What is the impact of EQ-5D-5L vs 3L on NICE Severity Weights?

REPORT BY

THE DECISION SUPPORT UNIT

February 2026

Allan Wailoo, Aline Navega Biz, Mónica Hernández Alava

Health Economics and Decision Science, Sheffield Centre for Health and Related Research (SCHARR), University of Sheffield.

Decision Support Unit,

SCHARR, University of Sheffield,

Regent Court, 30 Regent Street Sheffield, S1 4DA

Website https://sheffield.ac.uk/nice-dsu

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) External Assessment Group is based at the University of Sheffield with members at York, Bristol, Exeter, Leicester, Warwick, Swansea and the London School of Hygiene and Tropical Medicine. The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Centre for Health Technology Evaluation Programmes. Please see our website for further information www.nicedsu.org.uk.

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

ACKNOWLEDGEMENTS:

We are grateful to NICE colleagues Koonal Shah, Juliet Kenny, Thomas Jarratt and Michelle Green, for their input to the design of this study, provision of the underlying appraisal materials, and constructive comments on the analyses performed. Donna Davis provided expert project management and report formatting assistance.

Any errors are the responsibility of the authors alone.

This report should be referenced as follows:

Wailoo A, Biz AN, Hernández Alava M. What is the impact of EQ-5D-5L vs 3L on NICE Severity Weights? NICE DSU Report, 2026.

EXECUTIVE SUMMARY

NICE currently recommends the EQ-5D instrument be scored using the 3-level (3L) value set. A new value set for the 5-level version (5L) will soon be publicly available. Policymakers in the UK health system, including NICE, need to know the likely impact of switching from 3L to 5L and one aspect of that is the impact on the severity modifier NICE applies in its technology appraisals programme.

The severity modifier is based on estimates of absolute and proportional shortfalls (AS and PS): the expected numbers of discounted QALYs for patients under current care compared to those for the general population.

This report estimates the impact of switching from using the 3L to using the 5L by analysing 39 appraisal decisions in our main analysis and 18 appraisal decisions in a sensitivity sample. Reported impacts are the combined effects of differences in the descriptive systems and the value sets. These appraisals were conducted using NICE recommended 3L values for health states.

We estimated the likely impact of moving to 5L by mapping all health states in the cost effectiveness models using an updated version of a widely used mapping model. We also used estimates of 5L conditional on age and sex using data from the Health Survey for England 2018. This is a critical element of the calculation of AS and PS.

We found that moving from 3L to the 5L value set changed the severity weight that would be applied in very few case study appraisals. In the main analysis only one appraisal decision would receive a severity weight of 1.2 instead of 1.7. There are greater changes in AS than PS as a result of the move from 3L to 5L but this has little material impact on overall severity categories for decision making purposes because these are more driven by PS than AS and severity category boundaries are wide, given the current design of the severity modifier. Results were robust to the sample of appraisals used and the method for estimating 5L by age and sex for the general population.

Appraisals in the oncology category all had higher AS. Non oncology products experienced greater variability and for some case studies the change in AS was substantial and negative. The largest reduction in AS was 3.8 QALYs. Larger reductions in both AS and PS are observed in appraisals for patients with younger starting ages, on average, in the cost effectiveness models.

The impact of the move from 3L to 5L should consider the impact on severity in conjunction with the findings on cost effectiveness reported elsewhere.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AS	Absolute Shortfall
NICE	National Institute for Health and Care Excellence
PS	Proportional Shortfall
QALE	Quality Adjusted Life Expectancy
HRQoL	Health Related Quality of Life
QALY	Quality Adjusted Life Year
3L	EQ-5D-3L
5L	EQ-5D-5L
HSE	Health Survey for England
DSU	Decision Support Unit
TSD	Technical Support Document

1. INTRODUCTION

The preferred instrument to assess Health Related Quality of Life (HRQoL) in adults in NICE appraisals is the EQ-5D[1]. There are two versions of the EQ-5D. The original EQ-5D-3L (3L) allows respondents to indicate the degree of impairment for each of the five domains of health covered by the descriptive system at one of three levels (no problems, some problems, extreme problems). There is a value set for the UK for each of the 243 health states that the 3L instrument can describe[2] that has been in use for decades.

A 5-level version (5L) of the EQ-5D has been in development for some years. It allows respondents to indicate on 5-levels the degree of impairment. Whilst the descriptive system has become widely used in clinical studies, there is no accepted value set for use in the UK. NICE currently recommends responses to the 5L descriptive system be mapped to 3L using results from analyses developed by the Economic Evaluation Policy Research Unit (EEPRU) and published by Hernández Alava et al[3]. Mapping takes the observed responses to the 5L descriptive system (or the implied responses where we only have a mean 5L value reported) and estimates the responses that would have been observed had the 3L instrument been administered. It then applies the 3L value set to the distribution of 3L responses. Mapping is therefore the combined impact of the differences in the descriptive systems and the value sets for 3L and 5L.

A new value set for 5L has been produced by Rowen et al[4]. At the time of writing this is not yet in the public domain but, as part of the preparation process for switching, NICE requires information on the likely impact of doing so. We know from previous analyses that the new 5L value set moves the distribution of health state values closer to 1 and compresses them into a smaller range and that this has implications for cost effectiveness estimates[5].

Another element of the potential impact of switching to 5L relates to the severity modifier. NICE uses a severity modifier to attach greater weight to technologies where the condition is considered more severe. Severity is defined in terms of absolute and proportional shortfall (AS and PS), with additional weights of 1.2 and 1.7 applied in cases where the degree of severity is classified as high. See for example Wailoo et al[6] for an explanation.

This report provides details of analyses that quantify the expected impact of moving from 3L to the 5L value set on the severity weights used in NICE Technology Appraisals. We report how estimates of AS and PS, and the numbers of decisions that qualify for each AS, PS and overall severity category, change when using 5L compared to 3L in a sample of technology appraisal decisions.

