07 Time horizon

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

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

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

08 Choice of outcome measure

8.1 The quality-adjusted life year (QALY) should be used as a basic outcome. Other outcome can be used depending on the characteristics of the illnesses, drugs, and/or medical devices.

8.1.1 When QALY is used, life year (LY) should also be presented, if the healthcare technology has effects on survival.

8.1.2 As a rule, if QALY is not selected as an outcome unit, appropriateness must be discussed through a preliminary consultation that considers the characteristics of the drugs, medical devices, or other factors.

8.2 When QALY is calculated, the QOL score should be reflective of the value for a general population using questionnaires (EQ-5D, SF-6D, HUI, etc.), the standard gamble (SG) method, and the time trade-off (TTO) method.

8.2.1 If Japanese QOL scores are newly collected for a cost-effectiveness analysis, the use of an instrument with a scoring algorithm developed in Japan is recommended.

8.2.2 If data corresponding to item 8.2 are unavailable, it is acceptable to use mapping of other appropriate patient-reported outcomes (PROs). When using a QOL score obtained from mapping, the conversion into a QOL score via an appropriate method should be explained.

8.3 When the QOL score is assessed, the subjects' own QOL responses should be used.

8.3.1 Answers from a proxy (e.g., family member, caregiver) may be used only when responses cannot be obtained from the subjects.

8.3.2 Proxy responses from a healthcare professional should be considered in light of possible discrepancies from subjects' own responses.

8.4 As long as a QOL score that satisfies item 8.2 and 8.3 is available, the use of Japanese results is preferentially recommended.

8.4.1 If Japanese research is absent or insufficient but high quality research is available overseas, it is acceptable to use the data collected overseas.

09 Sources of clinical data

9.1 Calculations of the ICER and other factors should preferentially use effectiveness, safety, and QOL data derived from high quality research, with a high evidence level reflective of practical clinical results.

9.1.1 The selection of effectiveness, safety, and QOL data on the basis of a systematic review of Japanese and overseas clinical research is recommended. This review may also include unpublished clinical study/trial data if deemed appropriate.

9.1.2 Data with a high evidence level should be used preferentially. The use of data deemed appropriate from the viewpoints of research quality, target population, and generalization is recommended (for example, it is possible that the results of a randomized controlled trial may differ markedly from practical clinical results).

9.2 Japanese data should be used preferentially if there is evident heterogeneity in effectiveness and safety between Japanese and overseas data with similar levels of research quality and evidence.

9.3 If the results of a single study are used although a systematic review has revealed the presence of multiple clinical studies, an explanation is needed regarding the reason for the selection of the single study.

9.4 If direct comparison data are unavailable or the available data are deemed unsatisfactory in terms of research quality or evidence level, it is acceptable to conduct an analysis through indirect comparison.

9.4.1 If an indirect comparison is conducted, refer to item 5.3.2.

10 Calculation of costs

10.1 The analysis should cover the following range of costs, depending on the perspective of the analysis.

	"Public healthcare payer"	"Public healthcare and long-term care payer"	Consideration of broader costs
Public healthcare costs	●	●	●
Long-term care costs		●	(●)
Productivity loss			●

10.2 The estimation should include not only the costs of the target technology and the comparator, but also the costs of factors such as adverse events and related future events.

10.3 Regarding costs such as those of the target technology and the comparator, the medical resource consumption and unit costs should be reported separately.

10.3.1 However, this provision does not necessarily apply to cases wherein the costs of adverse events and related future events are analyzed using the results of a claim analysis, existing cost-of-illness studies, or similar research.

10.4 An analysis of public healthcare costs should include not only the portion of costs paid by the insurer, but also those paid by the government and patients as copayment (i.e., total public healthcare expenses).

10.4.1 Depending on the situation, it may be acceptable to present an additional analysis that includes the costs of health checkups, vaccinations, or similar procedures that are funded publicly and not reimbursed by Japan's national healthcare insurance system.

10.5 Unit costs should be derived to the extent possible from the latest medical fee schedule, drug price list, or similar resources. It is particularly essential to use the latest unit costs for the target technology or the comparator.

10.5.1 Unit costs at the time of analysis, rather than at the time of actual medical resource consumption, should be used.

10.5.2 If the use of unit costs at the same time point is difficult (e.g., analyses

using data from past cost-of-illness studies, analyses of claims data), it is acceptable to make adjustments such as multiplication by the medical payment system revision rate. Such adjustments may be omitted if a sensitivity analysis has shown the influence on results to be minimal.

