https://doi.org/10.7175/693.ch7

7 Data Sources and Accounting for Uncertainty

Albert Wertheimer
PhD, MBA, Professor, Nova Southeastern University Ft. Lauderdale, FL, USA

Chapter reviewed by

- Umberto Restelli. *Center for Health Economics, Social and Health Care Management, LIUC, Università Cattaneo, Castellanza (VA), Italy. Honorary Senior Lecturer, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa*
- Giovanni Tripepi. *Italian National Research Council, CNR · Institute of Clinical Physiology IFC, First Researcher, Reggio Calabria, Italy*

Abstract

To perform cost-effectiveness and cost-utility analyses, data may be retrieved from Delphi panels, the literature, or conducting *ad hoc* studies.
Studies from the literature may be observational or experimental. Data coming from several studies may be combined in meta-analyses, systematic reviews, or network meta-analyses. Further data may come from administrative databases. Data are elaborated by means of the most suitable techniques, such as classical statistics or Bayesian approach.
Several types and subtypes of biases may arise during and after the conduct of a study, thus the most suitable countermeasures must be undertaken to reduce errors, e.g., randomization, case matching, and propensity scoring. In pharmacoeconomics studies, the weight of each variable is tested by means of simple sensitivity analysis, probabilistic analysis, analysis of extremes, or threshold analysis.
Finally, several guidelines may be of help in assessing the reliability of the published studies that could be used as data sources.

7.1 Data Sources

In a cost-effectiveness analysis, two kinds of data are needed: clinical data values (i.e., probability values for the chance nodes in the decision tree, incidence of adverse events- or pathology-related events, associated utility values) and data to account for the healthcare resources used and the monetary costs associated with the consequences (branches) stemming from each chance node. Suppose that a healthcare professional wishes to know the answer to a certain question, such as: Is drug X an effective treatment for dis-

ease Y in population Z? There are three ways to find an answer: The first is to ask someone; the second is to look the answer up in the literature; and the third is to perform a study designed to find out the answer. All three ways of finding answers are used in pharmacoeconomics.

The first method—asking someone who might know—is institutionalized in the form of the augustly named "Delphi Panel", which is a panel of experts convened to provide their collective opinion. The Delphi Panel is named after the oracle of Delphi, famed in classical Greece for its cryptic pronouncements that could be interpreted as prophetic only in retrospect. Unless supported by an explicit marshalling of factual data and analysis, mere opinion does not qualify as evidence-based medicine. The second and the third methods are discussed in this Chapter.

A Cost-Utility Analysis (CUA) is performed in the same way as a cost-effectiveness analysis except that the unit of effectiveness is Quality-Adjusted Life-Years (QALYs) or another measure of utility. This analysis is better described in Chapter 3.

Clinical Epidemiology

Pharmacoeconomic studies usually require data both on costs and effectiveness. The effectiveness data are taken from epidemiological or medical research studies. The design of every medical research study can be classified according to a few fundamental mutually exclusive dichotomies (Table 7.1).

First, a study may be either observational or experimental. A clinical trial is an experiment—a test of an intervention—in which the investigator interferes with the normal course of events, usually by providing certain people with a treatment they would not otherwise have received. In an observational study, the investigator merely observes events and does not interfere with them. There cannot be, therefore, an "observational trial". Second, in respect to the time in which the investigator decides to perform the study, the events being studied may have already occurred, in which case the study is a retrospective one, or may not yet have occurred, in which case they will

	Study design		
Investigator involvement	Observational		Experimental
Time perspective	Prospective	Retrospective	Prospective
Time sampling	Longitudinal or cross-sectional	Longitudinal or cross-sectional	Longitudinal

Table 7.1. Concepts in clinical study design.

be studied as they happen, i.e., prospectively. Third, the observations may refer to a point in time, and are called "cross-sectional", or to two or more points in time, and are then called "longitudinal". An observational study may be longitudinal or cross-sectional. In the first case, it may be either prospective or retrospective, whilst cross-sectional studies are neither prospective nor retrospective, because exposure and outcome are measured at the same time (they are compared to snapshots [Grimes, 2002]). An experimental study, however, can only be longitudinal because, however brief in duration it might be, it is an analysis of cause and effect—the "cause" being the intervention which is tested, and the "effect" being the clinical outcome that is observed, which cannot, by definition, occur at the same point in time. Thus, a clinical trial is a prospective, longitudinal, experimental study.

Literature Analysis

The U.S. state and federal governments make available basic demographic information that is often needed in pharmacoeconomic analyses—e.g., the death rate from all causes by age and gender, death rates by cause, etc. The Census Bureau publishes past and projected population demographic data. The National Center for Health Statistics also collects and makes available ongoing national survey data addressing a variety of subjects. These include the National Health Insurance Interview Survey, the National Health and Nutrition Examination Survey, and the National Health Care Surveys.

Much of the information required in a pharmacoeconomic analysis is present in studies published in the medical literature. There are several considerations in extracting data from the literature: determining sources of information, defining a search strategy, categorizing the studies identified, assessing the internal validity of individual reports, assessing the representativeness of the sample of studies identified, and determining the external validity of the data, i.e., whether the data that are valid in the context of the published study are applicable to the setting of the pharmacoeconomic analysis.

Very often, several studies can be identified that yield range of values for the variable of interest. There are two approaches to this situation. First, if there are one or two large, well-designed studies (such as randomized controlled trials designed and performed under FDA scrutiny), it is reasonable to consider this as "best evidence" and disregard larger numbers of smaller, statistically underpowered studies or studies with less rigorous designs. Second, the results of several studies can be combined in meta-analyses, systematic reviews, or network meta-analyses.

Systematic Review and Meta-analysis

A **systematic review** collects all empirical evidence to answer a specific research question through a clear and systematic methodology, with a view to minimizing bias and thus providing more reliable findings from which conclusions can be drawn. Systematic reviews may contain meta-analyses. The term **meta-analysis** is used in a number of different ways but, in essence, it means the statistical pooling of data from several studies. In particular, the meta-analysis requires studies that are conceptually homogenous in design, interventions, and endpoints in order to provide more precise estimates of the effects of healthcare relative to the results of individual studies included within a review [Higgins, 2011].

