CORE2 HEALTH

Guideline for Preparing Cost-Effectiveness Evaluation to the Central Social Insurance Medical Council

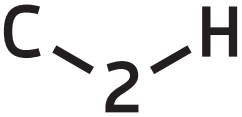
—

This version was approved by CSIMC on 17th January, 2024

The English version is a translation of the original version in Japanese. The Japanese version is preferentially applied in cases of discrepancy between the two versions.

Center for Outcomes Research and Economic Evaluation for Health, National Institute of Public Health (C2H)

Contents



1 Objectives

1.1 This guideline presents the standard methods to perform cost-effectiveness evaluations of medicines, medical devices, and regenerative medicine products selected by the Central Social Insurance Medical Council (selected product).

1.2 This guideline is applied to manufacturers' submissions and Academic Technology Appraisal Groups' analysis (review of submission and academic analysis).

2 Perspective

2.1 The perspective of the analysis should be specified. In particular, the perspective determines the costs to be included.

2.2 "Public healthcare payer's perspective" is the standard setting. It uses costs, comparators, and target populations to reflect the public healthcare insurance situation in Japan.

2.2.1 Even when an analysis is conducted from a perspective other than the "public healthcare payer's perspective," an analysis from the "public healthcare payer's perspective" should also be submitted.

2.2.2 There are products that are not covered by public healthcare insurance but are publicly funded, such as prophylactic procedures (e.g., health checkups and vaccinations). Analyses that consider these products can be submitted from the "public healthcare payer's perspective."

2.3 If consideration of public long-term care costs is important for the selected product, it is acceptable to submit an analysis from the "public healthcare and long-term care payer's perspectives."

2.4 If the introduction of a selected product influences productivity, then it is acceptable to submit an analysis that includes productivity losses from a broader perspective.

3 Target population

3.1 Patient groups for whom the selected product had indications at the time designated as an item for cost-effectiveness evaluation should be considered as the "target population."

3.1.1 In the case that a new indication is approved between the designation of the product and determining the framework of analysis, these patients are also included in the target population for cost-effectiveness evaluation.

3.1.2 New indication is added after the time when "3.1.1" is defined, and further evaluation is performed after the first evaluation is completed if it may influence the initial results.

3.2 If the product has multiple indications or subpopulations that differ in outcomes, doses, administration methods, or comparators, an analysis should be conducted for each population.

3.2.1 However, if "3.2" is difficult to achieve and the impact is limited, a part of the population can be omitted from the evaluation, considering the number of patients or disease characteristics. The exemption is determined based on an agreement between the manufacturer and the National Institute of Public Health/Academic Technology Appraisal Group in consultation.

3.3 The proportion of patients in each population should be estimated from a long-term perspective (based on the cumulative number of patients during the patent period). The current clinical status, utilization patterns, and other empirical data on the selected products should be considered.

3.3.1 If it is difficult to estimate, the proportion of patients may be calculated at a certain cross-sectional time point where the utilization pattern becomes stable.

3.3.2 Short-term negligible effects should be excluded (e.g., the impact of elective patients just after the selected product is introduced to the market).

4 Comparator

4.1 The comparator should be selected from candidates that are widely used in clinical practice and expected to be replaced by the selected product when the product is newly introduced to treat the target population. Among them, the product that results in the best outcome should be selected.

4.1.1 Products that are "widely used in clinical practice" are not decided only by the amount of actual consumption; rather, the products are used as a clinical standard, as specified in the clinical guidelines.

4.1.2 When deciding whether a product is "the best outcome," refer to the report on additional benefit published in the cost-effectiveness evaluation process.

4.2 If a single comparator cannot be determined based on "4.1," the selection of a comparator should be considered following the comparators in randomized controlled trials (RCTs), similar products for the official pricing and cost-effectiveness of candidate products, based on agreement in consultation.

4.3 Non-treatment or watchful waiting can also be used as a comparator.

4.4 Except for the cases described in "4.3," a comparator should be selected from products that are reimbursed by public healthcare insurance.

4.5 The reasons for the selection of a comparator should be sufficiently explained.

5 Additional benefits

5.1 Whether the selected product has additional benefits to the comparator should first be evaluated using empirical data.

5.2 The additional benefits to the comparator, as defined in Section 4, should be evaluated through a systematic review (SR) of RCTs. Results of unpublished clinical studies and trials may also be included in the SR.

5.2.1 When an SR is conducted, research questions (RQs) must be clearly presented. For example, RQs structured according to PICO (P, patient; I, intervention; C, comparator; and O, outcome) should be provided.

5.2.2 There may be products with action mechanisms (drugs) or function categories (medical devices) similar to the selected technologies or the comparator, which are expected to show equivalent outcomes. These products can also be included as interventions (I) or comparators (C) in the SR.