Beyond these analyses, further exploratory analyses were conducted to understand how these impacts were affected by the case studies included in the sample, the calculation methods used to estimate Quality Adjusted Life Expectancies (QALEs) in the general population, the type of technology under evaluation, and the starting age of patients in the cost effectiveness models.

2. DATA AND METHODS

2.1 Technology Appraisal samples

We extracted data from samples of NICE Technology Appraisals. We were provided with access to the executable cost-effectiveness models by NICE colleagues for calculations to be made. We reproduced the NICE appraisal committee preferred assumptions used to inform decision-making where possible and defaulted to the independent assessment group's preferred scenario in the small number of cases where this could not be achieved.

There are two sets of NICE appraisal decisions that have been used in this study.

The main analysis sample comprises 39 appraisal decisions from 37 appraisals, and matches the sample used for analysis of the impact of switching to 5L on incremental QALYs and cost effectiveness reported by Navega Biz et al[7]. These appraisals were selected pragmatically but aimed to cover appraisals spanning oncology, non-oncology where the intervention had an element of life extension and non-oncology where the intervention had no life extension. All the case studies reported using EQ-5D-3L utility values in their economic models at least for the main health states. This is referred to as the **"main sample".**

The main sample was originally selected to support another related impact analysis (the analysis of the impact of switching to the 5L value set on incremental QALYs and cost effectiveness, reported by Navega Biz et al[7]). It was not initially known if this sample was representative of the current Technology Appraisals programme in terms of the frequency with which the severity modifier is applied. An unrepresentative sample might lead to biased estimates of the impact on the numbers of appraisal decisions that would change AS, PS or overall severity category. Therefore, we compared the main sample to samples described in a previous DSU severity report[8]. NICE colleagues also compared the sample to the most recent monitoring data on the application of the severity modifier in Technology Appraisal decisions.

Compared to 68 appraisal decisions published between January 2022 and March 2024 (described as the "implementation sample" in the previous DSU report), the main sample has a higher proportion of appraisal decisions in categories 1.2 and 1.7 for AS. For PS, the new main sample has a higher proportion of appraisal decisions in category 1.2 but a lower proportion in category 1.7. Overall mean severity is higher in the new sample (1.133 vs 1.103).

When compared to the most recent monitoring data, the overall mean weight in the main sample was marginally higher than the cumulative mean weight across all appraisal decisions published since the implementation of the severity modifier (1.133 vs 1.116).

It was considered that the main sample was sufficiently representative of the Technology Appraisals programme that it should form the basis for the primary analyses.

The main sample was supplemented by a further 18 appraisal decisions from 12 appraisals. These were selected from appraisals where the appraisal was considered either to have met the criteria for additional severity weighting or was close to the boundary cutoffs. This is referred to as the **"sensitivity sample".** This is a relatively severe sample of appraisal decisions. For example, the mean severity weight applied in these appraisals is 1.267, higher than the main sample (1.133). The mean AS is 11.09, far higher than the mean of 8.02 in the main sample.

For the main analyses, where we are interested in how the distribution across severity categories is impacted by the move from 3L to 5L, we separate out these two samples.

For others, where the larger sample size is important and the interpretation of results is not undermined, we combine the samples. The sample is clearly indicated in the reporting of all results.

2.2 Changes made for 5L

We used the UK 5L value set, as yet unpublished, provided to us by the authors (Rowen et al)[4]. This allowed us to update the mapping model published by Hernández et al[9] and thereby predict what the mean 3L value for health states in the cost effectiveness models would have been had the 5L have been used. The mapping estimates the combined impact of the difference between the 3L and 5L descriptive systems and the value sets.

From the output of the case study cost effectiveness models, we obtained QALE for the comparator arm using 3L and 5L.

To estimate severity shortfalls (AS and PS) requires QALE from the comparator arm to be compared against QALE for the age and sex matched general population. There are 4 components to the calculation of QALE in the general population as detailed in Decision Support Unit (DSU) Technical Support Document (TSD) 23[10]. One of those 4 components is EQ-5D as a function of age and sex.

- For 3L, we followed the recommendations of TSD23 i.e. we used the estimates from the Health Survey for England (HSE), 2014. This recommendation came from previous DSU work[11]
- For 5L there are two options:
    1) Use the 3L values by age and sex and map them to 5L (referred to throughout as 5L Mapped).
    2) Use estimates derived from the most suitable national source where 5L has been directly observed by age and sex (referred to as 5L). This approach used HSE 2017-18, with modelling methods reported by Hernández et al[12].

The first approach, using mapping, has the advantage that differences observed are solely to do with the impact of 5L, although some degree of uncertainty is introduced by the reliance on mapping. The second makes interpretation of observed differences slightly more complex because it combines the impact of different source data (HSE

2014 vs 2017-18) and therefore the impact of different samples at different timepoints, with the impact of the different survey instrument and accompanying value set. We report results using both approaches below. Precedence is given to the second approach in this report as this reflects how QALEs and severity estimates will be calculated in real technology appraisals if the 5L value set is adopted. We include 5L mapped results to control for any impact of differences between the HSE samples.

AS and PS compare QALE for the general population with QALE for the patient population given current NHS care. The impact of shifting to 5L depends on the differential proportional impact on these two components.

Additional analyses sought to understand how these impacts on AS and PS were affected by the type of technology under evaluation, and the starting age of patients in the cost effectiveness models.