Two methods may be used to combine the results obtained by meta-analyses:

Example of Systematic Review and Meta-Analysis

An example of systematic review and meta-analysis is provided by an investigation of the potential benefit of ω-3 fatty-acid enriched Parenteral Nutrition (PN) vs standard (non-ω-3 fattyacid enriched) PN on nosocomial infection and mortality in adult hospitalized patients [Pradelli, 2020]. The Authors of the review included only randomized controlled trials published in English in peer-reviewed journals containing at least one predefined clinical outcome and laboratory parameters as reported in the study protocol. In total, 4495 publications were identified through searches on the main databases—such as MEDLINE, Embase, and Cochrane database—and 49 with at least one outcome of interest met inclusion criteria. Authors calculated a pooled risk ratio for the effect of ω-3 fatty-acid enriched PN on infections and mortality relative to standard PN. Finally, 24 studies (2,154 patients) reported any nosocomial infections: Compared with standard PN, ω-3 fatty-acid enriched PN significantly reduces the infection rate (RR 0.60, 95% CI 0.49-0.72; P < 0.00001), as shown in Figure 7.1. Simultaneously, there was a nonsignificant reduction in mortality rate (RR 0.84, 95% CI 0.65-1.07; P = 0.15), based on data reported by 22 studies (1,839 patients). This study, published in the *Journal of Parenteral and Enteral Nutrition*, is a good example of a meta-analysis performed under the procedures of a systematic review: Authors adhered to best practice, such as the prospective registration of methods in PROSPERO and reporting systematic reviews and meta-analyses according to the PRISMA statement. The statistical pooling was performed with studies that were conceptually homogenous in design, interventions, and endpoints. Purists might object, however, that the studies were heterogeneous in several ways and therefore not eligible for statistical pooling. This is an example of a meta-analysis that is also a systematic review in the tradition of evidence-based medicine.

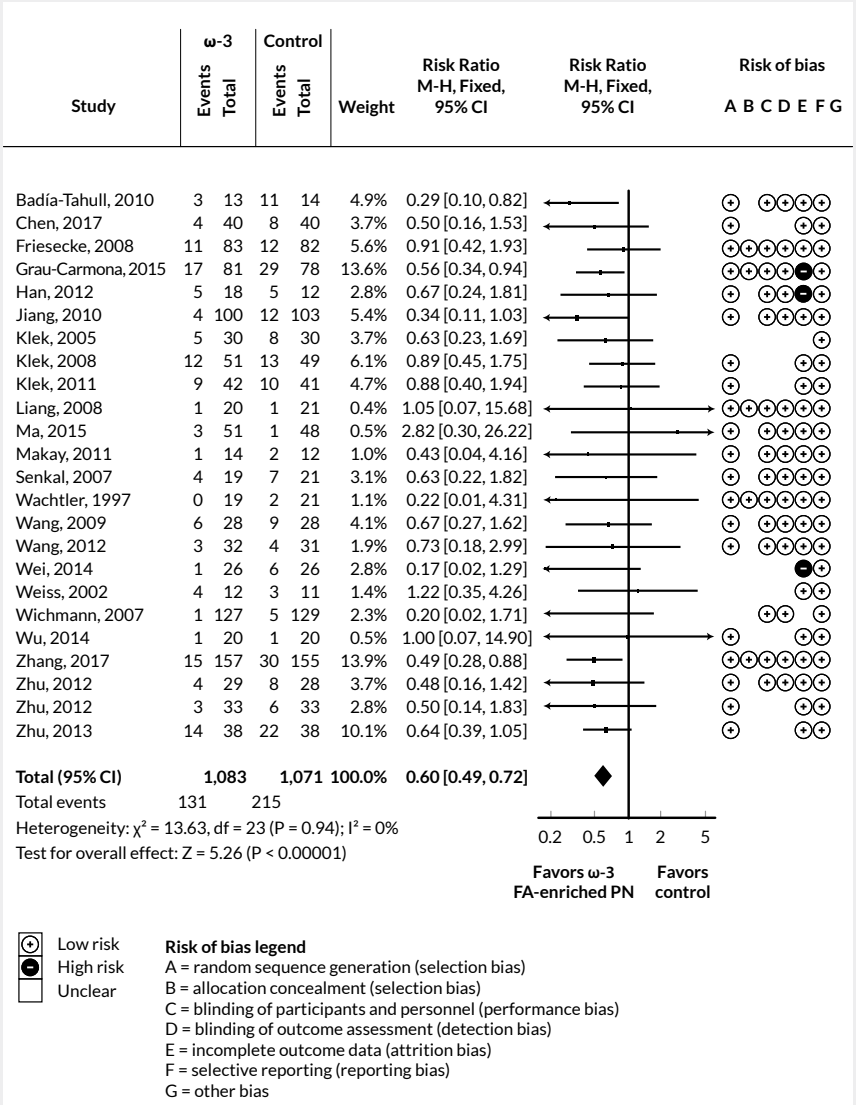

Figure 7.1. Infection rates. Forest plot of fixed effects meta-analysis showing individual study means, pooled estimates, and risk of bias for individual studies (Cochrane tool).

CI = Confidence Interval; FA = Fatty Acid; M-H = Mantel-Haenszel; PN = Parenteral Nutrition

- Fixed-Effect (FE) model, that assumes homogeneity among studies. It is used when the sampling variability (due to sample size for each study) is the only source of variability. For example, it is appropriate when results from randomized clinical trials, which follow a common protocol, are combined;
- Random-Effect (RE) model, that assumes heterogeneity among studies. It is adopted when there are two sources of variability: sampling variability (within study) and heterogeneity (measured between studies). Therefore, it is preferred when results from observational studies or potentially heterogeneous randomized clinical trials are combined and in general when the homogeneity assumption is no longer appropriate due to differences in design, analysis, studied population, and experimental condition.

In systematic reviews and meta-analyses, an evidence-based minimum set of items is generally used for reporting, i.e., Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The relevant checklist, flow diagram, statement, and explanation and elaboration are available at the website: https://prisma-statement.org/.

In addition, the International Prospective Register of Systematic Reviews (PROSPERO) gathers the protocols of systematic reviews [PROSPERO]. This procedures ensures that the reported analyses were not decided *post-hoc* to show something that was found "by chance" (e.g., with data mining techniques) or after *ad hoc* outcome selection.

One of the most useful sources for meta-analyses of trials evaluating the efficacy of treatments is the Cochrane Database of Systematic Reviews (www.cochrane.org).

Network Meta-Analysis

Differently from metanalyses, which contemplate direct comparisons (i.e., studies in which the same experimental drug is compared to the same comparator for the same outcome), network meta-analyses or Mixed Treatment Comparison (MTC) allow indirect comparisons of treatment effects. In fact, they enable us to combine trials comparing different sets of treatments and form a network of evidence within a single analysis [Caldwell, 2005]. Given that A, B, and C are three different treatments, the validity of this analysis is closely conditional to three assumptions [Salanti, 2013]:

- Homogeneity, which, in turn, includes the second and third assumptions.
- Transitivity, that means that one can learn about B versus C via A.
- Consistency, i.e., direct and indirect evidence are in agreement.

