Board meeting

25 September 2024

Review of the implementation of the severity modifier

Purpose of paper

For approval

Board action required

The Board is asked to:

* approve the recommendation that no change to the severity modifier is required at this time and to continue monitoring the mean weighting applied to technology appraisal decisions.
* approve the progression of further research on societal preferences in relation to severity.

Brief summary

The severity modifier was introduced in 2022. It was designed to be opportunity cost neutral in relation to the end of life modifier that it replaced, and to apply to a broader range of conditions than had benefited from the end of life modifier. This paper assesses whether the severity modifier is operating as intended.

Board sponsor

Dr Nick Crabb, Chief Scientific Officer

Helen Knight, Director of Medicines Evaluation

Background

The severity modifier was introduced in 2022 in order to place greater value on health benefits provided for people with more severe diseases. It replaced a previous “end of life” modifier, where providing additional weeks or months at the end of life for people with terminal diseases was valued more highly. The change to the severity modifier was designed to benefit a broader range of diseases and better reflect the available evidence on society’s preferences.

Responses to 2 public consultations showed stakeholders, including the life sciences industry, broadly supported the concept of a severity modifier. This was because they considered it to be fairer than the previous end of life modifier as it would apply to more severe disease areas outside of terminal cancer.

The change to the severity modifier had to remain opportunity cost neutral. This is important because a larger severity modifier would have displaced more care, and therefore more health benefits, for patients elsewhere in the NHS. This is a key principle NICE works under and any change to NICE methods that were cost inflationary would have to be approved by the Department of Health and Social Care.

Two levels of severity modifier weightings were introduced, 1.2 that values the health benefit at 20% higher than standard and 1.7 that values the health benefit at 70% higher than standard. Further information on the weighting profile is provided in the technical appendix.

Although there is evidence that society highly values health benefits in more severe conditions, further research is needed on society’s preferences on how much additional weighting to apply to health benefits for people with severe diseases.

Aim of this review

The aim of this review is to assess whether the severity modifier is operating as intended, including whether it has been opportunity cost neutral compared to the end of life modifier it replaced.

Findings

The severity modifier is operating as intended. It has been applied more widely than the end of life modifier it replaced and is resulting in a higher proportion of positive recommendations than end of life. In addition, both levels of severity (the 1.2 and 1.7 weightings) are being used and it has also contributed to positive recommendations for non-cancer as well as cancer medicines.

A severity weighting has been applied in 35% of all decisions, compared to 18% under end of life.

The proportion of positive recommendation decisions when additional severity weighting was applied (84.4%) is higher than when additional end of life weighting was applied (82.7%). The overall proportion of positive recommendation decisions, regardless of whether additional weighting was applied, has been higher since the severity modifier was introduced (83.5%) than the historical level when the end of life modifier was in place (82.5%).

The severity modifier has been applied to diseases including non-end-of-life cancers as well as non-cancer conditions that have dramatic and far-reaching impacts on patients, such as cystic fibrosis and chronic hepatitis D. Most of these would not have qualified for additional weighting under the end of life modifier.

Positive decisions involve treatments for cancer, including glioma and cancer of the breast, lung, bowel and stomach.

Since the introduction of the severity modifier:

* 98% of the decisions that did not attract additional weighting under the severity modifier wouldn’t have been eligible for additional weighting under the end of life criteria either.
* Only 1 of the 22 decisions estimated to have met the end of life criteria (5%) did not achieve additional weighting under the severity modifier. On the other hand, 11 of the 69 decisions estimated not to have met the end of life criteria (16%) achieved additional weighting under the severity modifier, 3 of which achieved a weight of 1.7.

In the majority of cases where the severity modifier was applied and where the end of life criteria were judged also to have been met, a weight of 1.2 was used (14 out of 21 decisions; 67%). For the remaining 7 decisions (33%), the 1.7 severity weighting given was the same as the maximum that would have been applied under end of life.

Under the severity modifier, the proportion of positive recommendation decisions relating to treatments for cancer has been greater than had been the case under the end of life modifier. The same is true when looking specifically at treatments for metastatic and/or advanced cancer.

Figure 1 shows how the mean weighting applied to technology appraisal decisions has changed over time since the introduction of the severity modifier. The mean weighting is calculated for each individual quarter (blue line), as well as cumulatively, where the mean weight for each quarter also includes any decisions published in earlier quarters (orange line). These are compared to the mean weights calculated for the historical dataset, applying either the severity modifier (solid black line) or the end of life modifier (dashed black line).