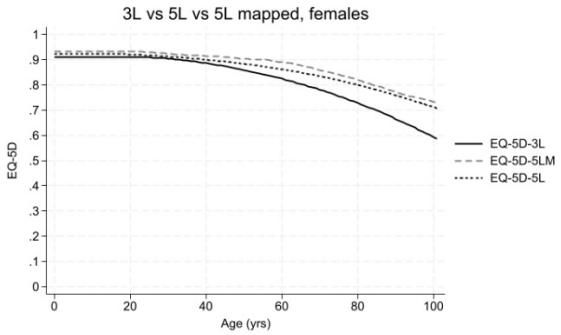
Analyses were conducted using Stata v19. All code was checked by a second member of the team. Calculations of AS and PS were additionally cross-referenced against the existing DSU severity calculator[13]. A second member of the team checked the implementation of changes to health state utility values in 5 case study cost effectiveness models.

3. RESULTS

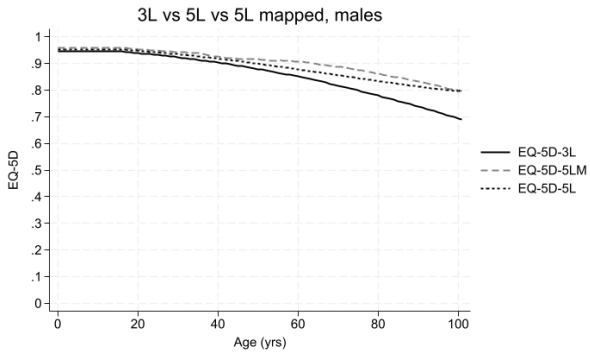
3.1 3L vs 5L vs 5L (mapped) utility by age

Figure 1: 3L vs 5L vs 5L (mapped) by age for a) females and b) males

a)



b)



One input to the calculation of AS and PS in subsequent sections is EQ-5D for the general population conditional on age and sex.

Figure 1 shows that 5L is higher than 3L at all ages, whether using mapped HSE 2014 estimates or HSE 2017-18 directly observed 5L. Both 5L variants yield similar estimates, indicating that the combined effect of mapping and sampling differences is minimal.

The difference between 5L and 3L is greater as age increases and as mean utility decreases. This is entirely consistent with expectations since the 5L value set yields a distribution of health states that are higher than those generated for the 3L (see for example Navega Biz et al[14] for more details on this issue).

Differences are greater for females than males. This is because females have a lower mean 3L than males, conditional on age.

Figure 2: QALE for 3L and 5L according to age for a) females and b) males

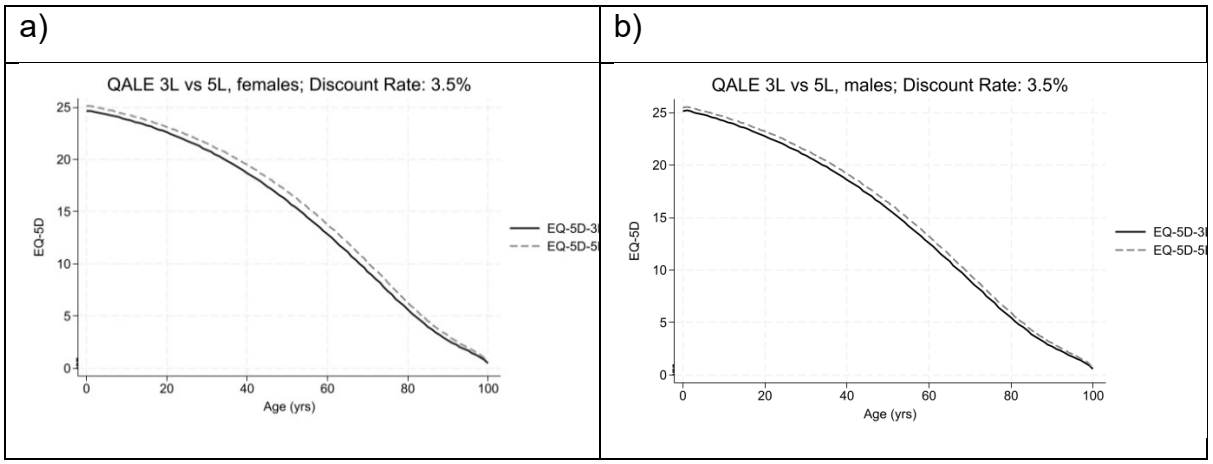


Figure 2 shows how these estimates translate into discounted QALE (dQALE) by age. In absolute QALE terms the difference diminishes with age, because life expectancy decreases. For females the range in the difference in dQALEs is 0.94 to 0.46 (comparing 3L with 5L). For males the range in the difference in dQALEs is 0.63 to 0.33 (comparing 3L with 5L).

3.2 The impact on severity of 3L vs 5L

3.2.1 Main comparison

Table 1: 3L vs 5L impact on absolute shortfall, proportional shortfall and overall severity weighting category in the main sample

	3L		5L	
Absolute Shortfall	**N**	**%**	**N**	**%**
Cat 1 (AS<12)	32	82.05	31	79.49
Cat 1.2 12≤AS<18)	5	12.82	7	17.95
Cat 1.7 (AS≥18)	2	5.13	1	2.56
Mean	8.02		7.73	
Cases greater under 5L			21	53.9
Proportional Shortfall				
Cat 1 (PS<0.85)	26	66.67	26	66.67
Cat 1.2 (0.85≤PS<0.95)	10	25.64	10	25.64
Cat 1.7 (PS≥0.95)	3	7.69	3	7.69
Mean	0.590		0.550	
Cases greater under 5L			1	2.6
Severity Overall				
Cat 1	23	58.97	23	58.97
Cat 1.2	12	30.77	13	33.33
Cat 1.7	4	10.26	3	7.69
Mean	1.133		1.121	

Table 1 reports the impact on severity when moving from 3L to 5L in the main sample. Few changes resulted in a different categorisation of severity. Considering severity overall, one appraisal decision that was in category 1.7 when using 3L switched to category 1.2. No category changes were observed in relation to PS, though mean PS was reduced from 0.59 to 0.55. All except one appraisal decision saw PS reduce. There was some slight recategorization observed in relation to AS: two more appraisal decisions in category 1.2 and one less in category 1.7. Mean AS was reduced from 8.02 to 7.73. Slightly more than half of appraisal decisions saw an increase in AS when moving to 5L (54%).