The main assumption in network meta-analysis is that relative within-trial treatment effects can be pooled, thus generalizing meta-analysis: Instead of solely considering RCTs conducted to investigate the same direct comparison, it infers on the network of available evidence.

The relevance in pharmacoeconomics is related to the possibility to inform simulation modeling about treatment comparisons that have not been (sufficiently) studied in clinical trials.

Example of Network Meta-Analysis

A Network Meta-Analysis (NMA) has been developed to determine the most effective therapy or combination of therapies in minimizing the exposure to homologous transfusion and the number of transfused Packed Red Blood Cells (PRBCs), while maximizing post-operative hemoglobin (Hb) during cardiac and thoracic surgery [Pradelli, 2016].

A systematic literature review up to July 2015 was performed via PubMed including Randomized Controlled Trials (RCTs), meta-analyses, and reviews. In addition, a non-systematic search was performed using Google Scholar. Investigated outcomes were Post-Operative (PO) Hb, Transfusion Rate (TR), and total number of transfused PRBCs. Only papers reporting both Hb (or Hct converted in Hb—conversion factor from Hct to Hb: 0.3389, R^2 = 0.8869, $p < 0.0001$) and TR or number of transfused PRBCs were considered. The choice of concurrently assessing clinical outcomes "three-dimensionally" aims to get a clearer picture of the perioperative blood management activity: since comparisons representing just one of the dimensions might be confusing and misleading (i.e., evaluating one therapy better than another based only on lower TR, while requiring more units transfused (PRBCs), or achieving a low level of Hb).

A random effects model with hierarchical structure was coded in WinBUGS according to the standard Bayesian Markov Chain Monte Carlo (MCMC) approach for indirect comparisons, multi-arm trials, and NMA. TR was modeled on the logit scale, while normal likelihood is used for continuous variables (PO Hb and mean number of PRBCs transfused).

A total of 86 RCTs were selected, comparing 48 different active strategies that were grouped into five broad categories (Figure 7.2 left):

- Auto-Transfusion (AT) of processed blood through the use of centrifugal cell washing (Cell Salvage—CS) or intraoperative blood processed using only an UltraFiltration device (UF) or unprocessed/unwashed blood (noCS);
- Administration of Antifibrinolytics (AA): aprotinin, tranexamic acid, e-aminocaproic acid, or desmopressin.
- The combined use of auto-transfusion and antifibrinolytics (AT+AA);
- Acute Normovolemic Hemodilution (ANH);
- No intervention (NT).

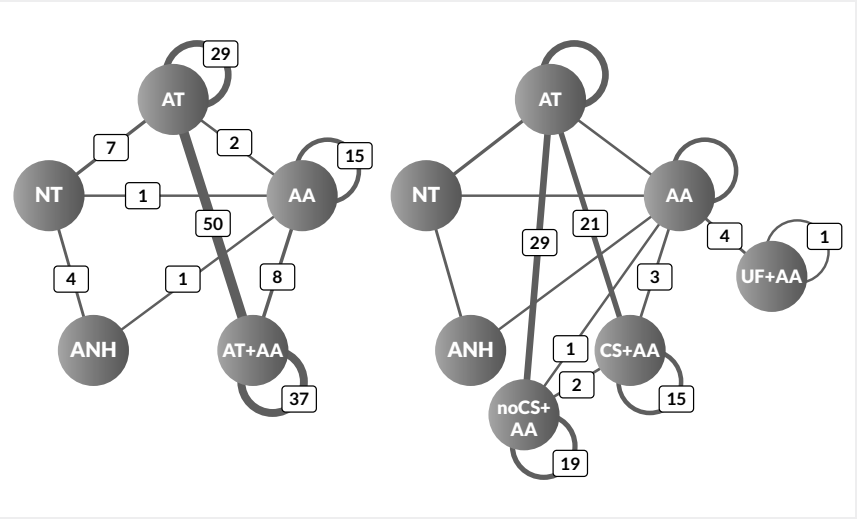

Figure 7.2. On the left, the evidence network resulting from systematic literature review (every edge between two nodes is labelled with the number of studies that compared the strategies represented by these nodes). On the right, the same network with AT+AA strategy blown-up into 3 strategies according to AT techniques (CS, noCS ,or UF).

AA = Administration of Antifibrinolytics; ANH = Acute Normovolemic Hemodilution; AT = Auto-Transfusion; CS = Cell Salvage; NT = no intervention; UF = UltraFiltration device

Furthermore, the category AT+AA was blown up into the 3 strategies according to AT techniques in order to investigate the specific effect of AT strategies (Figure 7.2 right).
AT+AA has the highest probability to be the best technique in reducing both TR and PRBCs transfused with high PO Hb, followed by AT alone. In combination with AA, the most effective AT strategy results CS (Table 7.2):

- The odds of TR are 0.36 (95% CrI 0.19–0.69) vs noCS+AA and 0.31 (95% CrI 0.11–0.88) vs. UF+AA with Bayesian p-value > 0.99;
- The amount of PRBCs transfused is -0.74 (95% CrI -1.41–0.02) vs noCS+AA and -0.9 (95% CrI -1.75–0.05) vs. UF+AA both with Bayesian p-value > 0.97;
- The level of PO Hb results quite the same for all 3 strategies.

According to NMA results, the Authors could conclude that the use of washed cell salvage in combination with antifibrinolytics is the optimum strategy to address perioperative blood loss. Also, replacing cell salvage with other autologous techniques such as unwashed cell salvage and ultrafiltration, or abolishing the combined use of antifibrinolytics, will increase the recourse to banked blood.

Clinical Studies

In extremis, when the required data are not available in the literature, a dedicated study must be performed. Observational studies are often undertaken to determine costs. However, a clinical trial is never undertaken with the sole purpose of determining cost-effectiveness. This is because a cost-effectiveness analysis is only considered **after** effectiveness has first been demonstrated. However, cost data may be collected prospectively in an effectiveness (or efficacy) trial, which is then called a "piggyback trial".