10.5.3 A scenario analysis using the prices of generic drugs should also be conducted if an influence on the results cannot be ruled out.

10.6 Unit costs of the targeted technology should be subjected to sensitivity analysis.

10.7 To enable a more appropriate evaluation of the influence of target technology introduction on other medical resource consumption, the estimation should be based on fee-for-service costs rather than diagnosis procedure combination (DPC)-based costs.

10.7.1 However, DPC-based costs may be used if a precise estimation of the costs is difficult (e.g., adverse events, related future events, and other factors) and the use of such expenses is unlikely to markedly affect the results.

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

10.9 Medical resource consumption should reflect the average quantity used (e.g., dose, body weight, height) and standard healthcare practices in Japan. If an appropriate reflection cannot be expected (e.g., overseas clinical study data, data from limited institutions), appropriate adjustment will be needed.

10.10 Costs include only related medical costs that are directly affected by the target technology, and do not include unrelated medical costs.

10.11 Calculations of medical resource consumption based on overseas data will require attention regarding possible differences in healthcare technology use between Japan and overseas countries. The unit costs in Japan should be reflected in the analysis.

11 Long-term care costs and productivity loss

11.1 Long-term care costs and productivity losses arising from an inability to perform work should not be included in the base-case analysis.

11.1.1 It is acceptable to include long-term care costs and productivity losses in additional analyses. However, judgments regarding the appropriateness of including productivity losses should take into account the possibility of working in the context of the illness characteristics, among other factors.

11.2 When long-term care costs are included in the analysis, these costs should be calculated based on the care level.

11.3 The amount utilized under public long-term care insurance should be based on the actual quantity of resources consumed. If this quantity is difficult to determine, it is acceptable to use the average amount utilized per beneficiary or similar data.

11.4 Decreases in productivity losses may be classified as follows:

(A) Decreases arising directly from healthcare technology (e.g., treatment-related shortening of hospital stay period);

(B) Decreases arising indirectly from outcome improvements (e.g., alleviation of illness, survival period extension).

When productivity loss is included in an analysis, only (A) should be included in the calculation of costs.

11.5 Productivity losses should be estimated using the human capital method. This method was designed to generate estimations based on the expected earned wage in the absence of illness.

11.5.1 The unit wage used for estimations of productivity loss should be the average wage across all industries, all 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), not to discriminate by income.

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

11.5.3 If the item described in 11.5.2 is difficult to perform, productivity loss should be calculated by multiplying the expected number of days (excluding holidays) or hours of work missed in the target population by the average wage across all industries, all ages, and both genders. A 100% employment rate should be assumed for those aged 18 years and older.

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

11.7 Time costs that are unrelated to a decrease in work should not be included in the cost estimations.

12 Discounting

12.1 Future costs and effectiveness should be discounted and converted into current values.

12.1.1 Discounting is not needed if the time horizon is 1 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 a sensitivity analysis and should be changed at a rate of 0–4% per year.

13 Modeling

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

13.2 When a model analysis is conducted, the validity of the model should be presented. For example:

(A) Internal validity: This addresses why a model with a given structure has been created, whether or not the natural course of illness has been sufficiently evaluated, whether or not the parameters used are appropriate, and other factors.

(B) External validity: This addresses whether or not the estimation yielded from the model is appropriate in comparison to other clinical data, and other factors.

13.3 The assumption used to create the model should be described clearly.

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

13.5 The model used and the calculation processes should be expressed in the form of electronic files and in a manner that can be understood by third-party experts and can change parameters.

14 Uncertainty

14.1 If the patterns of clinical practice or other factors are not uniform and this discrepancy could affect the results, analyses based on multiple scenarios should be conducted.

14.2 For situations wherein the uncertainty is large because of a long time-horizon, it is also necessary to perform a shorter-term analysis, such as an analysis of the period for which clinical study data are available.

14.3 If no available studies involve a comparison with the comparator according to the section section 5, particularly when a comparison has been made concerning results between single-arm studies, a sensitivity analysis with a sufficiently wide range is required because of the large uncertainty.

14.4 Sensitivity analyses are needed for parameters with large variances, those based on assumptions rather than actual data, those with possible heterogeneity between overseas and domestic data, and others.

14.5 A probabilistic sensitivity analysis is also desirable. In such a case, it is necessary to present the distribution used for analysis, scatter plots on the cost-effectiveness plane, and cost-effectiveness acceptability curves.