5.2.3 As outcome (O) in "5.2.1," the most appropriate clinical outcomes (e.g., a "true outcome") should be selected from clinical effectiveness, safety, and health-related quality of life (HRQOL).

5.2.4 A description of the inclusion/exclusion criteria, databases, search strategy formula, and research selection process (inclusion flow diagram) is required, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

5.2.5 It is acceptable to use existing reliable SR. In such cases, an existing review is used either directly or in combination with a new literature search. The existing literature is consistent with the RQs and covers the most recent literature.

5.2.6 If appropriate, the pooled results from pairwise meta-analyses should be presented. In such cases, the statistical method, heterogeneity, forest plots, pooled results, and confidence intervals should be reported.

5.2.7 A time point between determining the framework of analysis and the manufacturer's submission should be used as the cutoff date for the literature search in the SR.

5.2.8 There may be cases in which the results of new clinical trials are published after the cutoff date defined in "5.2.7." If evidence is important for cost-effectiveness evaluation (e.g., clinical trials with large sample sizes or reliable results different from current studies), the inclusion of these trials in the SR should be considered through academic technology appraisal group analysis.

5.3 When no studies are available based on the result of the SR described in "5.2," additional benefits are evaluated through an SR of comparative non-RCT (e.g., observational) studies based on "5.2."

5.3.1 More various biases tend to occur in the non-RCTs than in the RCTs. Sufficient explanation of research quality is required (e.g., study design, differences in background factors between groups, statistical analysis, sample sizes, and number of institutions).

5.3.2 Research quality is not necessarily high in studies that retrospectively use large-scale databases (real-world data), such as health insurance claims and registries. Therefore, adequate explanation should be given regarding the characteristics of the database, differences in the healthcare environment between Japan and other countries, definitions of variables and events, validity of the definition, validation survey on the definition, analytical methods, and extrapolation of the results to Japan.

5.4 As a general rule, the results of non-RCTs should be used to complement the results of an RCT to evaluate the benefits of the product.

5.5 When there are no RCT studies using the same comparator selected in Section 4 but there are RCT studies that compare the selected product to others, additional benefits can be evaluated through indirect comparison using SR results.

5.6 When there are only single-arm clinical studies on selected technologies, an indirect comparison should be performed based on the SR results of the selected technologies and the comparator.

5.7 When indirect comparison is performed, the following items should be considered:

5.7.1 When individual patient data are available, the difference of background factors should be adjusted using appropriate method, such as matching adjusted indirect comparison (MAIC).

5.7.2 A network meta-analysis (NMA) can be conducted when patient-level data cannot be used or when an NMA is a preferable method.

5.7.3 For analyses in "5.7.1" and "5.7.2," an anchored method is desirable.

5.7.4 Sufficient explanation should be given on the assumptions for indirect comparisons (e.g., heterogeneity of disease, severity and demographics, and the similarity of studies).

5.7.5 A naïve indirect comparison may be used when comparative study results are not available and when there is no other choice; in this case, intergroup comparability should be further explained in detail.

5.7.6 If more than one analytical method can be used, the reasons for selecting the methods (e.g., suitability of the assumption) should be explained; when necessary, sensitivity analyses with different methods should be performed.

5.7.7 If an NMA is performed, the following points should also be considered and explained. It is recommended to perform a sensitivity analysis as necessary.

(a) Network size (number of treatment groups)

(b) Treatment groups (drug classes or individual products).

(c) Handling different doses and modes of administration

(d) Handling drugs not approved in Japan

5.7.8 When a network meta-analysis is performed, data and programs reproducible by third parties should be submitted if possible.

5.8 Literature information should be presented and used to evaluate additional benefits (e.g., those identified in the SR or incorporated into an NMA).

5.9 There may be cases in which the detected studies based on "5.3" to "5.7" have serious problems with insufficient quality. However, the selected product is not expected to be inferior to the comparator. In such cases, the analysis described in Section 6 can be performed, assuming that the outcome of the selected product is equivalent to that of the comparator.

5.10 When there are no available clinical data on the selected product in humans, the analysis described in Section 6 can be performed, assuming that the outcome of the selected product is equivalent to that of the comparator, if appropriate. It is assume that products approved by the Pharmaceuticals and Medical Devices Agency (PMDA) have such an outcome.

5.11 When SR results obtained using the methods in "5.2" to "5.7" show that outcomes of the selected product are inferior to that of the comparator, no cost-effectiveness analysis is performed.

6 Methods of analysis

6.1 A cost-effectiveness analysis should be conducted. In this guideline, cost-effectiveness analysis is defined as an analysis that calculates cost and effectiveness separately, without converting effectiveness into monetary units.

6.2 If the evaluation described in Section 5 reveals additional benefits, the incremental cost-effectiveness ratio (ICER) should be calculated from the expected cost and effectiveness in each treatment group.

