



Journal of Health Economics and Outcomes Research

Methodology and Healthcare Policy



Quantifying Disease Severity in Health Technology Assessment in Japan: A Retrospective Analysis Using Quality-Adjusted Life-Year Shortfalls

Miwa Enami^{1*}, Akira Yuasa¹, Shunya Ikeda²

¹Japan Access & Value, Pfizer Japan Inc., Tokyo, Japan

²Department of Public Health, School of Medicine, International University of Health and Welfare, Narita, Japan

ARTICLE INFORMATION

Accepted November 10, 2025

Keywords: cost-effectiveness analysis; quality-adjusted life-years; health technology assessment; proportional shortfall; absolute shortfall; Japan

*Corresponding author:

Email address: miwa.enami@pfizer.com

➤ [Supplementary Material](#)

ABSTRACT

Background: Since 2019, Japan has implemented health technology assessments (HTAs) for selected drugs and medical devices. In the HTA system, the incremental cost-effectiveness ratio (ICER), calculated using quality-adjusted life-years (QALYs), is employed to guide price adjustments. However, the current system does not incorporate a quantitative assessment of disease severity.

Objectives: This study aimed to evaluate whether severity modifiers based on QALY shortfalls correspond to conditions currently granted special consideration, that is, those eligible for a higher ICER reference value (1.5× the standard), and to explore their implications for Japan's HTA system.

Methods: We retrospectively analyzed 32 drugs assessed under Japan's HTA up to March 2025. Absolute shortfall (AS) and proportional shortfall (PS) were calculated using age, sex distribution, and comparator quality-adjusted life expectancy estimates from manufacturer assessments and public assessments. Severity categories were defined as $\times 1.0$ ($AS \leq \times 1.2$ or $PS \leq 0.85$), $\times 1.2$ ($1.2 < AS < 1.8$ or $0.85 < PS < 0.95$), and $\times 1.7$ ($AS \geq 1.8$ or $PS \geq 0.95$). The concordance between severity classification and policy-based special consideration was then examined.

Results: Twenty-five matched target populations were identified. Mean AS and PS values did not differ significantly between manufacturer and public assessments, although manufacturers tended to report higher shortfalls. All cancer and pediatric cases were classified as $\times 1.2$ or $\times 1.7$, whereas 1 designated intractable disease was classified as having low severity ($\times 1.0$). Chronic and infectious diseases fell into higher severity categories despite not currently being subject to special consideration. Weighted mean severity values were comparable to those used in the UK's National Institute for Health and Care Excellence benchmarks.

Discussion: The findings revealed both alignment and misalignment between Japan's current HTA policy and severity classification. While cancer and pediatric diseases were consistent with the existing system, some serious diseases might have been overlooked, and the designated intractable disease might not align with quantitative severity criteria.

Conclusions: QALY shortfalls may serve as a complementary approach to identifying unmet health needs within Japan's HTA system. To ensure methodological robustness and social acceptance, broader validation, standardized estimation methods, and stakeholder consensus are necessary for effective decision-making.

BACKGROUND

Advances in medical technology have greatly improved human health, supported life functions, and extended life expectancy. However, healthcare systems face increasing financial pressures, largely due to demographic shifts such as population aging and the growing adoption of advanced medical technologies. Such pressures have raised concerns

about the sustainability of healthcare budgets; in response, health technology assessments (HTAs) have been integrated into pricing and reimbursement decision-making processes in many countries. However, HTA systems differ across the world.¹⁻⁵

Since April 2019, Japan has applied an HTA to adjust the prices of drugs and medical devices that meet its selection criteria.⁶ Within this framework, the prices of drugs and medical devices with expected high



sales or notably high unit costs are adjusted based on the incremental cost-effectiveness ratio (ICER) derived from cost-effectiveness analyses conducted by manufacturers and public research organizations. Manufacturer assessments are conducted by pharmaceutical or medical device companies to analyze the cost-effectiveness of their own products, following an agreed analytical framework established with an expert committee. Public analyses are conducted by public research organizations or academic groups to review and, when necessary, reanalyze the manufacturers' submissions.⁶ Quality-adjusted life-years (QALYs) serve as the measure of effectiveness in these cost-effectiveness analyses.⁷ The calculated ICERs are classified into 5 categories, ranging from under 2 million Japanese yen (¥) (US \$13 212)/QALY to ¥10 million (US \$66 063)/QALY and over. If the ICER is less than ¥5 million (<US \$33 031)/QALY, no price adjustment is made. In contrast, if the ICER is at least ¥5 million (\geq US \$33 031)/QALY, the price is reduced according to a coefficient assigned to each ICER category.⁶ Here, US \$1 is converted to ¥151.37, the average exchange rate in 2024.⁸ Meanwhile, drugs and medical devices for cancer, pediatric disease, and designated intractable disease⁹ are currently subject to special consideration, with a higher ICER reference value (1.5 \times standard) applied.⁶ However, this special consideration is a policy-based regulation rather than an approach grounded in a quantitative assessment of disease severity.