Administrative Databases

Medical information about patients, such as blood pressure, temperature, severity of illness, etc., is usually recorded on paper "charts" and is conveniently accessible only for individual patients or small groups of patients. An alternative is to "follow the money". Every healthcare transaction that takes place is documented using standard systems of codification, with the ulti-

	Transfusion rate (OR)	
	noCS+AA	UF+AA
CS+AA	0.36 (0.19–0.69) [>0.99]	0.31 (0.11–0.88) [0.98]
noCS+AA		0.87 (0.32–2.13) [0.67]
	PRBCs transfused (mean difference)	
	noCS+AA	UF+AA
CS+AA	-0.74 (-1.41–0.02) [0.97]	-0.90 (-1.75–0.05) [0.97]
noCS+AA		-0.16 (-1.07–0.73) [0.68]
	PO Hb (mean difference)	
	noCS+AA	UF+AA
CS+AA	0.23 (-0.17–0.62) [0.87]	0.23 (-0.97–0.88) [0.81]
noCS+AA		0.00 (-1.17–0.65) [0.41]

Table 7.2. Odds Ratios (OR) for the logit model and mean differences for the normal models for each pair of treatments and each outcome compared: Each cell gives the posterior mean with 95% central credible interval (CrI) in parentheses followed by the posterior probability that OR < 1 (or the mean difference <0).

AA = Administration of Antifibrinolytics; CS = Cell Salvage; Hb = Hemoglobin; PO = Post-Operative; PRBC = Packed Red Blood Cells; UF = UltraFiltration device

mate purpose of obtaining payment. This codified information can be used not only to track payments for healthcare services, but to build a picture of the healthcare services used by a patient or class of patients, and to infer a picture of the courses of diseases and the effects of the treatments.

The main types of data contained in administrative databases are:

- Demographic. Demographic data are usually limited to the patient's age and gender.
- Diagnostic. The diagnosis the patient receives is entered as a code using the International Classification of Diseases (ICD). The principal diagnosis and a limited number of comorbidities are recorded by their ICD codes and, for the healthcare services that use it, also with the relevant Diagnosis Related Group (DRG).
- Procedural. The exact nature of any procedure performed by a physician is documented in terms of a Current Procedural Terminology (CPT) code.
- Pharmaceutical. Each drug that is prescribed is specified by its National Drug Code (NDC). Drug consumption is reported in terms of dispensed packages.

These datasets also include information on the service provider: the physician's specialty, the setting (primary care, outpatient, inpatient), geographic

Example: A HMO Administrative Database Analysis

The healthcare costs of peptic ulcers and bleeding resulting from the prescription of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) for arthritis have been intensively investigated. Johnson et al. estimated the incidence of inpatient and outpatient gastropathies, the services provided to treat them, and the costs of those services for elderly members of a Health Maintenance Organization (HMO) located in the north-west United States [Johnson, 1997]. The data for morbidity and healthcare resource use were obtained from four automated databases maintained by the HMO: an outpatient pharmacy database, a hospital discharge database, a membership information database, and an outpatient utilization database. These four data sets were linked using the patient identifier tagging each record. The pharmacy and hospital discharge databases contained information for every prescription and inpatient stay, respectively, that occurred. The outpatient database represented a random sample of outpatient encounters abstracted from paper medical records. Costs to the HMO were estimated from the Medicare Cost Report, an aggregate report of costs that includes direct medical and overhead costs related to capital investment, general administration, etc., and which formed the basis for cost-based Medicare reimbursements to the HMO. The result of the study was that, for every dollar spent by the HMO on NSAID therapy for the elderly, another 35 cents was spent to treat the adverse gastrointestinal effects of the NSAIDs.

location, and, if inpatient, patient identifier, a unique number that identifies the recipient of the healthcare service.

It should be pointed out that the nature (public or private) and the information available in administrative databases varies according to the nation. For example, in the USA claims are almost the unique source. The claims are the reimbursement requests to Health Maintenance Organizations (HMOs—insurance structures that provide coverage through a network of physicians) made by providers such as hospitals, pharmacies, and healthcare professionals. The HMOs' databases are linked to the expenses incurred by each insured person. Because of the fragmented nature of the Healthcare System in the United States, however, collating the different datasets into one coherent whole may be problematic for any given population of patients. State Medicaid and Medicare datasets are certain comprehensive records for patients covered by this insurance systems, but they only apply to eligible indigent and elderly (65 years and over) patients, respectively. Staff model HMOs also may contain complete datasets. In the case of third-party payment systems, however, the datasets may be dispersed among different payer. Pharmacy datasets may be maintained by Pharmacy Benefit Management companies (PBMs), while numerous healthcare insurers may maintain hospital and primary care claims data. Collations of private sector healthcare insurance datasets can be purchased that offer a complete picture of the encounters of hundreds of thousands of patients with the healthcare system over defined time periods.

Conversely, in Italy and in other countries, data must be collected by several databases by using data linkage, i.e., gathering data on the same patient from different databases by means of a unique code, e.g., the fiscal code.

Example: A Private-Pay, Fee-for-Service Database Analysis

The healthcare costs associated with the treatment of depression with different classes of antidepressant drugs were estimated in a retrospective cohort study [Poret, 2001]. The data source was a proprietary database of medical and pharmacy claims collated from numerous private, fee-for-service healthcare insurers covering employees of corporate America. Individuals who had a new prescription for an antidepressant drug and a diagnosis of depression within a defined time period (the index period) were identified, and their healthcare resource use was tracked for the next six months. A diagnosis of depression was indicated by a relevant ICD-9 code in inpatient and outpatient records. Antidepressant drugs were identified by their NDC codes. Costs were compared for antidepressant drug classes using an intent-to-treat analysis, i.e., the patient was classified according to the initial antidepressant drug he or she was prescribed in the index period, regardless of whether he or she subsequently switched to another class during the follow-up period.

Financial Data

Administrative datasets contain the dollar (or euro, etc.) payments made for the healthcare resources used. These payments are not necessarily the same as the charges made by the payee because of negotiated fee schedules, capitation, etc. List prices of healthcare goods and services are published by State Medicare systems. In the United States, list prices for drugs are published as average wholesale prices and are also available as retail prices. Data on employee remuneration, which might be needed for an indirect costs analysis, are provided by the Bureau of Labor Statistics.

7.2 Statistical Analysis

Classical statistics is based on hypothesis testing. The hypothesis is made that the observations to be explained are the result purely of chance: This is the null hypothesis. A calculation is then made of the probability that the observations would arise under the null hypothesis and, if that probability is below an arbitrary threshold (most often 1 in 20, or 0.05), the null hypothesis is rejected. The results cannot be explained purely by chance and are said to be "statistically significant".