Table 2: 3L vs 5L impact on absolute shortfall, proportional shortfall and overall severity weighting category in the sensitivity sample

	3L		5L	
Absolute Shortfall	**N**	**%**	**N**	**%**
Cat 1 (AS<12)	13	72.22	10	55.56
Cat 1.2 12≤AS<18)	5	27.78	8	44.44
Cat 1.7 (AS≥18)	0	0	0	0.00
Mean	11.09		11.59	
Cases greater under 5L			16	88.9
Proportional Shortfall				
Cat 1 (PS<0.85)	5	27.78	5	27.78
Cat 1.2 (0.85≤PS<0.95)	9	50.00	10	55.56
Cat 1.7 (PS≥0.95)	4	22	3	16.67
Mean	0.883	22	0.870	
Cases greater under 5L			0	0.0
Severity Overall				
Cat 1	4	22.22	4	22.22
Cat 1.2	10	55.56	11	61.11
Cat 1.7	4	22.22	3	16.67
Mean	1.267		1.239	

The impact of moving from 3L to 5L in the sensitivity sample is reported in Table 2. A similar pattern to the main sample analysis is observed, even though this is a much more severe sample. One appraisal decision moves from overall severity category 1.7 to 1.2. There is again a reduction in the mean PS. However, unlike the main sample, AS sees an increase in mean and three appraisal decisions shift from category 1 to 1.2.

Taking both samples together, the evidence suggests that whilst there are greater changes in AS than PS as a result of the move from 3L to 5L, this has little material impact on overall severity categories for decision making purposes because these are more driven by PS than AS. This is consistent with previous work into the operation of the NICE severity modifier[15].

3.2.2 5L vs 5L (mapped)

There were only minor differences between using 5L and 5L mapped approaches, indicating that observed differences between 3L and 5L arise due to differences in the instruments and value sets rather than differences in the HSE samples.

There was no change to the distribution of appraisal decisions to overall severity categories or PS categories in either the main or sensitivity samples.

One appraisal from the main sample that was in category 1 for AS using 5L moved to category 1.2 when using 5L mapped. This was TA852 (review of TA669) which was 11.22 AS using 3L (category 1) but increased to 12.3 under 5L mapped (category 1.2) and was 11.9 under 5L (category 1). There were no AS category changes in the sensitivity sample.

Interestingly, in the main sample, mean AS is lower than 3L when using 5L (8.02 vs 7.73) but higher when using 5L mapped (8.02 vs 8.14). This pattern is not seen in the sensitivity sample. This is related to the fact that whilst 54% of main sample appraisal decisions saw a rise in AS when using 5L, the figure was 72% using 5L mapped. In the sensitivity sample, these figures were 89% vs 100% respectively.

3.2.3 Appraisals where severity categories changed

Table 3: Appraisals where absolute shortfall, proportional shortfall and overall severity weighting categories changed (main and sensitivity samples)

	AS category changed		PS category changed		Overall severity category changed	
	Main	**Sensitivity**	**Main**	**Sensitivity**	**Main**	**Sensitivity**
TA588	X				X	
TA866				X		X
TA642	X					
TA1063		X				
TA948(a)		X				
TA948(b)		X				

Table 3 lists all appraisals where a change in AS or PS weighting category was observed when moving from 3L to 5L.

In two cases, these changes also resulted in changes to the overall severity weight. The first of these was in the main sample.

- In TA588, “Nusinersen for treating spinal muscular atrophy (SMA)”, (late onset subgroup), AS was 18.6 (category 1.7) under 3L. This decreased to 14.8 under

5L (category 1.2). In this appraisal decision, the overall severity weight changed from 1.7 to 1.2. SMA is a rare genetic condition that had no alternative disease modifying therapies. The starting age in the economic model was 4 years.

The second case was in the sensitivity sample.

- In TA866, “Regorafenib for previously treated metastatic colorectal cancer” (in the comparison versus best supportive care), PS fell from 0.953 under 3L to 0.948 under 5L. This was a change in PS category (category 1.7 to category 1.2). The overall severity weight also changed from 1.7 to 1.2. The starting age in the economic model was 60 years.

In four other cases, changes to the AS category were observed but these did not result in changes to the overall severity weight. One of these cases was in the main sample.

- In TA642, “Gilteritinib for treating relapsed or refractory acute myeloid leukaemia”, AS was 11.5 (category 1) under 3L. This increased to 12.08 under 5L (just into category 1.2). Overall severity weight was unaffected because PS was category 1.2 under 3L and remained so under 5L.

The three other cases were in the sensitivity sample.

- In TA1063, “Capivasertib with fulvestrant for treating hormone receptor-positive HER2-negative advanced breast cancer after endocrine treatment”, AS was 11.4 (category 1) under 3L. This increased to 12.1 (category 1.2) under 5L. Overall severity weight was unaffected because PS was category 1.2 under 3L and remained so under 5L.
- In TA948 (a), “Ivosidenib for treating advanced cholangiocarcinoma with an IDH1 R132 mutation after 1 or more systemic treatments”, in the comparison versus best supportive care, AS was 11.99 (barely in category 1) under 3L. This rose above 12 to 12.7 (category 1.2) under 5L. Overall severity weight was unaffected because PS was category 1.7 under 3L and remained so under 5L.
- In the same appraisal, TA948 (b), in the comparison versus mFOLFOX, the same pattern was observed as for the comparison versus best supportive care. AS was 11.91 (category 1) under 3L. This rose to 12.6 (category 1.2) under 5L. Overall severity weight was unaffected because PS was category 1.7 under 3L and remained so under 5L.