6.3 In the following cases, only the expected cost and effectiveness in each group need to be presented. The ICER should not be calculated.

6.3.1 In cases where the product has better effectiveness and lower cost, the product is considered "dominant."

6.3.2 A cost comparison (the so-called cost minimization analysis or CMA) should be performed if the evaluation described in Section 5 does not demonstrate additional benefits. In such case, the results are "cost saving" (or equivalent) or "cost increase."

6.3.3 Even if an additional benefit is demonstrated by the evaluation in Section 5, the ICER may have large uncertainty. Minor changes in the parameters can manifest as significant changes in the ICER when the incremental effectiveness is only slightly positive and the incremental cost is almost zero. In these cases, the results are "equivalent in terms of cost and effectiveness".

6.4 If the selected product has multiple indications or subpopulations defined in "3.2" and/or "3.2.1," the ICER should be calculated for each indication or subpopulation.

6.5 If a cost-effectiveness analysis of a selected product published in an academic journal or reports by a health product assessment (HTA) agency are available, the results should also be presented.

7 Time horizon

7.1 The time horizon should be sufficiently long to evaluate the influence of the product on cost and effectiveness.

7.2 The same time horizon should be applied for cost and effectiveness.

7.3 The reason for setting this time horizon should be specified.

8 Choice of outcome measure

8.1 Quality-adjusted life years (QALY) should be used in principle.

8.1.1 When it is difficult to calculate the QALY and CMA is applied, other outcome measures can be used, if appropriate.

8.2 When the QALY is calculated, the QOL scores (utilities) should be measured using a preference-based measure (PBM).

8.2.1 The QOL scores should reflect the preference of the general population in Japan.

8.2.2 The Japanese version of the EQ-5D-5L is recommended as the initial choice for the PBM.

8.3 When the QOL scores are measured by the PBM, responses of the patients should be used.

8.3.1 Proxy responses (e.g., a family member or caregiver) may be used when the patients cannot respond.

8.3.2 Proxy responses from a healthcare professional should be avoided because of possible discrepancies from patients' responses.

8.4 If the QOL scores that satisfy "8.2" is available, the use of Japanese scores are preferentially recommended.

8.4.1 If Japanese research is insufficient but high-quality research is performed in other countries, the use of data collected in other countries is acceptable.

8.4.2 PBM data collected in other countries should be handled in a manner consistent with "8.2.1"; for instance, the Japanese value set is applied to these responses.

8.4.3 If patient-level data are not available, "8.4.2" is difficult to perform. If the results are significantly affected, explain whether the QOL scores are consistent with the Japanese scores. When necessary, consider adjusting the scores to the Japanese scale.

8.5 If data corresponding to "8.2" are unavailable, mapping of other appropriate HRQOL data is acceptable.

8.5.1 If mapping is implemented, the QOL scores obtained from the mapping function should be consistent with "8.2.1."

8.5.2 If there is insufficient conceptual overlap between the QOL instruments, consider whether the use of mapping is appropriate.

8.5.3 The development process and characteristics of the mapping formula should be reported in detail using, for instance, the MApping onto Preference-based measures reporting Standards (MAPS) checklist.

8.5.4 It is not recommended to use a mapping function on which details cannot be reported.

8.6 If it is difficult to collect QOL scores directly from patients, it is acceptable for the general population to evaluate the assumed health scenario using a standard gamble (SG), time trade-off (TTO), and discrete choice experiment (DCE) (the vignette method).

8.6.1 It should be noted that the QOL scores measured using the vignette method are significantly affected by the assumed health scenario provided to respondents.

8.6.2 Since the assumed health scenario is not obtained directly from the patients, its validity should be sufficiently explained. It is also desirable to seek input from patients and/or clinical experts.

8.6.3 If the results from the vignette method are used, the actual scenario shown to respondents should be presented.

8.6.4 Considering "8.6.1," scores obtained from the vignette method may lead to overestimation of cost-effectiveness, the limitations of these methods must be considered. For example, the same QOL score should be used for the same health state.

8.6.5 Web-based SG and TTO surveys are known for systematic biases in results; therefore, face-to-face surveys are recommended.

8.6.6 Scores using the vignette method present difficulties in investigating the differences between Japan and other countries. It is relatively easy to collect data; therefore, the use of Japanese scores is recommended.

8.7 In the case of analysis from the public healthcare and long-term care payer's perspectives, the QOL scores influenced by informal caregivers may be considered if actual data exist.

8.8 It is recommended to consider the QOL data with similar quality in the following order based on the principle in "8.2." When this order is not applicable, the reasons for this should be explained.