Note that the above procedure explores the role of random chance and in itself does nothing to assess the role of systematic error (discussed below),

Example: Classical Statistics

The efficacy and safety of ustekinumab (UST), a human IgG1 monoclonal antibody currently approved for several autoimmune diseases, has been recently evaluated in patients with moderately-to-severely active Ulcerative Colitis (UC) in a phase III, randomized, double-blind, placebo (PBO)-controlled study [Sands, 2019]. In this study, 15.6% of patients received 130 mg intravenous UST (50/320) and 5.3% of patients treated with PBO (17/319) achieved clinical remission at week 8 (chi-square test, p-value < 0.001). In this example, p-value represents the probability of obtaining an effect at least as extreme as the one we have measured (15.6% vs. 5.3%), assuming the truth of the null hypothesis, i.e., UST has the same efficacy of PBO. Hence, according to the test's result, if UST had no effect, we would observe such results but chance in less than 0.1% of experiments. The difference in clinical remission rate among patients treated with UST and those treated with PBO is "statistically significant" and we can conclude that UST is not equivalent to PBO, but it is likely to be better than PBO.

which is often more important than random error. Hypothesis testing is not particularly useful in decision analysis, where we need to know the probability of a certain event occurring (such as death from a myocardial infarction) under a certain set of circumstances (such as when a patient has already had one heart attack). The calculation of such conditional probabilities is referred to as **Bayesian analysis**. To the non preconditioned mind, the **Bayesian approach** may be more intuitive, if less conceptually sophisticated, than hypothesis testing.

Example: Bayesian Analysis

Classical (frequentist) statistics uses only experimental evidence to reject/confirm the null hypothesis, while Bayesian analysis includes also the previous knowledge of the phenomenon under study.
Let's consider the results of two different RCTs comparing dabrafenib, an antineoplastic BRAF-inhibitor, with placebo in patients suffering from metastatic melanoma (trial 1) and non-small cell lung cancer (trial 2) in terms of 1-year mortality (Table 7.3).

	Dabrafenib	Placebo	RR (95% CI)
Trial 1 (N = 600)	98/300 (32.7%)	147/300 (49.0%)	0.67 (0.51 to 0.87), p = 0.0021
Trial 2 (N = 200)	32/100 (32.0%)	44/100 (44.0%)	0.73 (0.45 to 1.17), p = 0.2067

Table 7.3. 1-year mortality in two different hypothetical randomized controlled trials comparing dabrafenib with placebo in patients suffering from metastatic melanoma (trial 1) and non-small cell lung cancer (trial 2).

CI = confidence interval; RR = relative risk

In both trials there is a reduction of 1-year mortality in patients treated with dabrafenib, but only in the first trial the efficacy of dabrafenib was proven (p under the standard threshold of 0.05).
The Bayesian interpretation of the second trial could be more informative:

- Both diseases are cancers associated with the BRAF gene mutation;
- Dabrafenib is a BRAF-inhibitor and proved effective in the first clinical trial with large sample size;
- Trial 2 is smaller than trial 1, but the effect of dabrafenib is similar to that observed in trial 1.

By using the information from trial 1, it is possible to adjust the comparison in the trial 2; the posterior probability that dabrafenib reduces 1-year mortality is > 90%.

7.3 Accounting for Uncertainty

Definition of Error

There are two kinds of error, namely random error and systematic error. Random error is variability in the result caused by random or unpredictable variability in the factors determining the result. Systematic error is a bias in the result caused by nonrandom variability in these factors.

To understand the distinction between the two kinds of error, imagine darts thrown at a dartboard. The random clustering of the darts around the bull's eye (which is the target) represents random error. The tighter the clustering around the bull's eye, the less random error there is. If the darts tend to cluster in the lower right quadrant of the board, for example, there is a systematic error or bias in the thrower's aim.

In this context, the word "error" means variability and does not imply that a preventable mistake has been made. Similarly, the word "bias" does not require a conscious or unconscious human motivation to alter the results.

Principal Sources of Error in Clinical Studies

Although many types of bias in the design and conduct of chemical studies have been described, Chalmers has pointed out that most fall into three important categories: selection bias at study entry, selection bias after study entry, and bias in assessing outcomes [Chalmers, 1989] (Table 7.4).

Selection bias at study entry may occur when a randomizing clinician knows in advance in which arm the patient will be allocated, thus consciously or unconsciously excluding unsuitable patients [Souter, 1997].

Example

Several randomly chosen human subjects were given the same dose of a drug. The response of the subjects to the single dose varied greatly and a graph of the drug response versus frequency described a bell curve. Among the factors causing this variability, there were genetically determined differences among subjects in their ability to absorb, metabolize, and eliminate the drug, and in the interaction of the drug with its tissue target. The subjects varied greatly in their body mass, blood volume, percent body fat, the metabolizing ability of their livers, and in a host of other attributes that modified the action of the drug.

Error	Controlled trial	Comparative observational study
Selection bias at study entry	Randomization	Case matching, propensity scoring
Selection bias after study entry	Intent-to-treat analysis	Intent-to-treat analysis, multivariate analysis
Bias in assessing outcomes	Observer blinding, subject blinding	

Table 7.4. Means of accounting for systematic error by study design.

Selection bias after study entry may occur if patients who left their assigned treatment group were not included in the analysis, since these patients were subject to some form of selection.

Bias in assessing outcomes occur when the researcher that has to assess outcomes and effects is aware of group allocation [Souter, 1997].

Reducing Error in Clinical Studies

The obscuring of the true result due to random error can be lessened in two ways: by increasing the sample size, and by reducing the variability in the sample. Approaches to reducing systematic error are discussed below.

Selection Bias in Subjects Entering the Study

Randomization

In a controlled clinical trial, random allocation of subjects to treatments eliminates systematic error at study entry. An imbalance in the allocation of subjects may remain—e.g., more females than males may be assigned to treatment A than to treatment B—simply because of random error, but this may be reduced by increasing the sample size. The process of random allocation is blind to both perceivable and unperceivable differences between subjects, and this is its principal virtue.

Randomization procedure should achieve the following objectives [Lachin, 1988]:

- Equal group sizes for adequate statistical power, especially in subgroup analyses;
- Low selection bias; the procedure should not allow an investigator to predict how subjects will be assigned in reviewing the previous pairings;

- Low probability of confounding (i.e., a low probability of "accidental bias"), which implies a balance in covariates across groups. If the randomization procedure causes an imbalance in covariates related to the outcome across groups, estimates of effect may be biased if not adjusted for the covariates.

No single randomization procedure meets those goals in every circumstance, so researchers must select a procedure for a given study based on its advantages and disadvantages [Roter, 1998].

Randomization procedures are described below.

Simple randomization. Intuitive and commonly used procedure, similar to "repeated fair coin-tossing", also known as "complete" or "unrestricted" randomization. It is robust against both selection and accidental biases. However, its main drawback is the possibility of imbalanced group sizes in small RCTs. It is therefore recommended only for RCTs with over 200 subjects.