Conversely, Norway, the Netherlands, and the United Kingdom (UK) incorporate quantitative severity, measured by absolute shortfall (AS) and proportional shortfall (PS) of future QALYs lost due to disease, into their HTA systems. Specifically, quantitative severity is used either to adjust the ICER threshold (as in Norway and the Netherlands) or as a multiplicative modifier applied to QALYs, as implemented by the UK's National Institute for Health and Care Excellence (NICE). In recent years, the QALY shortfall concept, a hybrid framework that integrates the fair innings principle and considerations of severity, has gained increasing attention. Empirical studies, such as that by Stolk *et al.*,¹⁰ have shown that public preferences often prioritize patients with greater QALY shortfalls, reflecting a societal inclination to favor treatments for individuals experiencing more severe health losses.¹¹ By integrating such societal perspectives into their HTA systems, Norway, the Netherlands, and the UK have established institutional mechanisms that promote fairer allocation of medical resources to patient groups with serious conditions.¹²⁻¹⁴

This study assesses the feasibility of incorporating severity modifiers into Japan's HTA system by examining the relationship between estimated QALY shortfalls, based on the NICE severity modifier

framework, and the diseases currently granted special consideration under Japan's HTA policy.

METHODS

Study Population for QALY Shortfall Estimation

We retrospectively estimated QALY shortfalls for drugs evaluated under Japan's HTA system, applying the severity modifier framework used by NICE. As of March 2025, HTA evaluations had been completed for 32 drugs, with results made publicly available on the Center for Outcomes Research and Economic Evaluation for Health (C2H) website.¹⁵ These drugs corresponded to 74 target populations. After excluding populations without additional clinical benefits, 49 target populations remained from manufacturer assessments and 48 from public assessments. To accurately calculate AS and PS, we selected only those populations with information on mean age and sex distribution or at least mean age, leaving 26 populations from manufacturer assessments and 26 from public assessments. For populations with missing sex data but available mean age, the sex distribution was imputed as 50% male and 50% female.

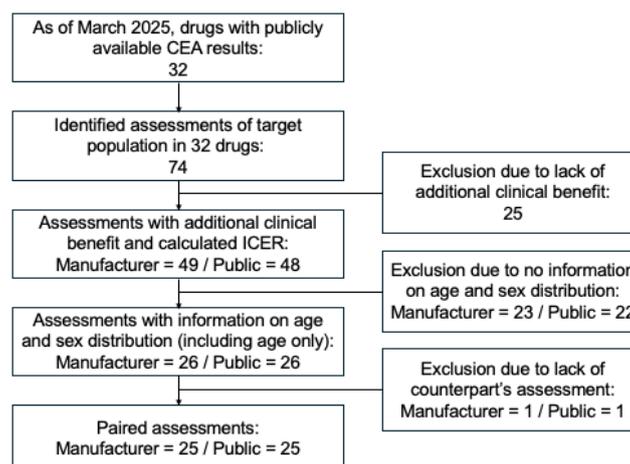
To determine how input parameters affect QALY shortfall estimates, we compared manufacturer and public assessments, which may differ in the data sources used to define target population characteristics. This comparison allowed us to examine how variations in population attributes, such as age and sex distribution, might influence severity classification outcomes. For this pairwise comparison, a set of 25 matched target populations was established as the study population (**Figure 1**). The information status on age and sex distribution is shown in **Figure 2**.

The mean age and sex distribution of the study population, along with patient quality-adjusted life expectancy (QALE) of the comparator under standard treatment, were systematically extracted from published manufacturer and public assessment documents. To ensure data accuracy, two researchers reviewed the extracted information. For any unclear points, a third researcher was consulted, and consensus was reached through discussion.

Definition and Calculation of AS and PS

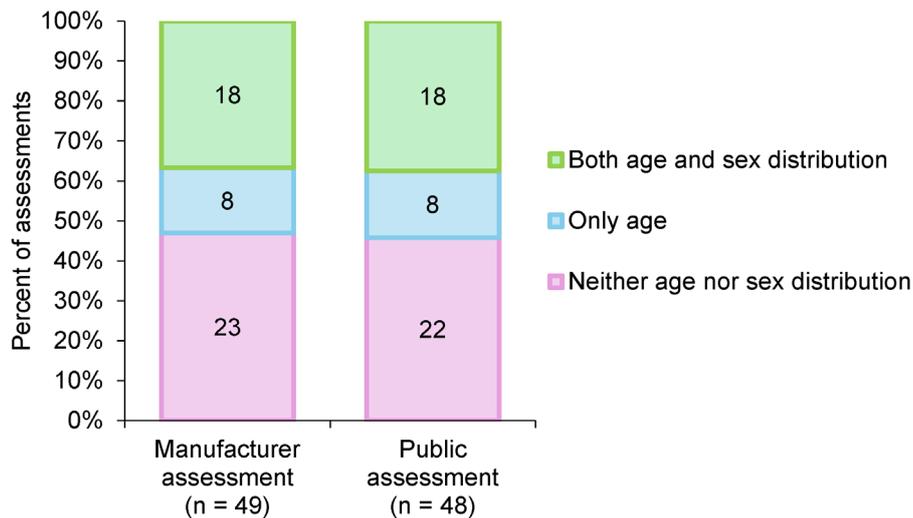
In this study, QALE was defined as the lifetime sum of QALYs. While a QALY represents 1 year of life adjusted for health-related quality of life, QALE reflects the total expected QALYs over an individual's remaining lifetime. The AS represents the total future health (measured in QALYs) and the PS represents the proportion of future health

Figure 1. Flow of Selection of Study Populations



Abbreviations: CEA, cost-effectiveness analysis; ICER, incremental cost-effectiveness ratio.