(a) Data collected using the Japanese EQ-5D-5L ("8.2").
(b) Data collected using a generic Japanese PBM with a Japanese value set other than the EQ-5D-5L.
(c) Data collected using a condition-specific Japanese PBM with a Japanese value set.
(d) Data collected in other countries using a PBM with patient-level data available; the priority of instrument is based on the concept from (a) to (c) ("8.4").
(e) Data collected in other countries using a PBM with patient-level data unavailable ("8.4").
(f) Data converted to QOL scores using a mapping function; for which, follow (a) to (e) in terms of the mapped instruments and mapping formula ("8.5").
(g) Data collected using the PBM with no Japanese value set.
(h) Data collected using the vignette method ("8.6").
(i) Other methods are not recommended unless there is an unavoidable reason and/or scientific justification.

The high-quality mapping formula developed in Japan may be preferable to (f).

8.9 The principles in "8.2," "8.3," and "8.8" shall not be applied to the QOL in children and adolescents. Particularly regarding the issue in "8.3," it will depend on the situation as to who should respond to the PBM.

8.9.1 The method and reason for the selection of a QOL instrument in children and adolescents should be explained.

9 Sources of clinical data (except costs)

9.1 When estimating outcomes and costs for the cost-effectiveness analysis, the treatment process used in the analysis should be presented together with its rationale.

9.1.1 The above treatment process should reflect the standard practice in Japan.

9.2 Calculations of the ICER should preferentially use effectiveness, safety, and QOL data (including parameters such as transition probability for model analysis) derived from high-quality research, with a high level of evidence reflective of clinical practice in Japan.

9.2.1 The selection of effectiveness, safety, and QOL data based on the SR is recommended. This review will also include unpublished clinical study/trial data, if appropriate.

9.2.2 Data with a high level of evidence should be preferentially used. Consideration of research quality, target population, and external validity is also recommended when applying data (e.g., it is possible that the results of an RCT may differ markedly from practical clinical results).

9.2.3 Data using additional analysis of existing studies and/or registry data can be used for this evaluation. In such cases, detailed information regarding the patient's background and statistical methods should be provided.

9.3 Japanese data should be used preferentially if there is evident heterogeneity between Japanese data and those of other countries.

9.4 Pooled data of both groups should be applied, if there is no statistically significant difference between groups.

9.4.1 If the same parameter values are not used in both groups owing to a lack of statistical power, show other supportive data and reasons as well as the size of the treatment effect (whether clinically important) and explain that the difference is interpretable.

9.5 If there are reliable and quantitative data on medical devices, analysis reflecting "learning effect" (i.e., improvement of treatment effect through the accumulation of clinicians' experience) or "product improvement effect" can be submitted in addition to analysis not considering the effects, upon agreement in consultation.

10 Calculation of healthcare costs

10.1 Only public healthcare costs should be included in the analysis from the public healthcare payers' perspective.

10.2 Healthcare costs of each health state include only related costs directly affected by the selected product. Do not include unrelated costs.

10.3 Healthcare costs of each health state should reflect the average resource consumption and standard clinical practices in Japan.

10.4 Claims databases established in Japan, which reflect actual clinical practice based on "10.3", should be used to estimate the costs of each health state. Cases in which it is difficult to define health states using only information from claims data, or insufficient data have been accumulated in the database, were excluded.

10.4.1 Definition of each health state and its rationale is required when claims data are used for cost estimation.

10.4.2 The methods and rationale for the method to estimate costs (including handling outliers and unrelated costs) should be presented.

10.5 Micro-costing (e.g., by medical fee schedule) based on the definitions of the standard clinical process can be used if it is difficult to estimate the costs of each health state using a claims database, or if micro-costing is more appropriate.

10.5.1 In the application of micro-costing, the rationale for costing should be shown based on "10.2." In this case, the claims database can be used to identify relevant items and/or estimate the amount of resource consumption.

10.5.2 When micro-costing is used, the amount of resource consumption and unit costs should be reported separately.

10.5.3 For the estimation of resource consumption in "10.5.2," the injection products should be calculated by the number of vials rather than by patient dosages.

10.6 The estimation should include not only the costs of the selected product and comparator but also the costs of adverse events and related future events.

10.7 Public healthcare costs should include not only the portion of costs paid by the insurer but also those paid by the government and patients (i.e., total public healthcare expenses).

10.7.1 Based on the principle in "2.2.2," the analysis should include the costs of health checkups, vaccinations, or similar procedures that are funded publicly but not reimbursed by public healthcare insurance in Japan.

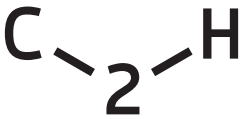
10.8 Unit costs should be derived from the latest medical fee schedule, drug price lists, or similar resources. It is particularly essential to use the latest unit costs for a selected product or comparator.

10.8.1 Even if existing cost-of-illness studies or analyses of claims data are used, the unit costs at the time of this evaluation and not at the time of the study should be applied. It is acceptable to make adjustments, such as multiplication by the revision rate of the medical payment system.