Figure 1: Mean weight over time, tracked on a quarterly basis

The mean weighting applied to technology appraisal decisions since the introduction of the severity modifier shows high variability when tracked on a quarterly basis, making it difficult to spot significant changes and trends.

Overall, the mean weighting applied to technology appraisal decisions since the implementation of the severity modifier has been very similar to the historical mean weight where the end of life modifier was applied (see the cumulative mean weightings in Figure 1. This indicates that opportunity cost neutrality has been achieved. Given the high variability, we plan to continue to monitor the mean weighting applied to technology appraisal decisions. If the cumulative mean weighting falls below 1.10 for 2 consecutive quarters, we would investigate the causes and identify whether corrective action was needed.

NICE is aware that the Association of the British Pharmaceutical Industry (ABPI) has also been monitoring the implementation of the severity modifier as part of its Continuous NICE Implementation Evaluation (CONNIE) initiative. CONNIE findings vary from those reported here. Through helpful cooperation from the ABPI, we understand and can reconcile the differences and consider the conclusions on opportunity cost neutrality compared to the end of life modifier to be sound. NICE’s analyses are based on a complete dataset and an appropriate unit of analysis. This is detailed in the technical appendix.

Clearly, the weighting applied to most products that would have met the end of life criteria is lower under the severity modifier. This is consistent with the design of the severity modifier – where the weighting is spread over a broader range of products addressing severe conditions whilst retaining opportunity cost neutrality. However the data shows this did not affect positive recommendations. Of 22 decisions under the severity modifier where it was judged that the previous end of life criteria would likely have been met, 20 decisions (91%) resulted in positive recommendations.

Conclusion based on this review

The severity modifier is operating as intended and has been applied to a wider range of diseases as it was designed to do. Based on the data to date, it has remained opportunity cost neutral compared to the end of life modifier and the proportion of positive recommendations has been higher.

We therefore recommend that no change to the severity modifier is required and that we should continue monitoring the mean weighting applied to technology appraisal decisions.

Further research is being commissioned to understand in greater depth the extent to which society places additional value on treatments for severe diseases.

Research on societal preferences

NICE has begun scoping further research into society’s preferences on how much additional weighting to apply to health benefits for people with severe diseases.

We intend to tender for this work in December and it will report back around 2 years after field work begins. The findings will support future further consideration of the severity modifier.

Review of the implementation of the severity modifier

Technical appendix

Background

Development of the severity modifier

NICE introduced the severity modifier in January 2022, following a review of the methods, processes and topic selection for health technology evaluations (‘methods review’ hereafter), including 2 stakeholder consultations.

The methods review found limited evidence that society places additional value on life-extending treatments at the end of life compared to other types of treatments. By contrast, greater evidence was found that society highly values health benefits in severe conditions, including conditions characterised by poor quality of life.

NICE therefore proposed replacing the end of life modifier with a new modifier based on severity and quality-adjusted life year (QALY) losses, using absolute and proportional QALY shortfall to quantify severity (AS and PS, respectively). This proposal was broadly supported by stakeholder consultees, including life sciences industry stakeholders.

When designing the modifier, an ‘opportunity cost neutral’ approach was taken whereby the weights applied to incremental QALYs previously invested in end of life treatments were reallocated to treatments for severe disease. This was considered important because a larger severity modifier would have displaced more technologies and care, and therefore more health benefits, for patients elsewhere in the NHS.

Opportunity cost neutrality is a key principle that NICE works under. NICE cannot introduce methods changes that would be cost inflationary without agreement from the Department of Health and Social Care. Further, any departure from opportunity cost neutrality should ideally be justified by evidence that society supports such reallocation of health.

A weighting profile was established (Table 1), calculated based on data from a retrospective review of health technology appraisal decisions. This included 3 different QALY weights (1, 1.2 and 1.7), applied based on AS and PS, whichever implies the greater severity level.

