaceutical industry and patient groups. Concerns included:

* the alignment with the evidence case and the HM Treasury Green Book
* links to policy developments
* the effects of maintaining the current discount rate on life sciences and innovation, specific innovations such as advanced therapy medicinal products (ATMPs), and on patients and patient access to health technologies.

In contrast, academic stakeholders, NHS organisations and NICE committee members were more supportive of retaining the current discount rate. They acknowledged the need for further work, consultation, and coordination with other health bodies to understand the impact on the overall healthcare system. We note stakeholders’ concerns. We maintain our view that there is a good case for changing the reference-case discount rate, and also recognise respondents’ views on the need for further work and coordination with the healthcare system.

Some stakeholders queried the policy and financial implications and interdependencies highlighted in the consultation. These issues were explored in detail during the first and second stages of the methods review. Some of the key issues are explored again in the following paragraphs.

* One of the most visible system implications of a change in discount rate for NICE health technology evaluations is the financial impact. Analyses in stage 1 of this review estimated that the impact would potentially be very large. Reducing the discount rate will make most technologies appear to be more cost effective (the effect is particularly large in technologies with upfront costs and long-term benefits, such as ATMPs; see section 58). In the absence of a change in the level at which technologies are considered cost effective, this would likely lead to higher prices for new health technologies. These additional costs would then impact on the healthcare system and displace care and services from other patients. Given NICE’s responsibilities to the NHS and patients (to consider the broad balance between the benefits and costs of providing health services or social care in England) and the tight funding and resource constraints facing the NHS (perhaps never more so as we respond to the COVID-19 pandemic), this impact must be taken into consideration.
* Several stakeholders noted the effect of the 2019 Voluntary Scheme in capping expenditure on branded medicines. They suggested that this would allow a lower discount rate to be used without affecting NHS budgets. This overlaps with similar arguments made about the severity modifier. Again, the 2019 Voluntary Scheme does not remove the need for robust evaluations nor eliminate displacement effects from additional spending on medicines. The financial effects of a lower discount rate would extend beyond the life of the current 2019 Voluntary Scheme. The 2019 Voluntary Scheme affordability mechanism operates based on a projected level of growth but the NHS has a statutory duty to operate within its overall financial envelope each year and it cannot defer payments to future years to compensate. So, increases in the prices of new medicines can still lead to displacement of care and services elsewhere in the NHS, even if the effects are later mitigated through other mechanisms. Moreover, many technologies considered in NICE health technology evaluations that use the reference-case discount rate are not covered by the 2019 Voluntary Scheme (including medical technologies with long time horizons, diagnostics and medicines that are outside the scheme). Therefore, although the 2019 Voluntary Scheme plays a valuable role in the healthcare system, the financial impact of the discount rate remains an important consideration for the methods review.
* Another important system-wide consideration is that of consistency between different parts of the health technology evaluation system, and in how different health technologies and health interventions are evaluated. Changing the discount rate for NICE health technology evaluations would create inconsistency, for example, between medicines evaluated by NICE and vaccines evaluated by Joint Committee for Vaccination and Immunisation (which uses a 3.5% rate), and between technologies considered in the NICE health technology evaluation programmes and treatments considered in the NICE Guidelines programmes. As previously explained, such interactions and interdependencies go beyond the reach of this methods review alone. For example, it would be inappropriate to introduce inconsistencies without system-wide discussion.

Stakeholders noted the importance of the discount rate for innovative technologies such as ATMPs and, by extension, rare diseases. This was explored in the stage 1 task and finish group report. No additional information was identified in the consultation. We note that the effect of discounting on ATMPs does not in itself indicate a change in discount rate is justified and that, to date, almost all ATMPs considered by NICE have been recommended using the current methods. Moreover, measures across the healthcare system, such as the Innovative Licensing and Access Pathway (ILAP) and the Accelerated Access Collaborative (AAC), combine with provisions elsewhere in the NICE methods (for example, uncertainty considerations and support for a comprehensive evidence base) to ensure the NHS welcomes valuable innovations and patients continue to benefit from them.

It was suggested that we could apply a 1.5% discount rate to selected technologies or circumstances, perhaps as a ‘pilot’ for changing the reference-case rate. Such an approach would carry the same policy and system implications and interdependencies as a change in the reference-case rate. This includes the financial implications, particularly if the selected technologies or circumstances were those in which the discount rate has a large effect (such as ATMPs). Therefore, a selected or pilot application of a 1.5% discount rate is not appropriate.

Overall, having explored the consultation comments in detail, we conclude that no new issues have been identified that were not previously considered, and therefore there is no case to depart from the previously established position. We recognise stakeholders’ concerns, and maintain our view that there is a case for changing the reference-case discount rate to 1.5%. However, we also maintain our view that this change in the reference-case discount rate cannot be implemented because of the system implications and interdependencies.