15 Reporting/publication

15.1 The results of the analysis should be reported in the style set (in Japanese) forth elsewhere.

15.2 The model and other parameters employed for the cost-effectiveness analysis should be submitted in the form of an electronic file, in accordance with item 13.5.

15.3 The analysis/review results should be made public. However, if some incorporated data are difficult to publish with regard to intellectual proprietorship protection, these data may be specified in advance. If the involved parties disagree over the extent of publication, the issue should be settled through discussions.

Additional benefit in effectiveness/safety

In a cost-effectiveness analysis, the additional benefit in effectiveness/safety relative to the comparator should be demonstrated before calculating the ICER. The endpoint of effectiveness used to demonstrate additional effectiveness/safety does not always need to be equal to the outcome unit used for the cost-effectiveness analysis, but should be clinically significant.

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 various outcome units (LY, event avoidance, etc.) other than QALY; (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 to each other in situations where the cost and outcome are estimated in different unit. For this reason, these types of analysis are collectively called ìcost-effectiveness analysesî in this guideline.

Discounting

In a cost-effectiveness analysis, a discount 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:

$$C_p = \frac{C_i}{(1+d)^{r-1}}$$

The same calculation can be used for effectiveness.

Dominant/dominated [extended dominance]

If a technology is lower in cost and equivalent or higher in effectiveness than the comparator is, the technology is called ìdominant.î If the technology is higher in cost but equivalent or lower in effectiveness relative to the comparator, the technology is called ìdominated.î

During an evaluation of multiple treatment technologies, there may be cases where Treatment 4 is located to the upper left (the area with higher ICER) of the straight line that joins Treatment 3 to Treatment 5, as illustrated below. Such a relationship is called ìextended dominance,î and in such cases there is no need to calculate the ICER for treatment technology 4, which is dominated by ìextended dominanceî. In this case, the ICER for Treatment 5 compared with Treatment 4 will always be smaller than that for Treatment 4 compared with Treatment 3. Therefore, if Treatment 4 is considered cost-effective, Treatment 5 is also cost-effective.

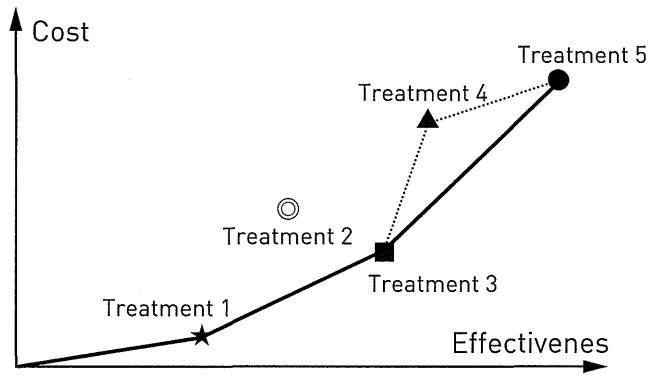

Figure: Dominated technology (Treatment 2) and extended dominance technology (Treatment 4)

	Effectiveness (QALY)	Cost (10,000 Yen)	Incremental effectiveness	Incremental cost	ICER (10,000 Yen/QALY)	
Treatment 1	1	50				
Treatment 2	1.5	200			Dominated	←(ICER not shown)
Treatment 3	2	150	1	100	100	←Compared to Treatment 1
Treatment 4	2.25	300			Dominated by extended dominance	←(ICER not shown)
Treatment 5	3	350	1	200	200	←Compared to Treatment 3

Table: ICERs of healthcare technologies 1 through 5

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**
IV a	**Analytical epidemiological study (cohort study)**
IV b	**Analytical epidemiological study (case-control study, cross-sectional study)**
V	**Descriptive study (case reports, case series)**
VI	**Views of expert committee or individual experts that are not based on patient data**

However, it has been often noted that the results from 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 ìhuman capital methodî is used to estimate productivity loss based on the wages originally expected to be earned. However, when viewed from a long-term perspective, 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%, other individuals are sure to work, instead of the 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, 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. A comparison of the ICER of treatment A with that of 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, CA: expected cost of treatment A, CB: expected cost of treatment B, EA: expected effectiveness of treatment A, EB: expected effectiveness of treatment B).

ICER is an indicator of the cost to acquire 1 unit of effectiveness. A lower ICER indicates higher cost-effectiveness.