Table 1: QALY weightings for severity

|  |  |  |
| --- | --- | --- |
| QALY weight | Proportional QALY shortfall | Absolute QALY shortfall |
| 1 | Less than 0.85 | Less than 12 |
| x1.2 | 0.85 to 0.95 | 12 to 18 |
| x1.7 | At least 0.95 | At least 18 |

Source: NICE health technology evaluations: the manual

During the development of the severity modifier, NICE identified that there would be a small number of topics that had previously met the end of life criteria but would not have received additional weighting with the severity modifier. However, the retrospective review showed that the application of the severity modifier would span a broader range of conditions than had benefited from the end of life modifier, including non-end of life cancer indications, and musculoskeletal, inflammatory and mental health conditions. It would not have been possible to design the severity modifier in a way that did not disadvantage any topics without it being cost inflationary.

The principle of exceptionality was emphasised throughout the methods review process. NICE’s position was that health benefits are of equal value (regardless of the characteristics of the technology and the people having those benefits), apart from in very exceptional circumstances. As was the case for the end of life modifier, the severity modifier was intended to be applied as an exception.

Initial data on frequency of use

Initial data on experience of using the new manual was reported to the NICE Board in December 2023. High level statistics on the use of the severity modifier were based on a sample of 52 appraisals (all topics that had been to at least 1 committee meeting and reached either published final or draft guidance stage). The Board paper noted that wider use of the 1.2 severity weighting had been anticipated, and that this would be the subject of a review in 2024-25.

Aim of this review

The aim of this review is to assess whether the severity modifier is operating as intended, including whether it has been opportunity cost neutral compared to the end of life modifier it replaced.

Commissioned DSU analysis

NICE commissioned the Decision Support Unit (DSU) to undertake an in-depth analysis of the use of the severity modifier in technology appraisals. NICE reviewed the draft DSU report and requested amendments and additional analysis as needed.

Full explanation of the DSU’s methods and findings, including statistical analysis methods, can be found in a [severity weights report on the DSU website](https://www.sheffield.ac.uk/nice-dsu/methods-development/severity-weights).

In brief, the DSU analysis was based on:

* A historical sample comprising 464 decisions from 269 technology appraisals published between January 2009 and March 2021 for which severity weightings could be estimated (decisions from appraisals that were conducted using cost comparison analyses rather than cost utility analyses would not include the necessary information, for example). This is referred to as the primary+ subsample in both the DSU report and this paper. Among these decisions were 423 that were based on health and cost outcomes over a lifetime horizon.

A ‘new methods’ sample comprising 68 decisions from 47 technology appraisals published between January 2022 and March 2024, excluding cost comparisons and decisions where severity was not calculated. This is referred to as the implementation subsample in both the DSU report and this paper. All 68 decisions were based on health and cost outcomes over a lifetime horizon.

Note that technology appraisal decisions differ from technology appraisal topics since one topic can include multiple decisions, for example in multiple technology appraisals where each decision relates to a different treatment or combination therapy option, or when there are different patient subgroups or comparator options that are considered separately.

Estimated and actual use of the severity modifier

The DSU analysis found minimal differences between estimated and actual use of the severity modifier.

The proportion of decisions meeting the criteria for a 1.2 weight in the implementation subsample was lower than the proportion estimated in the primary+ subsample, while for a 1.7 weight the proportion was higher. However, the differences are not statistically significant.

The differences in the application of the severity modifier across the primary+ and implementation subsamples vary depending on the type of QALY shortfall being measured – AS or PS – and the weight being applied.

The overall mean severity weight observed across the implementation subsample was 1.103. This compares to estimated mean weights of:

* 1.119 for the original historical dataset based on which the severity modifier was initially designed, covering April 2011 to November 2019 (described as the primary subsample)
* 1.116 for the extended dataset, covering January 2009 to March 2021 (the primary+ subsample)

Comparing the severity and end of life modifiers

The DSU analysis found a high level of consistency between eligibility for the severity modifier and eligibility for the end of life modifier.

The applicability of the end of life criteria to the 68 decisions that were published after implementation of the severity modifier in January 2022 was estimated by NICE staff. The report describes several caveats about the data underpinning the comparison of the severity and end of life modifiers.

Decisions considered to qualify for additional weighting under the end of life modifier almost always achieved additional severity weighting.

Summary of the DSU report findings

Table 2 summarises key findings of the DSU quantitative analyses.