10.8.2 Such adjustments may be omitted if the influence on results is minimal.

10.9 If generics of a comparator are already on the market, analysis using these costs should be also submitted.

10.10 Even if the costs of a selected product and/or comparator are included in bundled payment, the estimation should be based on a fee-for-service payment.



10.11 Future costs should also be estimated on the basis of current resource consumption and unit costs.

10.12 Calculations of resource consumption based on data from other countries will require attention regarding the possible differences in healthcare technology use between Japan and other countries. The unit costs in Japan should be considered in the analysis.

11 Public long-term care costs and productivity loss

11.1 Analysis from the public healthcare and long-term care payer's perspectives can include public long-term care costs. It is acceptable to include public long-term care costs only if they can be estimated using Japanese data.

11.2 When public long-term insurance care costs are included in the analysis, it is recommended that they be calculated based on the care level.

11.3 The amount used under public long-term care insurance should be based on the actual resource consumption. If consumption is difficult to determine, it may be acceptable to use the average costs per beneficiary.

11.4 Analysis including productivity loss can be performed in addition to the base-case analysis. However, regarding the appropriateness of including productivity losses, the disease characteristics that lead to the possibility of working are considered. Including productivity losses in additional analyses is acceptable only if they can be estimated using Japanese data.

11.5 Decreases in productivity losses may be classified as follows:

(a) Decreases arising directly from selected products (e.g., treatment-related shortening of hospital stay).

(b) Decreases arising indirectly from outcome improvements (e.g., improvement in illness and survival extension).

Only (a) is included when productivity loss is included in the analysis.

11.6 Productivity losses should be estimated using the human capital method. This method estimates the loss using the expected earned wage in the absence of illness.

11.6.1 The unit wage used for estimations of productivity loss should be the average wage across all industries, ages, and both genders or the average wage for each age group in all industries and both genders derived from the latest "Basic Survey on Wage Structure" (Wage Census).

11.6.2 Estimations of productivity loss require an actual investigation of the employment status of the target population (i.e., a measure of days or hours of work missed). The actual measured days or hours should then be multiplied by the average wage across all industries, ages, and both genders to estimate the productivity loss.

11.6.3 If the method described in "11.6.2" is difficult to apply, productivity loss should be calculated by multiplying the expected number of days (excluding holidays) or hours of work missed by the average wage across all industries, ages, and both genders. A 100% employment rate should be assumed for those aged 18 years and older. However, this method may overestimate productivity losses.

11.7 If other individuals (e.g., family members) experience productivity losses because of informal care, it is acceptable to count these productivity losses under the same conditions and methods as those used to calculate the patient's productivity loss.

11.8 Time costs that are unrelated to a decrease in work should not be included.

12 Discounting

12.1 Future costs and effectiveness must be discounted and converted into present values.

12.1.1 Discounting is not required if the time horizon is one year or less or is otherwise sufficiently short to ignore the influence of discounting.

12.2 Both cost and effectiveness should be discounted at a rate of 2% per year.

12.3 The discount rate should be subjected to sensitivity analysis and should be changed at the same rate of 0%–4% per year for both costs and effectiveness.

13 Modeling

13.1 To predict the prognosis and future expenses, it is acceptable to conduct a model analysis using a decision analytic model, Markov model, and/or other models in accordance with the principles described in Section 7.

13.2 Validity of the model should be presented. For example:

(a) Internal validity: Why has a model with a given structure been created, whether the natural history of illness has been sufficiently evaluated, and whether the parameters are appropriate?

(b) External validity: Whether the estimation yielded by the model is appropriate in comparison with other existing clinical data.

13.3 The assumption used for the model should be clearly described.

13.4 All parameters and data sources used for model should be presented.

13.5 The model should be submitted using electronic files. The model must be easily understood by third-party experts, and all main parameters (transition probability, QOL score, and healthcare costs) must be modifiable.

13.5.1 It is better that not only the total costs but also the breakdown (in the case of microcosting, medical resource consumption, and unit costs) can be changed. The Academic Technology Appraisal Group must be able to change the unit costs of the selected product and the comparator.

13.6 Half-cycle correction should be used in the Markov model if the length of the Markov cycle is long and its influence is not negligible.

13.7 The following should be noted when using a model in which the ICER fluctuates probabilistically and the ICER does not always produce the same value, such as microsimulation.

(a) Set seed for the generation of random number to ensure that the results are reproducible.

(b) Demonstrate that the results do not vary significantly depending on the seeds. Models with large variability may not be acceptable if the results are difficult to use for decision-making.

(c) Random errors in the ICER arising from probabilistic variations, such as the use of extreme outliers, must be accepted, unless the results are arbitrarily adopted.

(d) If the results vary widely, it is recommended to use the mean values from multiple trials instead of single-trial results.