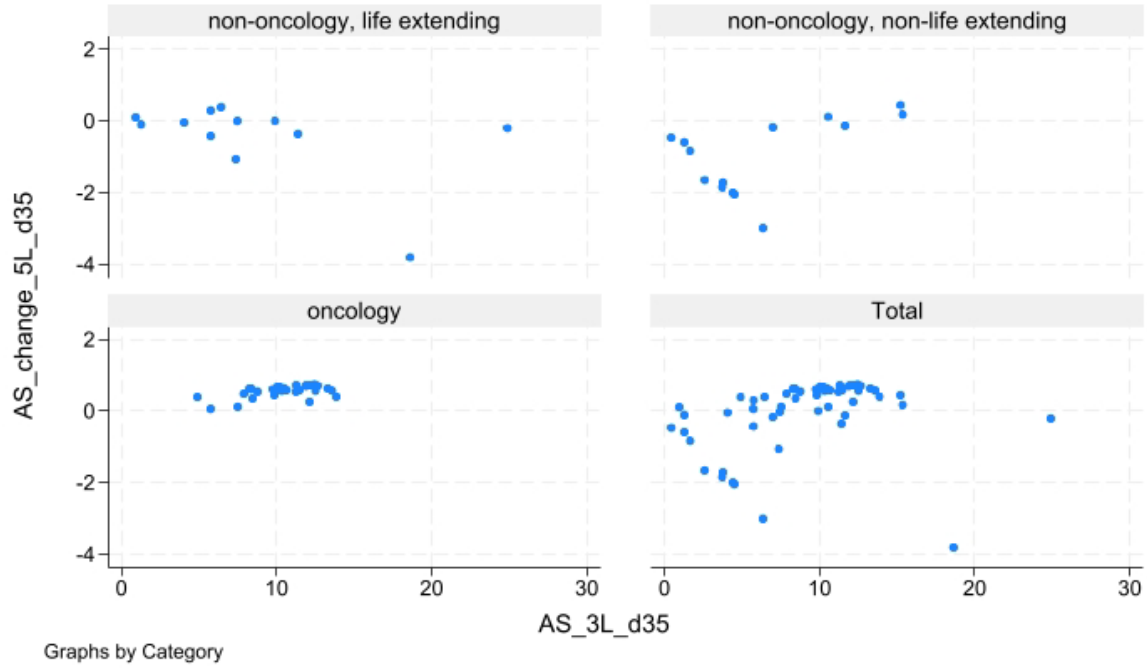
In two TA decisions the change from 3L to 5L led to a change in AS of 3 QALYS or more. Both cases were in the main sample. TA588, “Nusinersen for treating spinal muscular atrophy (SMA)”, (late onset subgroup), is listed above. In this case the large change in AS also resulted in a change in overall severity. The other case (TA935) did not lead to a change either in AS category or overall severity, despite the large magnitude change in AS.

- In TA935, “Secukinumab for treating moderate to severe hidradenitis suppurativa”, AS was 6.4 (category 1) under 3L and decreased to 3.4 under 5L (still category 1). PS was category 1 whether using 3L or 5L. Overall severity was unchanged.

3.2.4 Change in severity by baseline severity and technology type

We plot AS change against 3L AS in Figure 3, by appraisal technology type (oncology, non-oncology life extending, non-oncology no life extension). The plots merge the main and sensitivity samples together. In all oncology cases, the AS change is positive i.e. AS increases when moving to 5L from 3L. The pattern is more variable for other technology types but for non-oncology products the majority of appraisal decisions saw AS decrease as a result of moving from 3L to 5L. In those cases where AS increased, the change was relatively small. The maximum increase was 0.75 and every appraisal that saw a rise over 0.5 (n=23) was in the oncology category. Baseline AS has no clear relationship to the size or direction of AS change.

Figure 3: Absolute shortfall change against baseline absolute shortfall when moving from 3L to 5L



PS change against PS using 3L is shown in Figure 4 below. PS for oncology technologies is largely unchanged by the move from 3L to 5L. PS is reduced by the move to 5L for all oncology appraisal decisions (n=31) but the largest reduction is only 0.027.

Other technology types demonstrate variability in impact but with no obvious relationship to 3L PS. Non-life extending technologies decisions all have reduced PS, but the magnitude is variable. The largest reduction is 0.158 in this category. Eleven of twelve life extending technology decisions had reduced PS.

Figure 4: Proportional shortfall change against baseline proportional shortfall when moving from 3L to 5L