Table 2: Summary of the DSU report findings

| Research question | Key findings |
| --- | --- |
| Are there significant differences between expected and actual use of severity modifier? | The differences in the overall proportions of decisions in which a 1.2 or 1.7 weight was received are not statistically significant. |
| How does the severity modifier compare to the end of life modifier? | A high level of consistency between the severity modifier and the end of life modifier can be observed. As was expected, the majority of decisions qualifying for the severity modifier achieved a QALY weight of 1.2, which is lower than the equivalent weight had the end of life criteria been applied (1.7). |
| What is the impact of different QALY shortfall calculation methods on the use of severity modifier? | No substantial impact was identified from differences in calculation methods. |
| Is there a change in the distribution of weights for severity over time? | No clear pattern was identified in the decisions published from 2011 to 2020 but there is substantial variation across years. |
| What is the relationship between age and absolute and proportional shortfall? | Age is strongly related to absolute shortfall. The relationship with proportional shortfall is more complex. |

Additional analyses and further considerations

Updated results using extended implementation subsample

The commissioned DSU analysis included decisions published between January 2022 and March 2024 in its implementation subsample, reflecting the data available at the point of commissioning the work. We extended this dataset to include decisions published up until the end of July 2024, referred to hereafter as the implementation+ subsample. Analyses of the implementation+ subsample were undertaken by NICE and validated by the DSU.

The implementation+ subsample contained 91 decisions across 62 appraisals. It excluded cost comparisons and any decisions where AS and PS information was not available.

Decisions in the implementation+ subsample received a mean severity weight of 1.125. This compares to a mean severity weight of 1.103 for decisions in the implementation subsample, and 1.116 in the primary+ subsample. Table 3 compares the 3 subsamples in terms of mean weights and distributions across AS, PS and overall weighting categories. It is an adapted version of Table 2 in the DSU report.

Table 3: Distribution of AS/PS and combined severity weighting categories in technology appraisal decisions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Primary+ | Implementation | Implementation+ |
|  | n | 464 | 68 | 91 |
| Absolute shortfall | | | | |
|  | Mean | 9.31 | 7.79 | 8.62 |
| Cat 1  (AS<12) | % | 73.7 | 94.1 | 84.6 |
| Cat 1.2 (12≤AS<18) | % | 24.4 | 5.9 | 12.1 |
| Cat 1.7  (AS≥18) | % | 1.9 | 0.0 | 3.3 |
| Proportional shortfall | | | | |
|  | Mean | 0.61 | 0.62 | 0.64 |
| Cat 1  (PS<0.85) | % | 71.6 | 72.1 | 72.5 |
| Cat 1.2 (0.85≤PS<0.95) | % | 21.6 | 19.1 | 19.8 |
| Cat 1.7 (PS≥0.95) | % | 6.9 | 8.8 | 7.7 |
| Overall | | | | |
|  | Mean weight | 1.116 | 1.103 | 1.125 |
| Cat 1 | % | 62.5 | 70.6 | 64.8 |
| Cat 1.2 | % | 29.3 | 20.6 | 24.2 |
| Cat 1.7 | % | 8.2 | 8.8 | 11.0 |

The DSU analysis exploring severity weighting categories according to whether the end of life criteria were (or would have been) met was updated using the implementation+ subsample. As with the implementation subsample, this was based on the judgement of NICE staff retrospectively considering the available data. Results of this analysis are presented in Table 4, which is an adapted version of Table 5 in the DSU report. For 1 decision in the primary+ subsample, the committee did not consider eligibility for the end of life modifier (TA192), leaving 463 decisions.

Results using this extended dataset were consistent with those reported by the DSU and did not change the conclusions from the data.

Table 4: Severity weighting categories for technology appraisal decisions according to meeting the end of life criteria