(e) Do not use a model that requires longer time for a trial with a normal-performance PC because it makes reviews difficult to conduct.

14 Uncertainty

14.1 If the analysis setting has multiple scenarios and could affect the results, scenarios analysis should be conducted.

14.2 In the case where the uncertainty is large, resulting from a long time horizon, a shorter-time analysis is necessary, such as when the time horizon is limited to periods when clinical study data are available.

14.3 If no head-to-head studies are detected according to Section 5, particularly when indirect comparison data between single-arm studies are used, a sensitivity analysis with a sufficiently wide range is required to deal with the large uncertainty.

14.4 Sensitivity analyses are required for parameters with large variances, those based on assumptions rather than actual data, and those with heterogeneity between Japan and other countries.

14.5 When the variance of the estimator should be considered (parametric uncertainty), the range in the sensitivity analysis can refer to the 95% confidence interval of the estimator.

14.6 The validity of the parameters used in a base-case analysis must be explained if they involve large uncertainties and have a significant impact on the results. The impact of such uncertainty on decision-making should also be examined.

14.6.1 A threshold analysis may be useful when examining how decision-making is affected by parameters with large uncertainty.

14.6 A probabilistic sensitivity analysis (PSA) is desirable. In such cases, the distribution of the parameters used for the analysis, scatter plots of the cost-effectiveness plane, and cost-effectiveness acceptability curves (CEAC) must be presented.

Terminology

Additional benefit

In a cost-effectiveness analysis, the additional benefit relative to the comparator should be demonstrated before calculating the ICER. The endpoint of effectiveness used to demonstrate the additional benefit does not always need to be equal to the outcome used for the cost-effectiveness analysis but should be clinically significant. If additional benefit is judged to be shown, cost-effectiveness analysis should be performed. On the other hand, if no additional benefit is shown, cost of both treatment should be compared by so called "CMA".

Cost-effectiveness analysis

Economic evaluations of healthcare technologies are often divided into the following patterns: (a) cost-minimization analysis (CMA), in which the outcome is deemed equivalent and only cost is analyzed; (b) cost-effectiveness analysis (CEA), which uses outcome units other than QALY (LY, event avoidance, etc.); (c) cost-utility analysis (CUA), which uses QALY; and (d) cost-benefit analysis (CBA), which involves an evaluation of outcomes after conversion into monetary units.

However, CMA, CEA, and CUA can all be considered analogous in situations where the cost and outcome are estimated in different units. For this reason, these types of analysis are collectively termed "cost-effectiveness analyses" in this guideline.

Discounting

In a cost-effectiveness analysis, a discount at a constant rate is usually made to convert future costs and arising (or obtained) outcomes to current values. Costs converted to the current value after applying yearly discounts (Cp) can be calculated from the cost at i years later (Ci) and the discount rate (d) using the following equation:

$$C_P = \frac{C_i}{(1+d)^{i-1}}$$

The same calculation can be used for effectiveness.

Dominant/dominated

If a technology is lower in cost and equivalent or higher in effectiveness than the comparator is, the technology is called "dominant." If the technology is higher in cost but equivalent or lower in effectiveness relative to the comparator, the technology is called "dominated."

Evidence level

Diverse classification methods for evidence levels are available. MINDS (Medical Information Network Distribution Service) set forth the following classification:

I	Systematic review/meta-analysis of RCTs
II	From one or more RCTs
III	From a non-randomized controlled study
IVa	Analytical epidemiological study (cohort study)
IVb	Analytical epidemiological study (case-control or cross-sectional studies)
V	Descriptive study (case reports or series)
VI	Views of an expert committee or individual experts that are not based on patient data

However, it has been often noted that the results of experimental studies such as randomized controlled trials (RCTs) can differ from real-world clinical data. Economic evaluations of healthcare technologies should primarily use data with a high level of evidence, although consideration should be given to appropriate clinical data.

Human capital method

The "human capital method" is used to estimate productivity loss based on the wages originally expected to be earned. However, when viewed from a long-term standpoints, the inability of an individual to work does not always lead to a productivity loss because in a situation with an employment rate less than 100%, as other individuals are likely to work in place of the affected individual who is unable to work. For this reason, one view suggests that productivity losses should include only friction costs (e.g., based on the period needed to restore the initial production level). Wages should be originally estimated through an investigation of the period for which an individual was actually unable to work because of illness. If this estimation is difficult due to lack of data including housework, it is acceptable to set the employment rate at 100%. From the viewpoint of fairness, the mean wage across all industries, all ages, and both genders should be used as the unit wage, regardless of the actual unit wage for individuals.