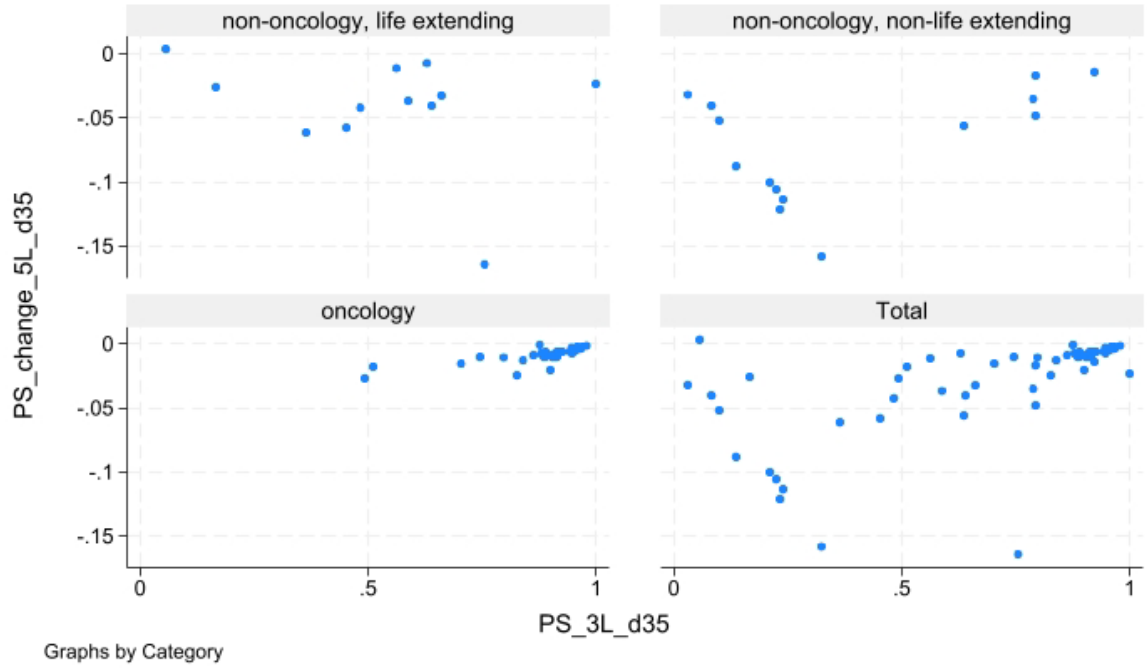
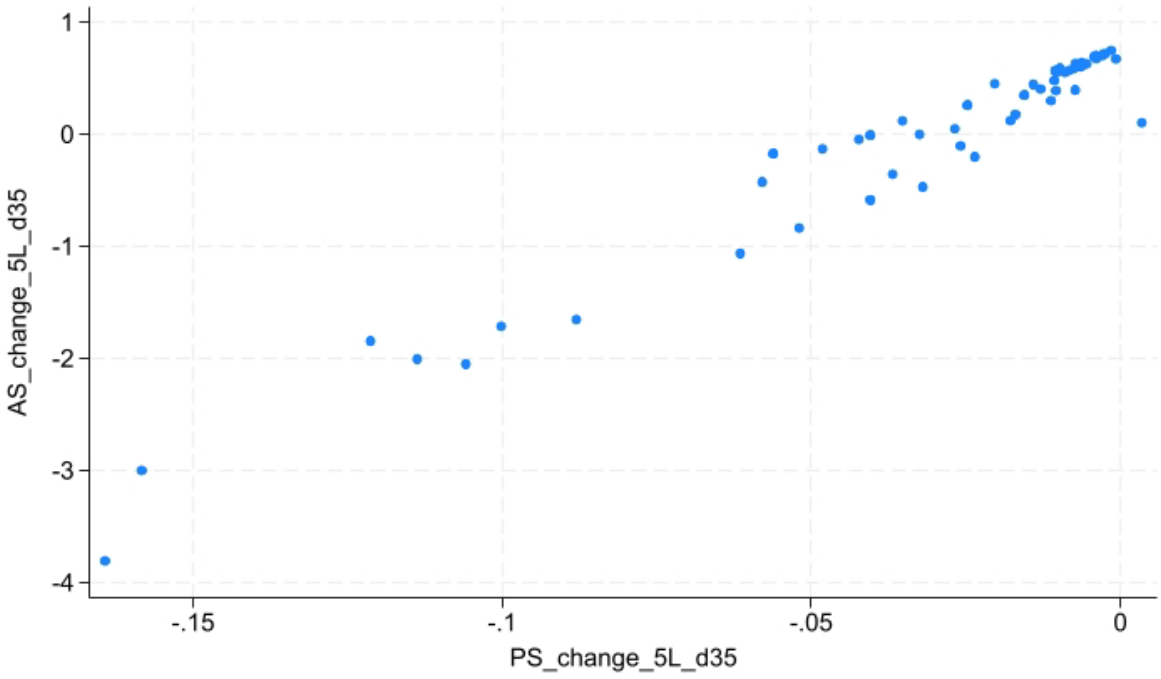


Figure 5 shows that AS and PS change are strongly, positively correlated, but there are a large number of case study appraisal decisions where AS change is positive but PS change is negative (36/57).

Figure 5: Absolute shortfall change vs proportional shortfall change when moving from 3L to 5L



3.2.5 Change in severity and baseline age

There is a clearer relationship between both AS and PS and the starting age of patients in the cost effectiveness models. Figure 6 shows scatterplots of AS and PS change against age in years. In both cases there is a clear positive relationship. Larger reductions in both AS and PS are observed in appraisals for patients with younger starting ages, on average, in the cost effectiveness models.