NICE will contribute to system-wide and policy discussions about discounting through established channels and ongoing collaborations with stakeholders and health system partners. To support our contribution to such discussions, we will prospectively collect evidence on the effects of discounting (for example, on ICERs and prices) in evaluations. This will be done as evidence collection to inform methods research, and as such must be separate from the evaluations themselves, but can use the models. When possible and appropriate, we will work with stakeholders to support this evidence collection.

In addition, stakeholders requested additional clarification on how the non-reference-case exception for the discount rate should be applied. On review, it is not possible to provide further, more definitive guidance on this situation because it strongly depends on the circumstances in individual evaluations and the deliberations by the committee. Overall, no new information was identified in the consultation, and no changes have been made.

Understanding and improving the evidence base

Broad improvements to the methods for sourcing, synthesising and presenting evidence were incorporated into the draft manual in line with the cases for change identified in the first phase of the review. This included, among others, improvements to evidence on HR-QoL and the costs of introducing health technologies, and further guidance on the use of real-world evidence.

Real-world evidence

Stakeholders welcomed the proposed additions to support greater use of real-world evidence in evaluations. They noted in particular the benefits for rare disease treatments and medical devices. Stakeholders sought reassurance on the appropriate use of real-world evidence by committees, and further guidance on different types, uses and qualities of real-world evidence.

Most of these queries relate to the implementation of the methods guidance, so do not trigger changes to the draft manual content. However, there was a need to refine the balance of the text on real-world evidence. This was to create a better balance both between prospective data collection and evaluation of existing data, and between different types of technology, contexts and use cases for real-world evidence (for example, medicines, devices, diagnostics). Updates to the manual have been made accordingly.

Real-world evidence remains an active area of interest for NICE, with significant work ongoing across the institute. In particular, the real-world evidence framework and associated activities will provide valuable additional guidance and continuing improvement in our use of such evidence. We have therefore adjusted the manual to direct readers to this work, avoiding unnecessary duplication in the manual and supporting stakeholders in accessing the most complete and up-to-date resources.

Health-related quality of life

The proposed hierarchy of alternative measures of HR-QoL (when EQ-5D is not suitable) was accepted. But stakeholders queried the evidence requirements to use alternative measures and the choice of measures in the hierarchy. We did not identify sufficient evidence to reconsider the preference for other (non-EQ-5D) generic measures of HR-QoL as the first step in the hierarchy. The methods correctly set clear evidence requirements for showing the EQ-5D is not appropriate. This is because of the importance of consistency in measuring HR-QoL across evaluations and the broad evidence of the suitability of EQ-5D in most cases. Nevertheless, we acknowledge that such evidence may be challenging to get in some circumstances, and clarification of the flexibility in the evidence requirements has been added. Methods for assessing the appropriateness of EQ-5D are documented in the manual and widely discussed in published literature. If further information is needed in future this may be considered as part of implementation and continuous improvement work. It may be appropriate to monitor use of the updated methods in practice to identify any additional challenges that may arise.

We note stakeholders comments on the importance of further refinement of the methods for health-related quality in children and young people and in carers. Both will be explored in future methods projects. No changes are made at this stage.

To support the change in preferred mapping algorithm for EQ-5D-5L, additional resources will be published by the DSU. It is noted that the limitations in the dataset that informs the preferred mapping algorithm identified while exploring the severity modifier (in particular, the effects of collecting the data during the COVID-19 pandemic) do not affect its validity for informing the mapping function. So, no change to the preferred mapping methods is needed.

Discussion of population HR-QoL data for the QALY shortfall calculations raises a similar issue for the adjustment of utilities that are extrapolated over long periods, to account for age and comorbidities. That is, it raises the question of which data source should be used to inform the age adjustment. Here, as for the shortfall calculations, we conclude that it is appropriate to specify that adjustment should be based on a recent and robust source of population HR-QoL. But we do not mandate a specific source (although we note that it is prudent to suggest that the ‘EEPRU dataset’ is unlikely to be suitable because of the effect of the COVID pandemic).

Surrogate outcomes

Stakeholders commented on the evidence requirements for surrogate outcomes. We consider that the manual provides appropriate detail on the evidence requirements, extrapolation between populations, and flexibility when the highest quality evidence may be unachievable. No further changes are needed. This might be a topic that could be considered for further activity (for example, training) in the future if needed.

Costs

Several issues relating to the costs of health technologies need additional consideration:

* using discounted prices for medicines other than patient access schemes (non-PAS prices)
* technologies that are not cost-effective at zero or low cost
* apportioning or adjusting costs in defined circumstances
* other issues such as VAT and non-NHS costs.