Figure 2. Information Status on Age and Sex Distribution in Manufacturer and Public Assessments



Note: Included are 49 target populations from manufacturer assessments and 48 from public assessments, each demonstrating additional clinical benefits.

(expressed in QALYs) that an individual is expected to lose as a result of their condition.^{12,16} AS and PS were calculated as follows:

$$AS = \text{Discounted QALE of the General Population} - \text{Discounted QALE of the Comparator}$$

$$PS = AS / \text{Discounted QALE of the General Population}$$

The QALE for the general population was estimated using the formula proposed by NICE,¹⁷ incorporating abridged life tables for 2019¹⁸ and EQ-5D-5L-based utility values specific to the Japanese population¹⁹ (Figure 3 and Supplementary Table S1). However, previous studies have not reported utility values for individuals younger than 16 years and older than 90 years. Therefore, we applied utility values for age 16 and 90 to age groups younger than 16 and older than 90, respectively. A 2% annual discount rate was applied in accordance with Japanese HTA guidelines.²⁰

Discounted QALE for the general population was calculated¹⁷ as follows:

$$\sum_{x=i}^{100} Lx \cdot Ux \cdot \frac{1}{(1+r)^{x-i}}$$

where Lx represents the number of life years lived between age x and age $x + 1$, adjusted using a half-cycle correction (ie, the average of survivors at ages x and $x + 1$); Ux is the age-specific health utility weight derived from EQ-5D-5L population norms; r is the annual discount rate (2%); and i is the starting age of the target population. Because Lx and Ux differ by sex, calculations were performed separately for males and females, and a weighted mean was computed based on the sex distribution of the target population.

The QALY weighting system for severity, as currently used by NICE, was then applied to the calculated AS and PS scores²¹ as follows:

QALY weight = $\times 1.0$ if $AS < 12$ or $PS < 0.85$

QALY weight = $\times 1.2$ if $\times 12 \leq AS < 18$ or $0.85 \leq PS < 0.95$

QALY weight = $\times 1.7$ if $AS \geq 18$ or $PS \geq 0.95$

Overall severity weights were assigned according to the higher value between AS and PS. Based on the assigned QALY weight, AS, PS, and overall severity were classified into three levels: $\times 1.0$, $\times 1.2$, and $\times 1.7$.²¹

Statistics

Descriptive statistics were used to summarize attributes of study populations, patient QALYs of the comparator, and estimated AS, PS, and severity weights. Continuous variables are reported as mean \pm SD and categorical variables as number and percentage. To identify characteristic differences between manufacturer and public assessments, we compared continuous variables using either the t test or Wilcoxon rank-sum test; categorical variables were compared using chi-square or Fisher’s exact tests, as appropriate. All analyses were conducted using Stata SE 17.0 (StataCorp LLC). A P value of $<.05$ was considered statistically significant.

Cross-Tabulation of Severity Classifications by Disease Category

To assess the alignment between severity classifications based on NICE’s QALY shortfall framework and Japan’s current policy on special consideration, we cross-tabulated severity categories by disease type (eg, cancer, pediatric diseases, and designated intractable diseases).

To further evaluate the alignment between severity classifications based on NICE’s QALY shortfall framework and Japan’s current policy on special consideration, agreement analyses were conducted using Cohen’s kappa. Specifically, 2×2 contingency tables were constructed to compare policy-based special consideration ($\times 1.0$ vs $\times 1.5$) with QALY shortfall-based severity categories ($\times 1.0$ vs $\geq \times 1.2$), and the percent agreement and Cohen’s kappa coefficients were calculated. The 95% confidence intervals (CIs) for Cohen’s kappa were estimated using bootstrap resampling (10 000 iterations) because standard functions in Stata 17.0 did not provide reliable confidence interval estimates given the small sample size. Analyses were performed separately for the manufacturer and public assessments, as well as for AS, PS, and overall severity.