Restricted randomization. To balance group sizes in smaller RCTs, some form of restricted randomization is recommended. The major types of restricted randomization used in RCTs are:

- Blocked randomization: The number of subjects in one group versus the other group and the block size are specified; subjects are allocated randomly within each block. For example, a block size of 6 and an allocation ratio of 2:1 would lead to random assignment of 4 subjects to one group and 2 to the other. Unfortunately, even if the block sizes are large and randomly varied, the procedure can lead to selection bias.
- Adaptive biased-coin randomization methods: relatively uncommon methods in which the probability of being assigned to a group decreases if the group is over-represented and increases if the group is under-represented. The methods are thought to be less affected by selection bias than permuted-block randomization.

Adaptive randomization. Less frequently, other two types of adaptive randomization procedures have been used in RCTs:

- Covariate-adaptive randomization: The probability of being assigned to a group varies in order to minimize covariate imbalance. Since only the first subject's group assignment is truly chosen at random, the method does not necessarily eliminate bias on unknown factors.
- Response-adaptive randomization, also known as outcome-adaptive randomization: The probability of being assigned to a group increases if the responses of the prior patients in the group were favorable.

Case Matching

In a comparative observational study, there may be considerable differences between the subjects who received treatment A and those who received treatment B. The investigator can control for perceivable (but not imperceivable) differences in subjects by balancing these characteristics between the two comparative groups. This is case matching.

Example of Case Matching

It is the year 2019. Patients who happened to receive treatment A at a clinic in the year 2017 were predominantly female and younger, while those who received treatment B were mostly male and older. The investigator compares treatment A with treatment B by picking, for example, a female in the 30-40 year age category who received treatment A and a (harder-to-find) female in the same age category who happened to have received treatment B. The investigator does the same for females in the 50-60 year age category and in the 70-80 year age category, etc., and for males in the same age categories. The investigator now has a group who received treatment A that is evenly balanced by age and gender with a group who received treatment B, and he may proceed with computation of the study outcome.

Propensity Scoring

In an analysis based on propensity scores, subjects in group A are matched with those subjects in group B who have the same propensity score, where the propensity score is the probability that a subject will be assigned to group B rather than group A based on a composite of observable determining characteristics [Joffe, 1999]. The study population is then divided into categories (usually quintiles) based on their propensity score and within each quintile the outcomes of individuals who received treatment A are compared with those who received treatment B. Propensity score matching is based on two technical assumptions [Robinson, 2004]:

- **Assumption 1** (Conditional Independence Assumption—CIA): There is a set X of covariates, observable to the researcher, such that after controlling for these covariates, the potential outcomes are independent of the treatment status. The CIA is crucial for correctly identifying the impact of the program, since it ensures that, although treated and untreated groups differ, these differences may be accounted for in order to reduce the selection bias. This allows the untreated units to be used to construct a counterfactual for the treatment group.

- **Assumption 2** (common support condition): For each value of X, there is a positive probability of being both treated and untreated. Then, a simple way of interpreting this assumption is the following: The proportion of treated and untreated individuals must be greater than zero for every possible value of X.

When these two assumptions are satisfied, the treatment assignment is said to be **strongly ignorable** [Robinson, 2004].

Selection Bias in Subjects After Entering the Study

In a randomized controlled trial, subjects initially allocated to receive (say) treatment A might subsequently leave their assigned treatment group and receive no treatment, an unplanned treatment, or switch to treatment B. In an intent-to-treat analysis, all subjects initially assigned to receive treatment A are included in the analysis of outcomes of group A. This ensures that the function of randomization—the control for selection bias—is preserved. A selection bias would be introduced if patients who left their assigned treatment group were not included in the analysis, since these patients were subject to some form of selection. The same intent-to-treat approach can be applied to retrospective studies.

7.4 Pharmacoeconomic Studies

Sensitivity Analysis

The sample decision analysis described earlier (see Chapter 4) contained an example of one-way sensitivity analysis, which is a type of simple sensitivity analysis, that, in turn, is one of several general approaches.

Simple Sensitivity Analysis

In simple sensitivity analysis, one study variable is varied over the range of likely values, while all other variables are held constant. If the variables are independent, a series of one-way sensitivities is informative. In two-way sensitivity analysis, the effects of varying two variables simultaneously are computed. Similarly, in three-way sensitivity analysis, three variables are varied over their likely ranges of values at the same time. Two- and three-way sensitivity analyses are more appropriate when study variables interact. Calcula-

tions are performed using a computer program and results are displayed graphically.

Results of one-way analysis can be represented by a tornado diagram, which is a special type of bar chart, where data categories are listed vertically

Example of Simple Sensitivity Analysis

A decision analytic DES model was built recently [Pradelli, 2012a] in order to compare the cost-effectiveness of two treatment alternatives for patients needing parenteral nutrition: Use of parenteral ω-3 enriched emulsions or standard fat emulsions. A one-way analysis is applied on the base case scenario to study which parameters are most influential on final results. Given that the main conclusion of the base-case simulation is dominance (parenteral ω-3 vs. standard fat emulsion), it was chosen to explore the effect of parameter value estimates on total incremental costs (Figure 7.3) by a deterministic one-way analysis.

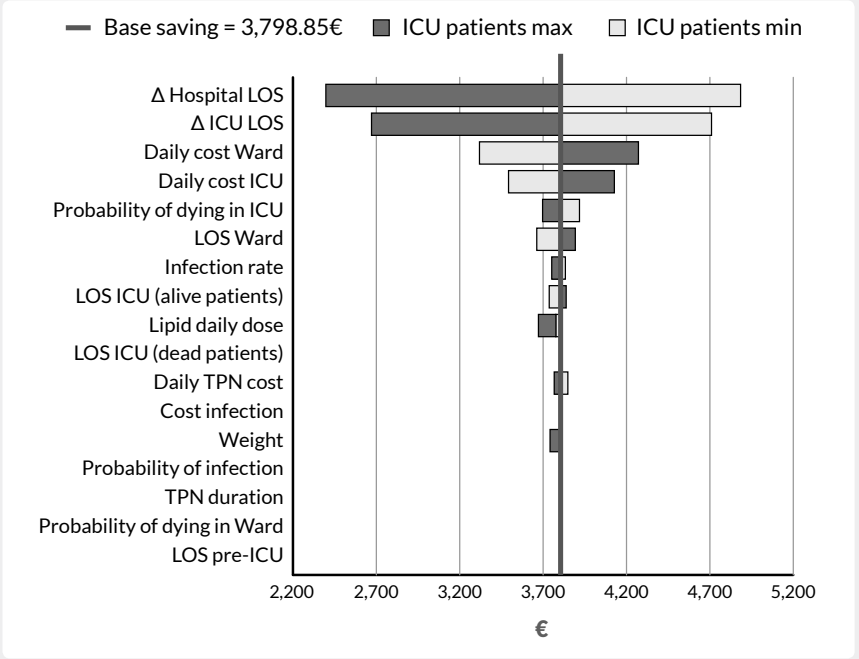

Figure 7.3. Example of a tornado diagram. The extremes of the 95% confidence interval were selected as minimum and maximum values of parameters; for variables without such intervals, a ±20% variation is applied to baseline values.