Figure 6a: Change in absolute shortfall by age

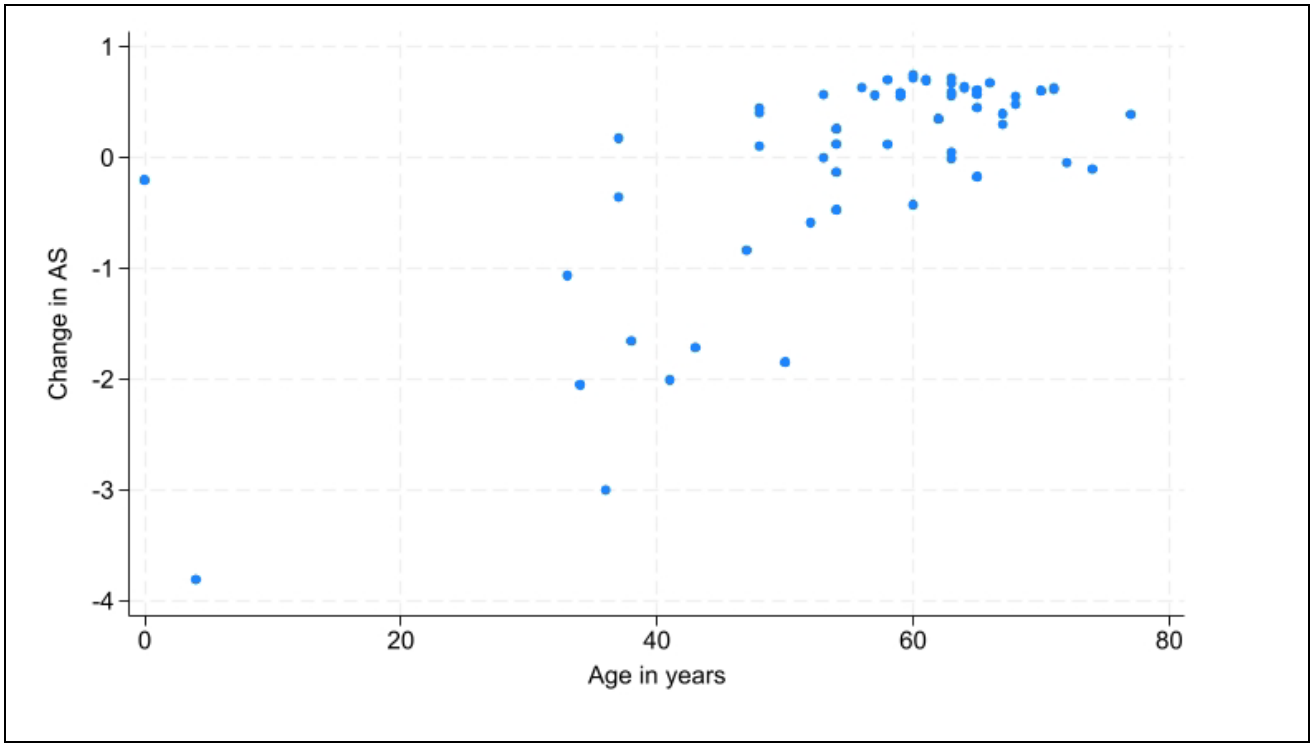
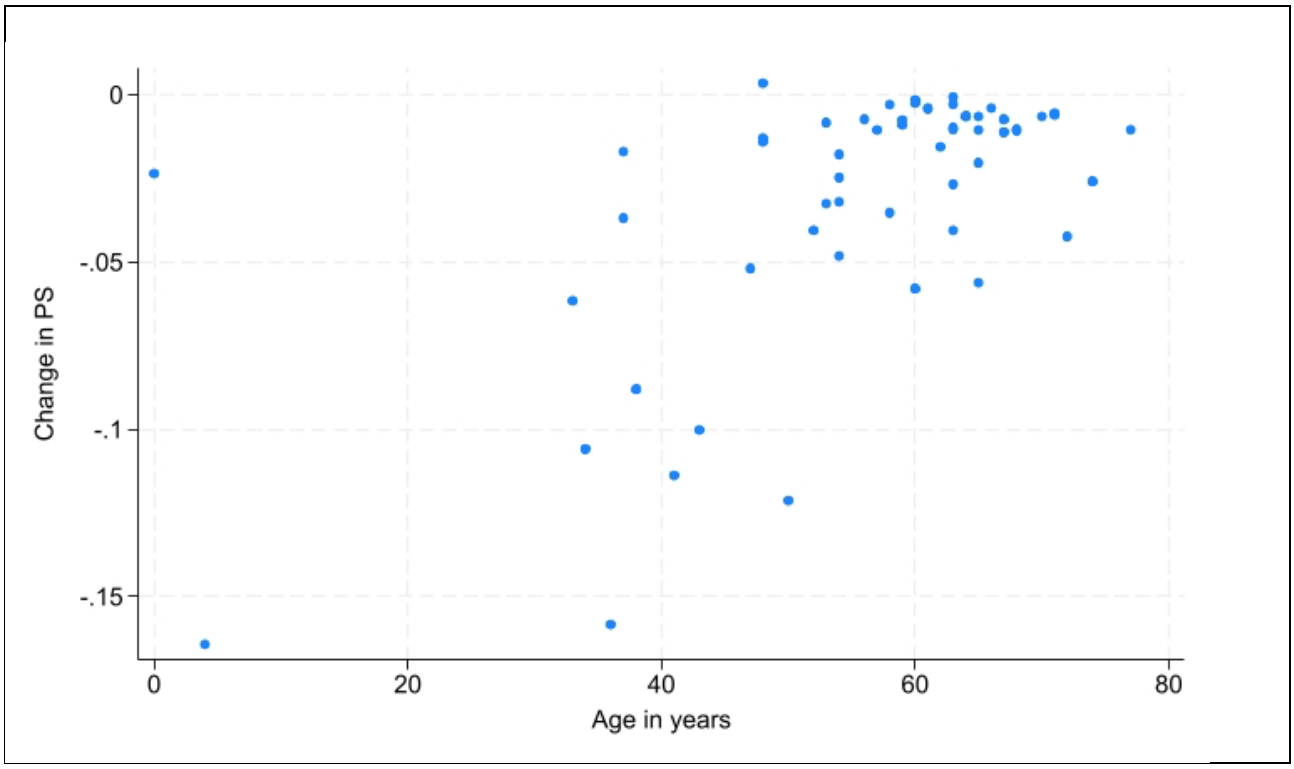


Figure 7b: Change in proportional shortfall by age



4. DISCUSSION

Using 3L versus 5L leads to very different estimates of HRQoL. We know that these two instruments cannot be considered interchangeable and that there can be substantial impacts on estimates of cost effectiveness. It is therefore important to be aware of other implications of adopting the new 5L value set for the UK. This report provides estimates of severity in terms of AS and PS, and using categories as defined in the NICE health technology evaluations manual.