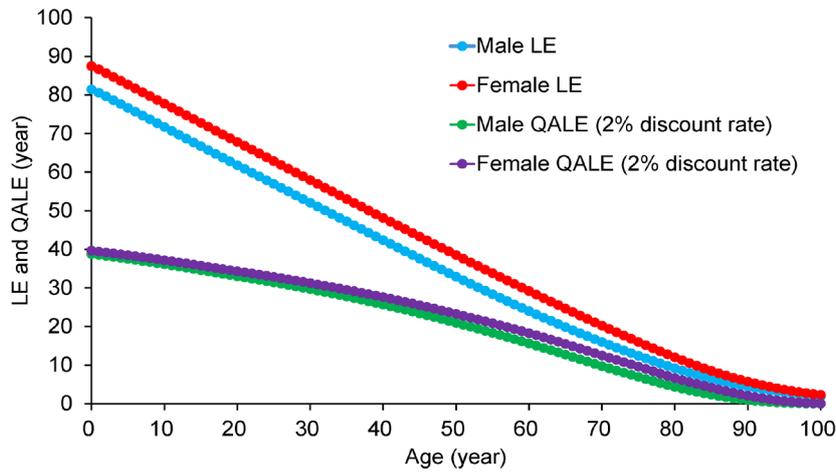
RESULTS

Comparison of AS and PS Between Manufacturer and Public Assessments

Table 1 summarizes the characteristics of study populations, estimated AS and PS, and the distribution of severity categories derived from manufacturer and public assessments ($n = 25$ each).

The mean age (\pm SD) of the study populations was 57.94 ± 11.21 years in manufacturer assessments and 60.94 ± 11.94 years in public assessments, with the latter group being slightly older ($P = .36$). The

Figure 3. Undiscounted Life Expectancy and Discounted Quality-Adjusted Life Expectancy at Each Age



Abbreviations: LE, life expectancy; QALE, quality-adjusted life expectancy.

Note: LE was derived from the abridged life tables for Japan, 2019,¹⁸ and QALE was calculated using Japan’s EQ-5D-5L values¹⁹ with a 2% discount rate.²⁰

Table 1. Characteristics of Target Population, QALYs of the Comparator, and Distribution of Absolute Shortfall, Proportional Shortfall, and Overall Severity Categories in Manufacturer and Public Assessments

Characteristic	Manufacturer Assessment (n = 25)	Public Assessment (n = 25)	P Value
Age (years), mean (SD)	57.94 (11.21)	60.94 (11.94)	0.36
Female proportion (%), mean (SD)	43 (16)	38 (22)	0.40
QALE of the comparator, mean (SD)	7.539 (6.14)	7.745 (6.54)	0.98
Absolute shortfall			
Mean (SD)	9.863 (5.452)	7.751 (5.002)	0.16
Weight, n (%)			0.19
	Cat. ×1.0 (AS <12)	15 (60.0)	18 (72.0)
	Cat. ×1.2 (12 ≤ AS <18)	7 (28.0)	7 (28.0)
	Cat. ×1.7 (AS ≥18)	3 (12.00)	0 (0.0)
Proportional shortfall			
Mean (SD)	0.581 (0.294)	0.521 (0.304)	0.48
Weight, n (%)			0.92
	Cat. ×1.0 (PS <0.85)	18 (72.0)	19 (76.0)
	Cat. ×1.2 (0.85 ≤ PS <0.95)	4 (16.0)	3 (12.0)
	Cat. ×1.7 (PS ≥0.95)	3 (12.0)	3 (12.0)
Overall			
Mean weight (SD)	1.200 (0.296)	1.116 (0.232)	0.27
Overall weight category, n (%)			0.53
	Cat. ×1.0	15 (60.0)	18 (72.0)
	Cat. ×1.2	4 (16.0)	4 (16.0)
	Cat. ×1.7	6 (24.0)	3 (12.0)

Abbreviations: Cat., category; QALE, quality-adjusted life expectancy; QALYs, quality-adjusted life-years; SD, standard deviation.

proportion of female participants was also similar: 43% in manufacturer assessments and 38% in public assessments ($P = .40$). The mean QALYs of the comparator were nearly identical: 7.539 in manufacturer assessments and 7.745 in public assessments ($P = .98$). These results indicate no substantial differences in age, sex distribution, or comparator QALYs between the two study populations.

However, manufacturer assessments tended to yield higher AS and PS scores. The mean AS (\pm SD) was 9.863 ± 5.452 in manufacturer assessments and 7.751 ± 5.002 in public assessments ($P = .16$). The mean PS (\pm SD) was 0.581 ± 0.294 and 0.521 ± 0.304 , respectively ($P = .48$). A similar trend appeared in the severity classification based on AS: 3 cases were classified in the highest severity category ($\times 1.7$) in manufacturer assessments compared with none in public

assessments. Based on combined AS and PS results, the overall severity weight placed 10 cases in manufacturer assessments into higher severity categories: 4 in $\times 1.2$ and 6 in $\times 1.7$. In comparison, 7 cases in public assessments fell into higher severity categories, with 4 in $\times 1.2$ and 3 in $\times 1.7$. The weighted mean of the three severity categories was 1.200 in manufacturer assessments and 1.116 in public assessments.

Association Between Severity Categories and Disease Categories

Table 2 summarizes the relationship between two types of severity classification: one based on AS and PS using NICE’s methodology and the other based on disease categories designated for special consideration under Japan’s HTA system. This system applies a higher ICER reference value (1.5× the standard) to cancer, pediatric diseases, and

designated intractable diseases. Among these, all cancer (n=6) and pediatric (n=1) cases were classified as either higher category $\times 1.2$ or the highest severity category $\times 1.7$ in the overall severity classification. In contrast, a designated intractable disease case (n=1) was classified as $\times 1.0$ in both manufacturer and public assessments. These findings suggest that the diseases granted a higher ICER reference value under the current Japanese HTA system do not always align with the severity defined by NICE's QALY shortfall framework.