Non-PAS prices for medicines

Stakeholders argued that, when using non-PAS discounts, the prices should be nationally available, stable and agreed for the duration of the appraisal. They noted that list prices may be appropriate when no discounts are available, and sought guidance on using confidential discounts.

It is important to reemphasise the critical principle, that the prices used in a health technology evaluation should reflect as closely as possible the prices that are paid in the NHS. We acknowledge that the term ‘nationally available’ (and our use of it in the draft manual) is potentially ambiguous and confusing. This term might be understood to refer to, for example, discounts that are negotiated at a national level, discounts that are available to everyone across the country, or to a product with various discounts across the country (particularly if everyone would be able to access a discount of at least a given level). Commercial Medicines Unit (CMU) prices are known to differ between regions, but they are still potentially valid prices to consider in evaluations. We have therefore removed the term ‘nationally available’ and refined the wording to avoid confusion. It is not appropriate to include a requirement that only discounts that are available to everyone in the country be considered. This is because this would not accurately reflect the price paid by the NHS when discounts vary between regions. Similarly, it is not appropriate to require prices to be stable and agreed for a particular period. But the duration of pricing arrangements should be considered by the committee (as specified in the draft manual). We acknowledge the omission of list prices as an appropriate source of prices when no discounts are available; this has been corrected. We recognise the transparency limitations of using confidential prices, but this reflects the reality of the healthcare market. NICE must respect and maintain confidentiality of negotiated agreements.

Some stakeholders queried the use of the highest, lowest and midpoint CMU prices, and some argued that weighted averages should be used. Although there is some appeal to a more sophisticated analysis of pricing and variability, it is unlikely to feasible in practice. Therefore, using both the lowest and highest values shows the range of available scenarios (the ‘best case’ and ‘worst case’ scenario) from which the committee can understand the sensitivity of the results to CMU price and the level of uncertainty and risk associated with the decision. The midpoint is a purely pragmatic scenario to use when testing other assumptions, variables and scenarios in the model.

Technologies that are not cost effective at low or zero cost

Stakeholders raised conflicting concerns about the proposed accommodations for non-reference-case analyses when a technology extends survival in populations for which the NHS is providing high-cost care. Some stakeholders expressed concerns that non-reference-case analyses would not be used in practice. Others were concerned that the permitted circumstances were too broad and would overly favour specific technologies at the cost of healthcare displaced elsewhere in the NHS. The concerns raised indicate that further consideration is needed.

We acknowledge an important limitation with the proposed wording, that it may imply a technology could be considered using a base-case analysis in which all background care costs removed. If this were considered by a committee as the only decision-making scenario, it would risk exposing the NHS to high levels of opportunity costs, and this was not the intention. Rather, it is more appropriate for committees to consider the cost effectiveness of a technology both with all costs included (in line with the reference case) and with background removed (in line with the non-reference case accommodation). We propose amendments to the manual to make clear that non-reference-case analyses should be presented as additional analysis alongside the reference case, and that committees should include both. The wording has also been strengthened to support implementation of the non-reference-case analyses, including details of what factors may go into the consideration of reference-case and non-reference-case scenarios. Given the complexity and unique circumstances in individual appraisals, the manual cannot be directive or exhaustive. This remains an area for committee deliberation and judgement.

Industry respondents explored other methodological options for this scenario, such as a modifier or higher cost-effectiveness threshold or attributing value between components of a combination regimen. This methods review has not identified an evidential or normative case to support applying additional weight to benefits in this circumstance. NICE has been clear throughout that commercial issues for combination technologies are outside the scope of the methods review. The core role of health technology evaluation decision is to establish whether a combination regimen as a whole represents an effective use of resources, and not the share of the cost received by different companies, which is a commercial issue.

Apportioning costs

The concerns raised about technologies that are not cost effective at low or zero cost also highlight parallel issues for circumstances in which costs may be adjusted or apportioned in non-reference-case analyses. Stakeholders expressed concern that non-reference-case analyses are not frequently accepted by committees. They argued that apportioned costs should be considered in the reference case. This is not appropriate because it departs from the principle that, in the reference case, the costs of introducing a health technology should be included in full and reflecting as accurately as possible costs to the NHS. Nevertheless, given the similarities between these non-reference case scenarios, it is reasonable to adopt a similar approach. That is, that both reference-case and non-reference-case analyses should be taken into account, with strengthened wording to support committees in implementing the approach. Changes have been made in the manual accordingly.

Other cost issues

Stakeholders queried the inclusion of VAT in budget impact analyses but exclusion from economic analyses. This has not changed from previous methods guides, and was not in the scope of the current review. If there is a need to reconsider this aspect of the methods, it would need to be done through a future modular update. If so, particular consideration may be needed for cost comparison analyses.