ICU = Intensive Care Unit; LOS = Length Of Stay; TPN = Total Parenteral Nutrition

instead of the standard horizontal presentation, and categories are ordered so that the largest bar appears at the top of the chart, the second largest appears second from the top, and so on. These graphs are so named because the final chart appears like a tornado (Figure 7.3).

Probabilistic Analysis

Instead of performing several one-way or multi-way analyses, it would be more desirable to perform an "every-way" analysis, i.e., simultaneously vary all the study variables throughout their ranges of likely values. A computer program can do this by starting with an imaginary cohort of, say, 1,000 patients and running them one by one through the decision tree, randomly assigning a likely value for the probability at each chance node. A distribution of outcome values for the entire 1,000 patients can be thus calculated, and a mean and 95% confidence interval (95% CI) computed. When distributions for the values of chance node probabilities are known, the computer can take these into account (rather than assuming a uniform distribution within a range of values).

The patient-level, probabilistic simulation (also called "Monte Carlo simulation") is performed by drawing parameter values from their probability distribution for each simulated individual, and allows to take into account two levels of uncertainty [Briggs, 2001]:

- The uncertainty on patient characteristics, which represents the effective heterogeneity among subjects;
- The uncertainty about model parameters, to represent the cognitive uncertainty on values derived from experimental measurements.

In cost-effectiveness analysis, results of probabilistic analysis can be represented by a scatterplot: Incremental costs and benefits are plotted on the cost-effectiveness plane and the grade of dispersion of the cloud of points can provide a visual indication of model stability (Figure 7.4).

Black square indicates ICER in the base case while grey points are simulated-patients in the probabilistic sensitivity analysis. The black ellipse represents the 95% confidence ellipse and the slope of each straight lines represents possible Willingness To Pay (WTP) thresholds. The fraction of points under the lines are simulated patients within the threshold associated with each line (as shown in Figure 7.4); more precisely the area of the 95% confidence ellipse under each line corresponds to the probability that the treatment is cost-effective given each WTP threshold.

To understand the uncertainty around the mean incremental cost and benefit, it is useful to plot, together with the 1,000 results of probabilistic simu-

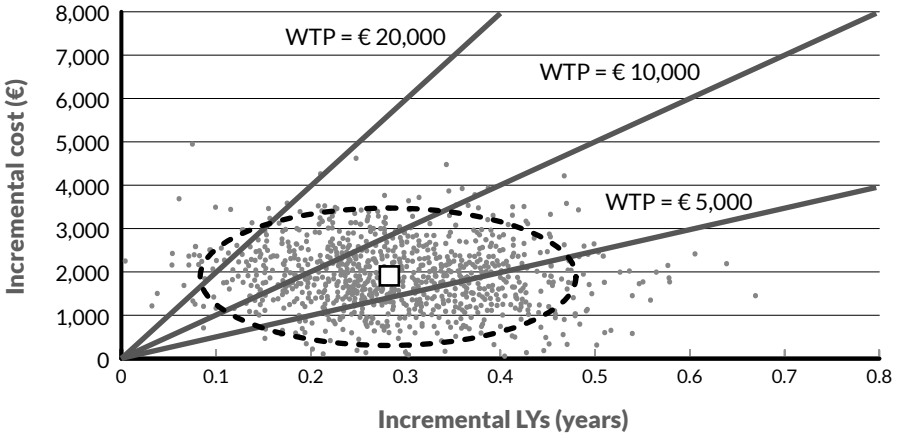

Figure 7.4. Scatterplot of the incremental costs and benefits on the cost-effectiveness plane. Modified from [Pradelli, 2014].

LY = Life-Years; WTP = Willingness To Pay

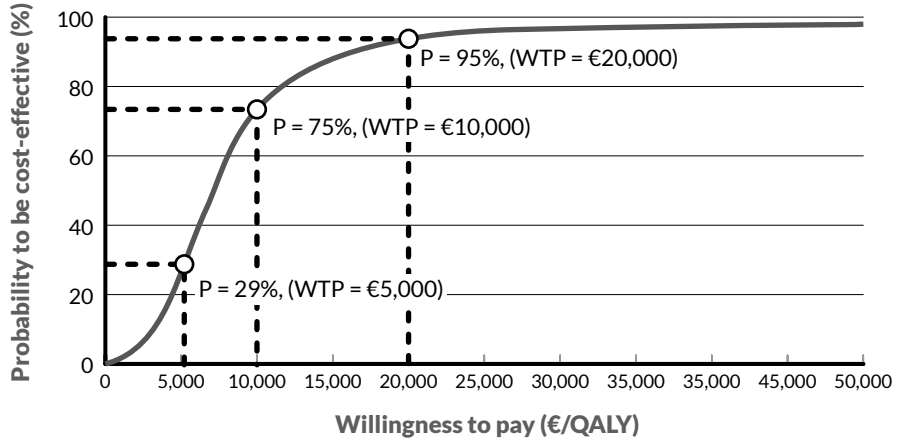

Figure 7.5. Cost-effectiveness acceptability curve. Modified from [Pradelli, 2014]. Apixaban versus standard therapy has an estimated 29%, 74%, or 95% probability of being cost-effective if the decision maker is willing to pay (WTP) up to 5,000€/QALY, 10,000€/QALY, or 20,000€/QALY respectively.