Analysis of agreement using Cohen's kappa provided additional insights into this alignment. For manufacturer assessments, agreement was moderate to substantial for AS and overall severity (percent agreement = 84.0%; $\kappa = 0.66$; 95% CI, 0.34-0.97 for each) and moderate for PS (86.4%; $\kappa = 0.58$; 95% CI, 0.11-1.05), indicating some

discrepancies between the Japanese reference value and NICE's QALY shortfall-based severity classifications. In contrast, public assessments showed stronger alignment, with almost perfect agreement for AS and overall severity (96.0%; $\kappa = 0.90$; 95% CI, 0.56-1.08 for each) and substantial agreement for PS (92.0%; $\kappa = 0.80$; 95% CI, 0.41-1.03) (**Supplementary Table S2**).

Notably, 3 of 17 cases outside the disease categories currently granted special consideration in Japan exhibited high AS and PS scores in manufacturer assessments. These included 2 infectious diseases and 1 chronic disease, each classified as either $\times 1.2$ or $\times 1.7$ in overall severity.

The relationship between AS and PS for each target population is further illustrated using scatterplots for both manufacturer and public assessments (**Figure 4**).

Table 2. Distribution of Absolute Shortfall, Proportional Shortfall, and Overall Severity Categories by Disease Category in Manufacturer and Public Assessments

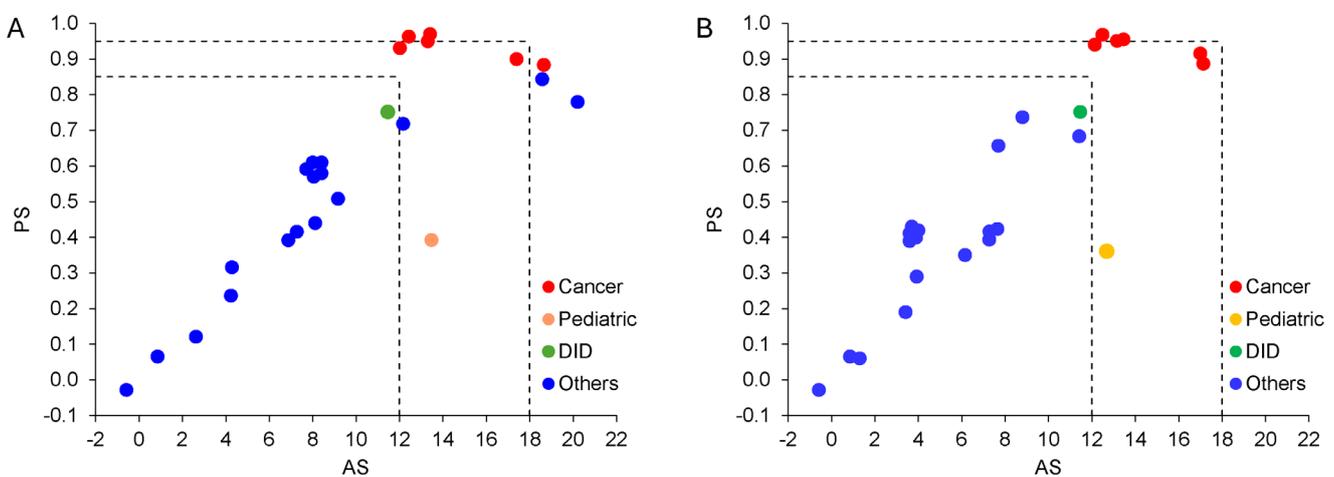
Disease Category, n (ICER Threshold)	Manufacturer Assessment (n = 25)						Public Assessment (n = 25)					
	Total N	Disease with Special Consideration			Others (n = 17) ($\times 1.0$)	Total N	Disease with Special Consideration			Others (n = 17) ($\times 1.0$)		
		Cancer (n = 6) ($\times 1.5$)	Pediatric (n = 1) ($\times 1.5$)	DID (n = 1) ($\times 1.5$)			Cancer (n = 6) ($\times 1.5$)	Pediatric (n = 1) ($\times 1.5$)	DID (n = 1) ($\times 1.5$)			
AS	Cat. $\times 1.0$	15	0	0	1	14	18	0	0	1	17	
	Cat. $\times 1.2$	7	5	1	0	1 ^a	7	6	1	0	0	
	Cat. $\times 1.7$	3	1	0	0	2 ^b	0	0	0	0	0	
PS	Cat. $\times 1.0$	18	0	1	1	16	19	0	1	1	17	
	Cat. $\times 1.2$	4	3	0	0	1 ^a	3	3	0	0	0	
	Cat. $\times 1.7$	3	3	0	0	0	3	3	0	0	0	
Overall	Cat. $\times 1.0$	15	0	0	1	14	18	0	0	1	17	
	Cat. $\times 1.2$	4	2	1	0	1 ^a	4	3	1	0	0	
	Cat. $\times 1.7$	6	4	0	0	2 ^b	3	3	0	0	0	

Abbreviations: AS, absolute shortfall; Cat., category; DID, designated intractable disease; PS, proportional shortfall.