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Primary+ (n=464) | | | | Implementation+ (n=91) | | | |
| EoL Yes | | EoL No | | EoL Yes | | EoL No | |
|  | n | % | n | % | n | % | n | % |
| Total | 81 | 17.5 | 382 | 82.3 | 22 | 24.2 | 69 | 75.8 |
| **Absolute shortfall** | | | | | | | | |
| Cat 1 (AS<12) | 43 | 53.1 | 299 | 78.3 | 17 | 77.3 | 60 | 87.0 |
| Cat 1.2 (12≤AS<18) | 37 | 45.7 | 75 | 19.6 | 5 | 22.7 | 6 | 8.7 |
| Cat 1.7 (AS≥18) | 1 | 1.2 | 8 | 2.1 | 0 | 0.0 | 3 | 4.3 |
| **Proportional shortfall** | | | | | | | | |
| Cat 1 (PS<0.85) | 11 | 13.6 | 321 | 84.0 | 2 | 9.1 | 64 | 92.8 |
| Cat 1.2 (0.85≤PS<0.95) | 54 | 66.7 | 45 | 11.8 | 13 | 59.1 | 5 | 7.2 |
| Cat 1.7 (PS≥0.95) | 16 | 19.8 | 16 | 4.2 | 7 | 31.8 | 0 | 0.0 |
| **Overall** | | | | | | | | |
| Mean weight | 1.281 | | 1.081 | | 1.350 | | 1.054 | |
| Cat 1 | 7 | 8.6 | 283 | 74.1 | 1 | 4.6 | 58 | 84.1 |
| Cat 1.2 | 58 | 71.6 | 77 | 20.2 | 14 | 63.6 | 8 | 11.6 |
| Cat 1.7 | 16 | 19.8 | 22 | 5.8 | 7 | 31.8 | 3 | 4.3 |

Since the introduction of the severity modifier:

* 98% of the decisions that did not attract additional weighting under the severity modifier wouldn’t have been eligible for additional weighting under the end of life criteria either.
* Only 1 of the 22 decisions estimated to have met the end of life criteria (5%) did not achieve additional weighting under the severity modifier. On the other hand, 11 of the 69 decisions estimated not to have met the end of life criteria (16%) achieved additional weighting under the severity modifier, 3 of which achieved a weight of 1.7.

In the majority of cases where the severity modifier was applied and where the end of life criteria were judged also to have been met, a weight of 1.2 was used (14 out of 21 decisions; 67%). For the remaining 7 decisions (33%), the 1.7 severity weighting given was the same as the maximum that would have been applied under end of life.

It is true that for the majority of decisions where the end of life modifier would have likely applied, a 1.2 weight was applied under the severity modifier (rather than up to the maximum 1.7 weight that would have been used under end of life). However, this shift was to be expected as it would not have been possible to allocate additional weightings to a broader range of decisions, within the principle of achieving opportunity cost neutrality, without some decisions attracting lower weighting than would have applied historically.

Within the implementation+ subsample, the proportion of decisions that were considered likely to have met the end of life criteria and achieved the highest 1.7 weighting for severity was substantial (32%), indicating that NICE still gives precedence to drugs that extend life in its decision making.

Of the 59 decisions that did not attract additional weight with the severity modifier, only 1 was estimated to have met the end of life criteria. This topic was for a first line non-small cell lung cancer (NSCLC) treatment, dabrafenib with trametinib for treating advanced BRAF V600 mutation-positive NSCLC. It received a positive recommendation without a severity weight.

Of the 32 decisions that had a severity weighting higher than 1 applied, 11 (34%) would not have met the end of life criteria for additional weight. Of these, 8 decisions had a 1.2 severity weighting applied by committee, with the remaining 3 receiving a 1.7 severity weighting. These decisions came from appraisals of treatments for hepatitis D, cystic fibrosis, and a range of cancer conditions including leukaemia and glioma.

Impact of the severity modifier on proportion of recommendations

We examined the proportions of positive recommendations (covering decisions recorded as recommended, optimised, or recommended with a managed access arrangement) associated with decisions in the historical primary+ subsample and the implementation+ subsample:

* 383 of the 464 decisions in the historical primary+ subsample were positive recommendations (82.5%). Of the 81 decisions where additional end of life weighting was applied, 67 were positive recommendations (82.7%).
* 76 of the 91 decisions in the implementation+ subsample were positive recommendations (83.5%). Of the 32 decisions where additional severity weighting was applied, 27 were positive recommendations (84.4%).
* Considering appraisals of treatments for cancer specifically, the proportion of positive recommendation decisions was 75% in the historical primary+ subsample, compared to 80% in the implementation+ subsample.

Considering appraisals of treatments for advanced and/or metastatic cancer specifically, the proportion of positive recommendation decisions was 69% in the historical primary+ subsample, compared to 81% in the implementation+ subsample.

Considering appraisals of treatment for non-cancer conditions, the proportion of positive recommendation decisions was 88% in the historical primary+ subsample, compared to 90% in the implementation+ subsample.