Incremental cost-effectiveness ratio

The incremental cost-effectiveness ratio (ICER) is the incremental cost divided by the incremental effectiveness. ICER of treatment A compared with B is calculated using the following equation:

$$ICER = \frac{IC}{IE} = \frac{C_A - C_B}{E_A - E_B}$$

IC: incremental cost
IE: incremental effectiveness
C_A: expected cost of treatment A
C_B: expected cost of treatment B
E_A: expected effectiveness of treatment A
E_B: expected effectiveness of treatment B

ICER is an indicator of the cost to acquire one unit of effectiveness. A lower ICER indicates higher cost-effectiveness.

Indirect comparison

When clinical studies yield results for "A vs. B" and "A vs. C," an estimation of the results for "B vs. C" in which no direct comparison is available from the head-to-head results is called an "indirect comparison." If no head-to-head study involving an appropriate comparator is available, an indirect comparison may occasionally be used.

The following conditions must be satisfied to enable indirect comparison: the results for "A vs. B" must also be applicable to the population "A vs. C" and the results for "A vs. C" must also be applicable to the population "A vs. B." This is called an "assumption

of similarity." When an indirect comparison is performed, it is necessary to test this assumption and to use appropriate statistical methods (for example, adjusted indirect comparison rather than naïve indirect comparison). This approach also enables analyses based on more advanced methods such as network meta-analyses (or multiple treatment comparisons; MTCs).

Mapping

When preference-based measure-determined QOL scores are unavailable, it is sometimes advantageous to use PRO data to calculate the QOL score used for cost-effectiveness analysis. The conversion of scores between measures is called "mapping." Mapping is acceptable as a second-best method when no other data are available but should be performed only after sufficient assessment of the statistical validity.

Meta-analysis

Meta-analysis is a method by which the results from a systematic review are integrated statistically to yield integrated values or their confidence intervals. If the heterogeneity is small, a fixed-effect model is usually used. If the heterogeneity is large, random-effect or Bayesian models are usually employed. The results are often depicted as forest plots. If a comparison is made among multiple treatments rather than between two treatments (pairwise comparison), a "network meta-analysis" is used, employing different methods (ref. "Indirect comparison").

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) is a technique used to determine the distributions of incremental cost, incremental effectiveness, and ICER by applying model parameters to the distribution. The results of a PSA are usually shown as a scatter plot on the cost-effectiveness plane and as a cost-effectiveness acceptability curve (CEAC), defined as $f(\gamma) = Pr(\gamma \cdot IE - IC > 0)$ (IC: incremental cost, IE: incremental effectiveness, γ: willingness to pay).

Productivity loss

Depending on the perspective, a loss resulting from the inability to perform work/housework because of illness (or benefit from early recovery) may be counted as a cost (i.e., productivity loss) but is not included in the base-case analysis. It is acceptable to consider not only the loss experienced directly by the patient but also losses experienced by family members or others arising from the need to provide nursing or informal care. According to this guideline, however, an indirect productivity loss resulting from an improvement in the patient's health states (e.g., survival period extension) is not included in productivity loss to avoid double counting (i.e., counting a factor as both effectiveness and costs). Only a productivity loss directly attributable to the healthcare technology (e.g., shortened hospital stay) is permitted for inclusion.

Quality-adjusted life year

A quality-adjusted life year (QALY) value is calculated by multiplying the life years (LYs) by the QOL score. A QOL score of 1 indicates full health, whereas 0 indicates death. If an individual has survived for two years under a health states with a QOL = 0.6, the LY is two years and the QALY is $0.6 \times 2 = 1.2$ (equivalent to 1.2 years survival under full health). If the QOL score changes over time, the QALY is represented by the area under the curve of the QOL score over time, as illustrated in the figure below.

Quality of life (QOL) score (utility)

The health states (i.e., value obtained from the health states) is scored using a one-dimensional scale ranging from 0 (death) to 1 (full health). Negative scores, reflective of a health states "worse than death," are also possible.

QOL scoring methods can be categorized as follows: (1) direct methods that evaluate health states under a hypothetical situation (or about himself/herself), including the standard gamble (SG) and time trade-off (TTO) methods, and (2) indirect methods that calculate QOL scores from patients' responses to QOL questionnaires using a scoring algorithm.

The QOL score used for cost-effectiveness analysis cannot always be calculated from any patient-reported outcome (PRO) or QOL data. Cost-effectiveness analysis can utilize only QOL scores determined using a preference-based measure developed for QALY calculation, as described below.

The EQ-5D (EuroQol 5 dimension) is one currently available measure for which a scoring algorithm has been developed in Japan.