^aHeart failure.

^bFungal infection.

Figure 4. Scatterplots of Estimated AS and PS Values of Target Populations: (A) Manufacturer Assessments; (B) Public Assessments



Abbreviations: AS, absolute shortfall; DID, designated intractable disease; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; PS, proportional shortfall.

Notes:

1. Each dot represents one of the 25 target populations analyzed in each assessment. The dots are color-coded by disease category to visualize the distribution of AS and PS values.
2. Cancer, pediatric disease, and DID are disease categories designated for special consideration currently under the Japanese HTA system.
3. Dashed lines indicate QALY shortfall thresholds used as severity modifiers, as defined by NICE.

DISCUSSION

This study identified alignment and misalignment between Japan's current policy on special consideration for certain diseases and severity classifications based on QALY shortfalls estimated from manufacturer and public assessments.

Although the mean AS and PS scores were similar between the two assessments, notable differences appeared in the distribution of severity categories. For AS, 3 cases (12%) in manufacturer assessments were classified in the highest category ($\times 1.7$), while no such cases were observed in public assessments. In these discrepant cases, the manufacturer-assessed populations were consistently younger: 41.3 vs 59.9 years, 49.5 vs 68.3 years, and 59.6 vs 68.8 years. Correspondingly, comparator QALYs were also higher in manufacturer assessments: 5.706 vs 5.282, 3.44 vs 3.137, and 4.76 vs 4.01 QALYs. These differences suggest that public assessments, which reflect older populations with shorter life expectancy, show smaller health disparities compared with the general population, resulting in lower AS values. While age is a major factor, the observed discrepancies also stem from differences in data sources and modeling assumptions. Manufacturer assessments typically draw on clinical trial data, which often include younger and healthier participants, whereas public assessments often rely on real-world sources such as the National Database of Health Insurance Claims and Specific Health Checkups of Japan, which consists of all patients, including older individuals and those with multiple comorbidities.^{22,23} Assumptions about long-term survival and utility also vary between models and directly affect QALY estimates. Thus, QALY shortfall calculations are highly sensitive to target population characteristics and modeling choices, with age being especially influential.

Most cancer cases were classified in the higher or highest categories ($\times 1.2$ or $\times 1.7$) in AS and PS analyses. Although only one pediatric case was included, it was also classified as $\times 1.2$. These results support the current policy of granting special consideration to certain diseases in Japan's HTA. By contrast, the designated intractable disease case was classified as low severity ($\times 1.0$), suggesting misalignment between policy-based special consideration and quantitative severity assessments based on QALY shortfalls. This implies the possibility of a reduced level of special consideration for diseases requiring ethical and social consideration.

It is noteworthy that 3 cases (ie, 1 chronic disease and 2 infectious diseases) not currently eligible for special consideration under Japan's system achieved high AS and PS scores in manufacturer assessments and were classified as $\times 1.2$ or higher using NICE's QALY weighting system. This finding suggests that the current policy may overlook certain diseases that merit prioritization from a QALY shortfall perspective.

The agreement analyses using Cohen's kappa indicate a high degree of alignment between the Japanese reference value and NICE's QALY shortfall-based severity classifications. Notably, the public assessments demonstrated a stronger alignment than the manufacturer assessments. Moreover, the weighted mean of the three severity categories derived from Japanese manufacturer assessments (1.200) and public assessments (1.116) was close to the benchmark value (1.119) used in designing NICE's severity modifier. All three figures were calculated using the same method, applying severity weights ($\times 1.0$, $\times 1.2$, and $\times 1.7$) to the distribution of disease severity. In NICE's case, this benchmark was based on a substantially larger dataset containing 364 HTA evaluations conducted between April 2011 and November 2019, distributed as 223 (61.3%) in severe weight $\times 1.0$, 111 (30.5%) in $\times 1.2$, and 30 (8.2%) in $\times 1.7$.²⁴ However, the distribution of disease severity in Japan differs from that in the UK, and the numerical similarity between the two countries may therefore be coincidental, reflecting differences in HTA objectives and the types of products evaluated. Importantly, NICE identified AS

and PS and the corresponding QALY weights through a retrospective analysis of 364 HTA decisions, which served as the primary evidence base for determining thresholds and weightings for the new modifier.²⁴ Japan should develop its own benchmark values that reflect the specific goals of its HTA system, supported by a sufficient number of validated cases to ensure methodological robustness.