Results from appraisals included in the sample suggest that moving from 3L to 5L is unlikely to change the overall severity category used for decision making purposes in the vast majority of appraisal decisions. This is because the change from 3L to 5L is predicted to have a greater impact on AS than PS and it is PS that determines overall severity weighting in the majority of TA decisions. For both AS and PS, severity categories are wide. This means that even relatively large changes in AS or PS may not span category boundaries. Large changes in AS were observed in some sample appraisals. For example, two case studies referred to in section 3.2.3 showed reductions in AS of three QALYs or more. In one case (TA588) this reduction led to a change of AS category from 1.7 to 1.2, and a change in overall severity weighting from 1.7 to 1.2. In the other case, the 3L-based AS estimate was already below the threshold for any additional weighting, and the lower estimate produced under 5L therefore made no difference to the overall severity weighting.

We tested the impact of different sources to calculate QALE: directly observed 5L responses in the HSE 2017-18 and mapped 3L responses from the HSE 2014. Only very minor differences were observed indicating that observed differences between 3L and 5L arise due to differences in the instruments and value sets rather than differences in the HSE samples.

We looked for evidence that the impact on severity of moving from 3L to 5L differed according to technology type (oncology, non-oncology life extending, non-oncology no life extension) for completeness. It is worth noting that AS and PS calculations are not based on evidence relating to the health technology under appraisal. Therefore, whilst

previous work on the impact of 5L on cost effectiveness identified differences in the impact on health technologies that are life extending versus those that are not, we may not expect this distinction to be relevant here unless treatment effects are correlated to other relevant characteristics of disease severity.

In all oncology cases, AS increases when moving to 5L from 3L but by a relatively small amount (mean 0.55 QALYs). The pattern was more variable in other technology categories. PS for oncology technologies is largely unchanged by the move from 3L to 5L. Eleven of twelve non-oncology, life extending technology decisions had reduced PS because of the move from 3L to 5L. However, the magnitude of the change was small in most cases.

The study has some limitations. The sample is inevitably limited in size due to the complexity and time required to implement the necessary calculations underpinning AS and PS for each appraisal decision. There were 39 appraisal decisions in the main sample and 18 in the sensitivity sample. In addition, the samples were chosen pragmatically, partly to ensure consistency with the corresponding impact assessment examining how the move to 5L may affect cost-effectiveness estimates. All analyses are based on mapping between 3L and 5L. Given the nature of the question, there is no obvious feasible alternative to predict the likely impact of moving from 3L to 5L. The mapping that has been used here is based on a very large UK sample, innovative study design and rigorous analytical techniques. We are confident that the key findings in this report are generalisable but, given the limitations, there are inevitable uncertainties around the estimated impacts.

In summary, the move from 3L to 5L is unlikely to have a significant impact on use of the severity modifier across the NICE Technology Appraisals programme as a whole, given the current cutoffs and combination of AS and PS used by NICE. However, in rare cases individual technologies may be substantially affected by the change, particularly where AS reduces. These are likely to be non-oncology products, but these could fall in either the life extending or non-life extending categories. In addition, were there to be any consideration of changes to the AS/PS category cutoffs, or the number of categories, the impacts highlighted in this report in relation to the changes in estimated AS and PS would be increasingly relevant.

The impact of the move from 3L to 5L should not consider the impact on severity in isolation but in conjunction with the findings on cost effectiveness as reported by Navega Biz et al[7].

REFERENCES

[1] NICE (2022) NICE Health Technology Evaluations: The Manual

[2] Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 Nov;35(11):1095-108. doi: 10.1097/00005650-199711000-00002. PMID: 9366889

[3] Hernández Alava, M., Pudney, S. and Wailoo, A. (2020) Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English population study. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. Report 063. DOI: https://doi.org/10.15131/shef.data.25219157

[4] Rowen D, Mukuria C, Bray N, Carlton J, Longworth L, Meads D, Oluboyede Y, O'Neill C, Yang Y. (2026) The new EQ-5D-5L Value Set for the UK. DOI: https://doi.org/10.1016/j.jval.2026.03.008

[5] Navega Biz A, Hernández Alava M, Wailoo, A. (in submission) "Switching from EQ-5D-3L to EQ-5D-5L in England: the impact in NICE technology appraisals", Value in Health

[6] Wailoo A, Bell Gorrod H, Dunning L, Kenny J, Leckenby E, Shah K. (2026) "The NICE experience of designing and utilising severity weights", Health Policy Open, Vol 10

[7] Navega Biz A, Hernández Alava M, Wailoo, A. (2026) EQ5D-5L adoption issues: the impact of switching from EQ5D-3L to a new value set for EQ5D-5L for model-based cost-effectiveness estimates in NICE technology appraisals. Policy Research Unit in Economic Evaluation of Health and Social Care Interventions. Universities of Sheffield and York. Report 079. DOI: 10.15131/shef.data.31889425

[8] Wailoo, A. Severity Weights in NICE Technology Appraisals. 2024. Available at https://sheffield.ac.uk/nice-dsu/methods-development/severity-weights

[9] Hernández Alava, M., Pudney, S., Wailoo, A. (2022) "Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK Population Study", Pharmacoeconomics

[10] Wailoo, A., A guide to calculating burden of illness for NICE Evaluations: Technical Support Document 23, 2024 [Available from http://www.nicedsu.org.uk]

[11] Hernández Alava M., Pudney S., Wailoo A. Estimating EQ-5D by Age and Sex for the UK. NICE DSU Report. 2022. Available at https://www.sheffield.ac.uk/sites/default/files/2022-02/DSU%20Age%20based%20utility%20-%20Final%20for%20website.pdf

[12] Hernández Alava M, Wailoo A. (forthcoming) "EQ-5D-5L Age Adjustment Estimates Using the Health Survey for England" EEPRU Report

[13] DSU Severity Shortfall calculator available at https://sheffield.ac.uk/nice-dsu/tsds/severity-shortfall-tsd

[14] Navega Biz A, Hernández Alava M, Wailoo, A. (in submission) “Switching from EQ-5D-3L to EQ-5D-5L in England: the impact in NICE technology appraisals”, Value in Health

[15] Wailoo, A. Severity Weights in NICE Technology Appraisals. 2024