It should be noted that the implementation+ subsample contains a larger proportion of cancer topics than the primary+ subsample. Cancer topics have been more likely to result in negative recommendations than non-cancer topics. This disparity may explain why the overall proportion of positive recommendation decisions is very similar across the subsamples despite the implementation+ subsample containing higher proportions of positive recommendations for both cancer and non-cancer decisions (a statistical phenomenon known as Simpson’s paradox).

The statistics on recommendation rates reported above are subject to limitations. Decisions were categorised according to whether they related to advanced and/or metastatic cancer by NICE staff, and this categorisation did not capture blood cancer or brain cancer conditions which are staged differently from other cancer conditions. Missing historical data has led to further uncertainty.

Notwithstanding these limitations, there is no evidence to suggest that replacing the end of life modifier with the severity modifier has meant that positive recommendations are less likely to be reached.

Impact of the severity modifier on mean QALY weights

We used the implementation+ subsample to examine whether the QALY weights applied in NICE’s decisions had changed much on average since the introduction of the severity modifier.

Under the severity modifier, decisions in the implementation+ subsample received a mean severity weight of 1.125. This compares to a mean weight of 1.122 for the historical primary+ subsample under the end of life modifier (giving a difference of 0.003), and 1.116 for the primary+ subsample had the severity modifier instead been applied.

If NICE had continued with the end of life modifier, decisions in the implementation+ subsample would have received an estimated mean weight of 1.169. As above, this was based on the judgement of NICE staff retrospectively considering the available data. Due to the conservative approach taken to making these judgements, it is more likely to be an overestimate than an underestimate.

Additional information on mean weights across the different subsamples is presented in Table 5. For 1 decision in the historical primary+ subsample, the committee did not consider eligibility for the end of life modifier (TA192).

It appears that the impact of implementing the severity modifier on mean weights has been minimal, as there is no evidence to reject the hypothesis that there is no difference in mean weight between the subsamples. However, the size of the implementation+ subsample is small.

These findings are subject to the same limitations as the DSU’s results related to end of life, but they provide some evidence to refute the assertions made by some stakeholders that the adoption of severity modifier has been ‘conservative’ and is making it harder for patients to access new treatments.

In order to detect a much larger difference of at least 0.05 QALYs, we estimate that over 400 decisions in the implementation+ subsample would be needed.

Table 5: Comparison of mean weights across subsamples

|  |  |  |
| --- | --- | --- |
| Subsample | Modifier applied | Mean weight |
| Primary  n=364  Apr 2011 to Nov 2019 | End of life (actual) | 1.125 |
| Severity (estimated) | 1.119 |
| Primary+  n=464  Jan 2009 to Mar 2021 | End of life (actual) | 1.122 |
| Severity (estimated) | 1.116 |
| Implementation  n=68  Jan 2022 to Mar 2024 | Severity (actual) | 1.103 |
| Implementation+  n=91  Jan 2022 to Jul 2024 | Severity (actual) | 1.125 |

Figure 2: Mean weight over time, tracked on a quarterly basis

Figure 1 shows how the mean weighting applied to technology appraisal decisions has changed over time since the introduction of the severity modifier.

The mean weighting is calculated for each individual quarter (blue line), as well as cumulatively, where the mean weight for each quarter also includes any decisions published in earlier quarters (orange line). These are compared to the mean weights calculated for the historical dataset, applying either the severity modifier (solid black line) or the end of life modifier (dashed black line).

The mean weight for the implementation+ subsample shows high variability when tracked on a quarterly basis, fluctuating above and below the mean weights for the primary+ subsample.

In 3 quarters, the mean weight based on cumulative decisions in the implementation+ subsample under the severity modifier was greater than the historical mean weight under the end of life modifier (1.122), in 3 quarters it was lower, and in 1 quarter it was the same (to 3 decimal places).

Alternative methods for assessing opportunity cost neutrality

The analyses that underpinned the design of the severity modifier used individual decisions as the unit of analysis. Opportunity cost neutral options were identified by considering the average QALY weight per decision. The analyses did not account for the relative size (for example, population size or budget impact) of each decision.