When considering the incorporation of severity modifiers based on QALY shortfalls into Japan's HTA framework, a carefully designed stepwise approach is essential to address several key issues. First, it is necessary to build ethical and social consensus on the justification for priority setting, particularly in balancing the needs of older vs younger populations and patients with severe vs mild conditions. Second, standardized methods for estimating QALY shortfalls must be developed. A particular challenge is the lack of Japan-specific utility values for target populations and comparators under standard treatment,²⁵ which limits the accuracy of QALY shortfall estimates for the Japanese population. Third, to incorporate AS and PS results into adjustment of ICER reference value, it is necessary to specifically design and validate appropriate reference value and their corresponding adjustment coefficients. This process should be supported by further validation across a broader and sufficiently large set of HTA cases to ensure robustness and generalizability.

Several countries have explored candidate QALY shortfalls as a means for quantifying disease severity. Ultimately, Norway, the Netherlands, and the UK (NICE) adopted AS, PS, and a combination of both, respectively. In these countries, the introduction of QALY shortfall has been carefully considered not only as a tool for improving healthcare efficiency but also as a mechanism for applying ethical principles, such as "Fair Innings" and the "Rule of Rescue," to decisions about healthcare resource allocation.^{13,14,17,26} For example, during system development, NICE solicited public comments from a broad range of stakeholders, including patient organizations, the pharmaceutical industry, and the general public, and repeatedly deliberated on the ethical, clinical, and policy implications of the proposed changes.²⁷ Such a process that prioritizes transparency and consensus building is critical for strengthening societal acceptance of the institutional framework.

This study has several limitations. First, approximately half of the reported HTA cases were excluded, as only those with available information on age and sex distribution were analyzed. As a result, the number of matched manufacturer and public assessments was limited to 25. Among these, only one case involved a pediatric disease and, one involved a designated intractable disease, which constrains the interpretation of results for these disease categories. Thus, the small sample size and limited disease representation may constrain the external validity of our findings, as they may not fully capture the characteristics of HTA cases in Japan. Future research should validate these findings using a larger and more diverse dataset to enhance generalizability. Second, the QALY of the general population was estimated using the 2019 abridged life table¹⁸ and EQ-5D-5L-based utility values¹⁹ specific to the Japanese population. These values were uniformly applied across all cases. However, QALY estimates vary depending on the choice of life table, utility values, discount rate, and calculation method, and AS and PS values may differ under alternative assumptions. NICE has investigated how methodological and data source differences influence outcomes and emphasized the importance of standardizing calculation methods.²¹ Furthermore, to ensure accuracy and transparency in QALY shortfall estimation, future HTAs should extract age and sex data for target populations using a standardized reporting format.¹⁷ Third, sensitivity analyses were not conducted to test the robustness of the results against variations in key assumptions and input parameters. Such analyses could offer additional insights. However, this study primarily aimed to examine the conceptual alignment between Japan's qualitative

severity considerations and NICE's quantitative severity categories, rather than to validate parameter-level robustness. Future research should incorporate sensitivity analyses when developing Japan-specific shortfall estimation methods and severity coefficients. Finally, because public assessments follow manufacturer submissions, any differences between them may have been influenced by this sequence. Nonetheless, an independent appraisal is conducted to provide a comprehensive evaluation using both assessments.

CONCLUSION

This study suggests that QALY shortfalls provide a complementary approach to identifying unmet health needs that are overlooked under the current system. However, such a shift would not be uniformly advantageous: while it could extend benefits to previously neglected populations, it may simultaneously reduce advantages for others who have been recognized under the existing framework. Therefore, integrating QALY shortfalls into policy will require further validation based on a broader range of HTA cases. A transparent, stepwise process of consensus building among stakeholders is essential for ensuring methodological robustness and social acceptability.

Acknowledgments: Medical writing was supported by Takahiko Murata, PhD, of WysiWyg Co. Ltd., Tokyo, Japan, and was funded by Pfizer Japan Inc., in accordance with Good Publication Practice (GPP3) guidelines (<http://www.ismpp.org/gpp3>).

Disclosures: M.E. and A.Y. are full-time employees of Pfizer Japan Inc. S.I. has no conflicts of interest to declare.