Indirect comparison

When clinical studies yield results for ìA vs. Bî and ìA vs. C,î an estimation of the results for ìB vs. Cî (no direct comparison available from the head-to-head results) is called an ìindirect comparison.î 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 ìA vs. Bî must also be applicable to the population ìA vs. C,î and the results for ìA vs. Cî must also be applicable to the population ìA vs. B.î This is called an ìassumption of similarity.î 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 naÔve 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 useful to use PRO data to calculate the QOL score. This conversion of scores between measures is called ìmapping.î Mapping is acceptable as a second-best method when no other data are available, but should be performed only after a sufficient assessment of 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, the fixed-effect model is usually used. If the heterogeneity is large, the random-effect model or Bayesian model is usually employed. The results are often depicted in forest plots. If a comparison is made among multiple treatments rather than between two treatments (pairwise comparison), a ìnetwork meta-analysisî is used, and different methods are employed (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 $f(\gamma) = Pr(\gamma \cdot IE - IC \rightarrow 0)$ (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 patientís health status (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 2 years under a health status with a QOL = 0.6, the LY is 2 years and QALY is 0.6 × 2 = 1.2 (equivalent to 1.2-year 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.

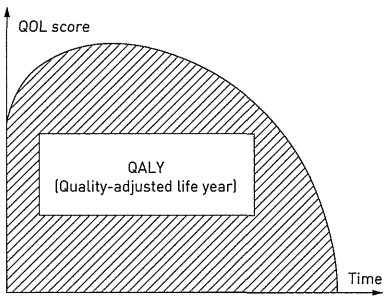

Figure: Conceptual diagram of QALY

Quality of life (QOL) score

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

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

The QOL score used for a cost-effectiveness analysis cannot always be calculated from a patient-reported outcome (PRO) or QOL data. A 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 of the currently available measures for which a scoring algorithm has been developed in Japan.

- Patient-reported outcome (PRO)
  - Symptom scale: Measure for evaluating patientsí 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 for 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 ìsensitivity analysis.î Sensitivity analyses can be further classified as a one-dimensional sensitivity analysis (only one parameter is changed), two-dimensional sensitivity analysis (two parameters are simultaneously changed), and PSA (simultaneous uncertainty in multiple parameters; see the ìProbabilistic sensitivity analysisî Section).

Systematic review

Systematic review is a method by which the literature is comprehensively searched about a specific topic and the results are evaluated/reported without bias if at all possible. This method was defined by Minds as follows: ìWhen defined from the aspects of practical actions, systematic review means ësearching 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 biasesí.î

Systematic review is often confused with meta-analysis. The results yielded from a systematic review do not always require statistical integration; this type of systematic review is also known as a ìqualitative systematic review.î 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 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 uncertainty and (b) parameter uncertainty. 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 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 a 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 status 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 medical costs (i.e., those directly affected by the target healthcare technology) and unrelated medical costs (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.

2. 費用データの分析方法に関する検討

1) 費用対効果評価のための標準的ツールおよびデータソースの確立：NDBを用いた費用分析

福田治久、池田俊也、赤沢学、白岩健、福田敬

1. 背景

費用対効果評価を実施する上で、健康状態別の医療費に関するデータが不可欠である。医療費は診療過程や診療報酬制度の影響を強く受けることから、『中央社会保険医療協議会における費用対効果評価の分析ガイドライン』において、「10.9 医療資源消費量は、日本における平均的な使用量(用法用量,体重,身長等)や標準的な診療過程を反映している必要がある。」や「10.11 海外データを用いる際には、資源消費量について、国内外における医療技術の使用実態等の違いに配慮する必要がある。単価は国内のものを反映させること」が求められている。しかしながら、本邦においては、費用対効果評価に利活用可能な健康状態別の医療費データが整備されていない。

本研究では、費用対効果評価の際に必要となる費用データについて、レセプト情報・特定健診等情報データベース(NDB)を用いて推計を行うことにより、基盤的なデータを整備することを目的に実施した。

2. 方法

(1) データベース

本研究で使用したデータは、レセプト情報等の提供に関する申出（提供依頼申出者：厚生労働省保険局医療課長　宮嵜雅則）により抽出された6疾患NDBである。平成21年4月から平成26年3月の5年間における医科、調剤、DPCのレセプト情報を使用した。

(2) 分析対象患者

解析対象となる疾患は、脳卒中、非小細胞性肺癌、虚血性心疾患、大腸癌、肝炎、糖尿病の6疾患である。平成21年4月から平成26年3月のいずれかの時点で、6疾患のいずれかに関する傷病名が一度でも出現するハッシュ値を抽出し、当該ハッシュ値に一致する、対象期間中の全レセプトデータを抽出した。したがって、定義した傷病名を有する全患者が分析対象患者となる。