There could be multiple decisions within a single appraisal, for example in multiple technology appraisals where each decision relates to a different treatment or combination therapy option, or when there are different patient subgroups or comparator options that were considered separately. Using decisions rather than topics as the unit of analysis made full use of the granularity of the available data and better reflected how committees make decisions in practice. To allow meaningful comparisons with the historical dataset, the analyses in this review also used individual decisions as the unit of analysis.

We are aware that the Association of the British Pharmaceutical Industry (ABPI) has been monitoring the impact of the new manual via its [Continuous NICE Implementation Evaluation (CONNIE)](file:///C:\Users\dcoombs\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\EM34ONS7\Reviewing%20implementation%20in%20practice%20of%20the%20NICE%20Health%20Technology%20Evaluation%20Manual%20-%20CONNIE%20August%202024%20(abpi.org.uk)) initiative. ABPI has reported statistics on the use of the severity modifier, derived using data submitted by companies for completed appraisals and using each topic as the unit of analysis.

We acknowledge the CONNIE findings and note two key differences why they differ from the results reported above.

First, the CONNIE database relies on company data submissions and does not include data on all completed technology appraisals. The latest version of CONNIE has data on 70 completed technology appraisals (including 9 cost comparisons) from submissions made to ABPI by 4 September 2024. According to the ABPI this represents 76% of the completed eligible appraisals over the time period covered.

By comparison, the NICE review included data for all technology appraisals undertaken using the new methods that were published up until the end of July 2024, excluding cost comparisons and decisions where severity weighting could not be calculated due to missing data. As noted above, the implementation+ subsample contains data on 91 decisions from 62 topics.

1. Cost comparison analysis compares the costs and resource use associated with a technology with that of its comparator(s). The effects of the technology and comparator(s) on health outcomes are captured in the clinical effectiveness evidence and are not included in the cost comparison analysis.
2. It is not possible to apply a severity weight to cost comparison evaluations as health outcomes are not captured in the model. Application of societal value would instead be indirect. If a comparator had received additional weighting or a higher threshold when evaluated by NICE, e.g. via end of life or severity weighting, or the highly specialised technologies process, the health benefits for the new technology would be valued equally to the past evaluation. This value is not considered or captured as part of the evaluation process.
3. The same considerations would be applied to evaluations captured in the historical dataset, which included only decisions where the incremental cost effectiveness ratio and QALY shortfall could be calculated. It is therefore appropriate and aligned to the methodological development of the severity modifier to exclude decisions based on cost comparisons from the implementation+ subsample analysis.

Second, the CONNIE analyses used topics as the unit of analysis, whereas the NICE review used decisions. ABPI have advised that in calculating the average QALY weight, they used an averaged weighting for topics where a different weight was granted for different subgroups (not accounting for patient population sizes). For multiple technology appraisals, ABPI included each medicine as a separate input.

We conducted further exploratory analysis to investigate how using topics as the unit of analysis would affect the average severity weight across the primary+, implementation and implementation+ subsamples:

* The primary+ subsample included 464 decisions across 269 topics. The average weight per topic under the end of life modifier was 1.157, compared to a per decision weight of 1.122. 1 topic was excluded from the analysis when considering the end of life modifier.
* When applying the severity modifier criteria to the primary+ subsample, the mean severity weight per topic was 1.139 compared to a per decision weight of 1.116.
* The implementation subsample included 68 decisions across 47 topics. The mean severity weight per topic was 1.078 compared to a per decision weight of 1.103,

The implementation+ subsample included 91 decisions across 62 topics. The mean severity weight per topic was 1.102 compared to a per decision weight of 1.125.

As noted above, there are multiple reasons why decisions could be considered separately within a technology appraisal. Since the historical dataset that informed the design of the modifier used decisions rather than appraisal topics, to allow meaningful comparisons with that dataset the relevant unit of measurement is decisions.

Any change to the severity modifier based on the reported ABPI CONNIE results would be cost inflationary. As noted previously, NICE is not in a position to implement actions that are cost inflationary without the agreement of the Department of Health and Social Care.

Limitations

The DSU report highlights several important limitations regarding the completeness and subjectivity of the data used in its quantitative analyses.

While limitations were identified in the historical data that support the current analyses, these were also known and accepted at the time of the methods review, which is why the design of the modifier, including the weights and cutoffs, was described as pragmatic in documents supporting the introduction of modifier and approved by the NICE Board in 2021.