Funding: This research and manuscript development were funded by Pfizer Japan Inc.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Sorenson C, Drummond M, Bhuiyan Khan B. Medical technology as a key driver of rising health expenditure: disentangling the relationship. *Clinicoecon Outcomes Res.* 2013;5:223-234. doi:10.2147/ceor.S39634
2. de Meijer C, Wouterse B, Polder J, Koopmanschap M. The effect of population aging on health expenditure growth: a critical review. *Eur J Ageing.* 2013;10(4):353-361. doi:10.1007/s10433-013-0280-x
3. Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *Eur J Health Econ.* 2018;19(1):123-152. doi:10.1007/s10198-017-0871-0
4. Sharma D, Aggarwal AK, Downey LE, Prinja S. National healthcare economic evaluation guidelines: a cross-country comparison. *Pharmacoecon Open.* 2021;5(3):349-364. doi:10.1007/s41669-020-00250-7
5. Fontrier AM, Visintin E, Kanavos P. Similarities and differences in health technology assessment systems and implications for coverage decisions: evidence from 32 countries. *Pharmacoecon Open.* 2022;6(3):315-328. doi:10.1007/s41669-021-00311-5
6. Hasegawa M, Komoto S, Shiroiwa T, Fukuda T. Formal implementation of cost-effectiveness evaluations in Japan: a unique health technology assessment system. *Value Health.* 2020;23(1):43-51. doi:10.1016/j.jval.2019.10.005
7. Edlin R, McCabe C, Hulme C, Hall P, Wright J. *Cost Effectiveness Modelling for Health Technology Assessment: A Practical Course.* Springer; 2015.
8. The World Bank Group. Official exchange rate (LCU per US\$, period average). Accessed August 11, 2025. <https://data.worldbank.org/indicator/PA.NUS.FCRF>
9. Kanatani Y, Tomita N, Sato Y, Eto A, Omoe H, Mizushima H. National registry of designated intractable diseases in Japan: present status and future prospects. *Neurol Med Chir (Tokyo).* 2017;57(1):1-7. doi:10.2176/nmc.st.2016-0135
10. Stolk EA, Pickee SJ, Ament AH, Busschbach JJ. Equity in health care prioritisation: an empirical inquiry into social value. *Health Policy.* 2005;74(3):343-355. doi:10.1016/j.healthpol.2005.01.018
11. Watters SM. Fair innings as a basis for prioritization: an empirical perspective. In: *Prioritization in Medicine: An International Dialogue.* Springer; 2016:179-196.
12. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. Accessed July 13, 2025. <https://www.nice.org.uk/process/pmg36>
13. Norwegian Ministry of Health and Care Services. Principles for priority setting in health care: summary of a white paper on priority setting in the Norwegian health care sector. Accessed July 13, 2025. <https://www.regjeringen.no/contentassets/439a420e01914a18b21f351143ccc6af/en-gb/pdfs/stm201520160034000engpdfs.pdf>
14. National Health Care Institute. Disease burden in practice: the theory and practice of calculating disease burden in package assessments (in Dutch). Accessed July 13, 2025. https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2018/05/07/ziektelast-in-de-praktijk/Ziektelast+in+de+praktijk_definitief.pdf
15. Center for Outcomes Research and Economic Evaluation for Health. Drugs and medical devices selected after the systematization of cost-effectiveness evaluation since April 2019 (in Japanese). Accessed July 13, 2025. <https://c2h.niph.go.jp/results/item.html>
16. Richardson J, Iezzi A, Maxwell A, Chen G. Does the use of the proportional shortfall help align the prioritisation of health services with public preferences? *Eur J Health Econ.* 2018;19(6):797-806. doi:10.1007/s10198-017-0923-5
17. Wailoo A. Technical support document 23: a guide to calculating severity shortfall for nice evaluations Accessed July 13, 2025. <https://www.sheffield.ac.uk/media/57671/download>
18. Ministry of Health, Labour and Welfare. Abridged life tables for Japan, 2019. Accessed July 13, 2025. <https://www.mhlw.go.jp/english/database/db-hw/lifetb19/dl/lifetb19-06.pdf>
19. Shiroiwa T, Noto S, Fukuda T. Japanese population norms of EQ-5D-5L and Health Utilities Index Mark 3: disutility catalog by disease and symptom in community settings. *Value Health.* 2021;24(8):1193-1202. doi:10.1016/j.jval.2021.03.010
20. Center for Outcomes Research and Economic Evaluation for Health. Guideline for preparing cost-effectiveness evaluation to the central social insurance medical council (version approved on January 17, 2024). Accessed July 13, 2025. https://c2h.niph.go.jp/tools/guideline/guideline_en_2024.pdf
21. Wailoo A. Severity weights in NICE technology appraisals. Accessed July 13, 2025. <https://sheffield.ac.uk/nice-dsu/methods-development/severity-weights>

22. Tanaka S, Seto K, Kawakami K. Pharmacoepidemiology in Japan: medical databases and research achievements. *J Pharm Health Care Sci.* 2015;1:16. doi:10.1186/s40780-015-0016-5
23. Yasunaga H. Updated information on NDB. *Ann Clin Epidemiol.* 2024;6(3):73-76. doi:10.37737/ace.24011
24. National Institute for Health and Care Excellence. Review of methods for health technology evaluation programmes: proposals for change. Accessed August 19, 2025. <https://qna.files.parliament.uk/qna-attachments/1364167/original/NICE%20Review%20of%20Methods%20-%20Proposals%20for%20Change.pdf>
25. Iwatani T, Inoue E, Tsugawa K. Validation of the predictive accuracy of health-state utility values based on the Lloyd model for metastatic or recurrent breast cancer in Japan. *BMJ Open.* 2021;11(12):e046273. doi:10.1136/bmjopen-2020-046273
26. Angelis A, Harker M, Cairns J, et al. The evolving nature of health technology assessment: a critical appraisal of NICE's new methods manual. *Value Health.* 2023;26(10):1503-1509. doi:10.1016/j.jval.2023.05.015
27. Association of the British Pharmaceutical Industry. NICE methods and process review outcome: key messages. Accessed July 13, 2025. https://www.abpi.org.uk/media/uuyhdfym/nice-methods-process-review-post-consultation-key-messages-february-2022_final.pdf