Patient-reported outcome (PRO)

- **Symptom scale:**
  Measure for evaluating patients' own subjective symptoms, etc.
- **Health-related quality of life (HRQOL)**
  - **Health profile measure:**
    Assessing the QOL of patients (primarily multi-dimensional assessment)
    - **Generic:**
      Applicable to all diseases (e.g., SF-36)
    - **Disease-specific:**
      Applicable to specific disease (e.g., FACT and EORTC for cancer)
  - **Preference-based measure:**
    Allowing measurement of QOL score that can be used to calculate QALYs (e.g., ED-5D, SF-6D, and HUI)
- **Other PRO scales**

Sensitivity analysis

When uncertainty is present, its influence on the results can be evaluated by changing the parameter in a "sensitivity analysis." Sensitivity analyses can be further classified as a one- (only one parameter is changed) and two-dimensional (two parameters are simultaneously changed) sensitivity analyses, as well as PSA (simultaneous uncertainty in multiple parameters; see "Probabilistic sensitivity analysis").

Systematic review

A systematic review is a method by which the literature is comprehensively searched regarding a specific topic and the results are evaluated/reported without bias if at all possible. This method was defined by MINDS as follows: "When defined from the aspects of practical actions, systematic review means 'searching studies on a given clinical question comprehensively, grouping studies of identical quality on each research design and analyzing/integrating them being accompanied by evaluation of biases'."

A systematic review is often confused with meta-analysis. The results of a systematic review do not always require statistical integration; this type of systematic review is also known as a "qualitative systematic review." In cases where the integration of results is deemed appropriate, a meta-analysis of the systematic review results is needed.

Regarding the reporting style for a systematic review (meta-analysis), the style presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement has been used as a standard and can be used as a reference.

Uncertainty

Various types of uncertainty accompany cost-effectiveness analyses.

Broadly, heterogeneity is a type of uncertainty that indicates a situation lacking uniformity in terms of the comparator, healthcare

patterns, targeted patients, and other factors. This differs from the uncertainty in the narrow sense, as explained below. This is not a technical problem related to statistics or health economics but rather arises from real-world variety. If such heterogeneity is present, a sensitivity analysis based on multiple scenarios is recommended.

Uncertainty in the narrow sense can be divided into (a) model and (b) parameter uncertainties. Model uncertainty can result from (a)-1 methodological uncertainty and (a)-2 model structure/assumptions.

Methodological uncertainty, mentioned in (a)-1, arises from the theoretical impossibility of setting uniform methods for the estimation of the discount rate and productivity loss, measuring the QOL score, and other parameters. To avoid this type of uncertainty, it is important to conduct an analysis in accordance with common and standard procedures. If results such as the discount rate are markedly affected, uncertainty should be evaluated through one-way sensitivity analysis.

Uncertainty arising from the model structure/assumption, as mentioned in (a)-2, is caused by the method used to model the health states and treatment processes, selection of parameters for incorporation into the model, assumptions regarding predictions of long-term prognosis beyond the observation period, and other factors. This uncertainty should be evaluated in a sensitivity analysis.

Parameter uncertainty, as mentioned in (b), arises from uncertainty inherent in the parameter estimation. For example, if 10 of 100 subjects develop events during a clinical study, the true incidence rate might not be 10/100 = 0.1 in the whole population. To deal with this type of uncertainty, which is attributable to statistical inference, it is useful to conduct a PSA in addition to a deterministic sensitivity analysis.

Unrelated medical costs

Medical costs can be divided into related (i.e., those directly affected by the selected technology) and unrelated (i.e., those affected indirectly through survival extension or those not related to the illness). For example, a hypertension treatment that reduces the incidence of cardiovascular disease and stroke will extend life expectancy, possibly leading to an increase in unrelated medical costs (e.g., costs related to dementia, diabetes, and hemodialysis). These unrelated costs are not included in the cost.

Vignette method

This method is used to evaluate QOL scores in which respondents read a hypothetical health scenario (vignette) so that they can imagine their health state. SG, TTO, DCE, and other direct methods are used for evaluation. Careful consideration is required to avoid creating an arbitrary health scenario because it significantly affects QOL scores. The health scenario is not directly obtained from the patients.

Abbreviations

CBA:
Cost-benefit analysis

CEA:
Cost-effectiveness analysis

CEAC:
Cost-effectiveness acceptability curve

CMA:
Cost-minimization analysis

CSPBM:
Condition specific preference-based measure

CUA:
Cost-utility analysis

DCE:
Discrete choice experiment

EQ-5D:
EuroQol 5 dimension

HRQOL:
Health-related quality of life

ICER:
Incremental cost-effectiveness ratio

MAIC:
Matching-adjusted indirect comparison

MTC:
Multiple treatment comparison

PBM:
Preference-based measure

PRO:
Patient-reported outcome

PRISMA:
Preferred Reporting Items for
Systematic Reviews and Meta-Analyses

PSA:
Probabilistic sensitivity analysis

QALY:
Quality-adjusted life year

RCT:
Randomized controlled trial

RQ:
Research question

SG:
Standard gamble

SR:
Systematic review

TTO:
Time trade-off