Of all the limitations discussed in the DSU report, the one we consider most relevant is the size of the implementation subsample (n=68 decisions). For this reason, we extended the analyses to include decisions published up until the end of July 2024. While the sample size remains small in comparison with the primary+ subsample (91 decisions versus 464 decisions), the analyses undertaken make best use of the data available at the time of writing.

Reporting on the use of the severity modifier has changed over time. In the first year of implementation monitoring, all topics that had gone to at least 1 committee meeting were included in reports. However, the latest analysis includes only published final guidance to ensure accuracy of the findings.

The approach of including only published final guidance also has limitations as complex topics or topics that are appealed take longer to reach publication.

Conclusion based on this review

The severity modifier is operating as intended and has been applied to a wider range of diseases than the end of life modifier, as it was designed to do. Based on the data to date, it has remained opportunity cost neutral compared to the end of life modifier.

The results and discussion points presented in the DSU report do not indicate that NICE’s approach to the design of the modifier lacked validity. Nor do they indicate that use of the modifier is leading to outcomes that differ from what was originally intended when the pragmatic approach was adopted, i.e. that the levels for AS and PS would represent an appropriate level of severity at which to apply a QALY weight, and which capture a suitable range of cancer and non-cancer topics, within the principle of exceptionality, and without increasing resource use or displacing more health compared to the use of the end of life modifier.

Further analyses undertaken by NICE, including analyses using a more up-to-date dataset, have led to the same conclusion.

Plans for further work

Further monitoring of the use of the severity modifier

Given the high level of variability observed over time, we will continue monitoring the mean weighting applied to technology appraisal decisions. If the cumulative mean weighting were to fall below 1.10 for 2 consecutive quarters, we would investigate the causes and identify whether corrective action was needed. Any actions taken would depend on the findings of the investigation.

Research on societal preferences

When the severity modifier was designed, the underlying analyses and pragmatic approach were considered acceptable in the absence of robust evidence on the extent of societal value from treating severe conditions.

The existing evidence suggests that there is broad support for placing greater value on health gains in severe conditions ([see paper by Skedgel et al.](https://www.valueinhealthjournal.com/article/S1098-3015(22)00105-X/fulltext?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS109830152200105X%3Fshowall%3Dtrue), for example). However, the results of published studies are subject to experimental framing effects and biases, and vary depending on the concepts used to define severity or the methodology used. Further work is needed to understand the degree to which society favours treating severe diseases, and how categorisation, grouping, other equity factors, and type of health gain may influence societal preferences.

We are aware of an ABPI-funded study conducted by the Office of Health Economics that has found evidence of societal concern for treating severe conditions that differs from that currently embodied in the NICE modifier, and that starts at an earlier degree of severity than is reflected in the current cutoffs. It should be noted that severity weighting has been applied in over a third of decisions in the implementation+ subsample. Expanding the severity modifier to make even more conditions eligible for additional weighting would run counter to the principle of exceptionality. Nevertheless, we welcome such research in this important area and will carefully consider the findings.

As part of NICE’s own commitment to pursue further research on societal preferences in relation to severity, a first stage of research was commissioned through the National Institute for Health and Care Research (NIHR) Policy Research Programme. The focus was to provide recommendations to NICE on how best to elicit severity weights from a UK population. This work was conducted through the Policy Research Unit in Economic Methods of Evaluation in Health and Social Care (EEPRU) in 2023. [The EEPRU report by Keetharuth et al. was published in 2024.](https://orda.shef.ac.uk/articles/report/Techniques_for_eliciting_societal_preferences_for_severity_for_use_in_health_technology_assessments/25406713/1?file=45030970) It presents a critical review of relevant literature and findings from an expert elicitation exercise.

The report recommends that while there are several techniques that could be used to elicit societal preferences, NICE should not specify a single technique as all have limitations. Research should instead be undertaken using multiple approaches. A reasonable attempt to triangulate and reconcile the different results can then be made where the goal is not to obtain a precise number but to achieve broad agreement.

Following publication of the EEPRU report, we have begun scoping further research on society’s preferences on how much additional weighting to apply to health benefits for people with severe diseases. We intend to tender for this work in December 2024. It will report back around 2 years after field work begins. The findings will support future consideration of the severity modifier.

© NICE 2024. All rights reserved. [Subject to Notice of rights](https://www.nice.org.uk/terms-and-conditions#notice-of-rights).

September 2024