QALY = Quality-Adjusted Life-Year; WTP = Willingness To Pay

lation, 95% confidence ellipse that represents the region in the incremental cost-effectiveness plane with a 95% probability of containing the mean incremental cost and incremental benefit [Nixon, 2010]; those simulations that

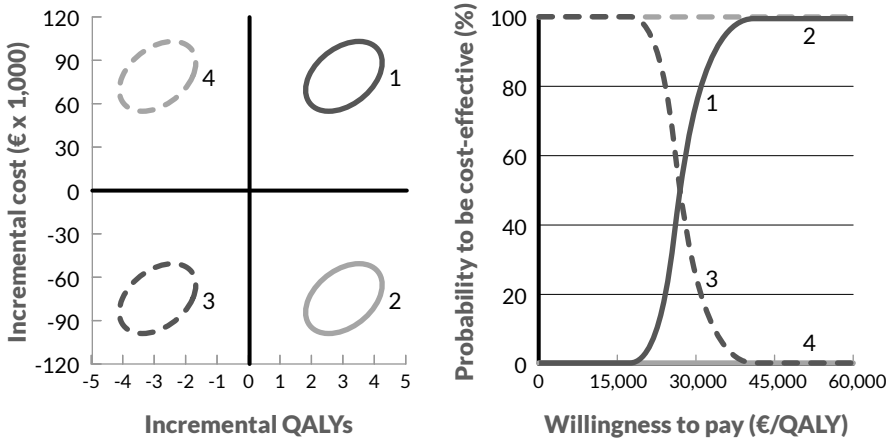

Figure 7.6. When the probability that the joint distribution of incremental cost and benefit extends beyond a single quadrant is negligible (i.e., the 95% confidence ellipse lies completely in one of the four quadrants of the incremental cost-effectiveness plane), CEAC can assume 4 different configurations: (1) The traditional "S"-shaped curve: The new treatment is more costly and more effective (the ellipse is contained in the first quadrant) and CEAC is an increasing function of WTP; (2) the ellipse is contained in the second quadrant, i.e., the new treatment is dominant (less costly and more effective) and CEAC is constantly equal to 1 since the entire distribution involves cost-savings and health gains; (3) and (4) are the inverse of case (1) and (2) respectively.

CEAC = Cost-Effectiveness Acceptability Curve; QALY = Quality-Adjusted Life-Year; WTP = Willingness To Pay

do not fall into the ellipse region can be seen as outliers. The 95% confidence ellipse can be seen as the two-dimensional analogous of the 95% confidence interval. Furthermore, the orientation of the ellipse represents graphically the correlation between incremental cost and benefit (Figure 7.4).

Another indication of results reliability is given by the Cost-Effectiveness Acceptability Curves (CEACs) [Briggs, 2001], where the WTP for a unit benefit gained from a hypothetical decision-maker (i.e., the cost-effectiveness threshold considered as acceptable) is placed on the x-axis and the probability that an intervention is cost-effective compared with the alternative, given the observed data, on the y-axis (Figure 7.5).

The CEAC, for each value λ of the WTP, is determined as the proportion of simulations in the incremental cost-effectiveness plane, falling to the south-east of a ray through the origin with slope equal to λ [Fenwick, 2004]. The

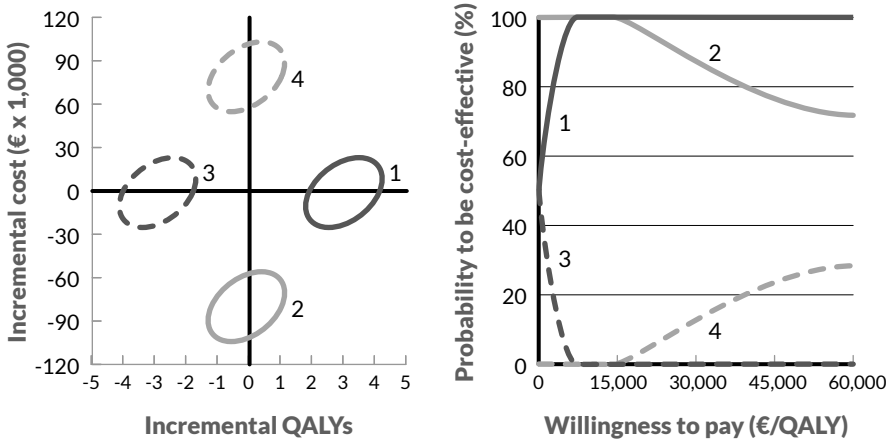

Figure 7.7. When the 95% confidence ellipse occupies two quadrants, i.e., it crosses either x-axis or y-axis, CEAC can assume 4 less standard configurations: (1) The new treatment is more effective but not always less costly and CEAC does not start from 0, since confidence ellipse is partially in the dominance quadrant and asymptotes to 1 because all the distribution involves health gains; (2) the new treatment is cost-saving, but not always more effective and CEAC starts from 1 because the entire density involves cost-savings, but it decreases asymptotically to a value lower than 1 because not all of the joint density involves health gains (here only 50%); (3) and (4) are the inverse of case (1) and (2) respectively.

CEAC = Cost-Effectiveness Acceptability Curve; QALY = Quality-Adjusted Life-Year

classic illustration of the CEAC is an "S"-shaped curve starting from the origin and asymptotic to 1 (as seen in Figure 7.5), but this is just one of the possible shape that CEAC can take. All possible CEAC shapes depend on the configuration of the 95% confidence ellipse in the incremental cost-effectiveness plane (Figures 7.6, 7.7, and 7.8).

Best and Worst Case Scenarios

This form of sensitivity analysis is more formally called "**analysis of extremes**". A best-case estimate of cost-effectiveness would combine the lower extreme estimate of costs and the upper extreme estimate of effectiveness. Similarly, a worst-case estimate would combine the upper extreme of costs and the lower extreme of effectiveness. Compared to a Monte Carlo simulation, this type of analysis is crude, but it is useful in some circumstances.

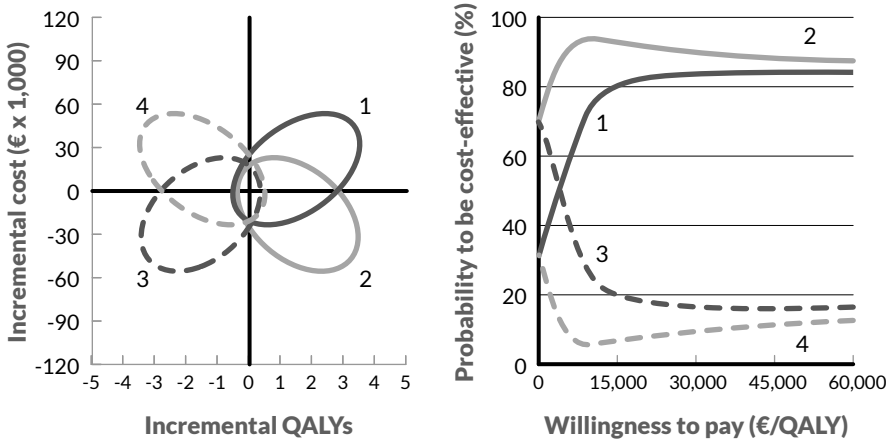

Figure 7.8. If the confidence ellipse cover 3 or 4 quadrants, the corresponding CEAC depends on the quadrant mostly occupied by the joint distribution: (1) if the ellipse is mainly in the first quadrant (more costly, more effective), the CEAC does not start from zero (as case 1 in Figure 7.5) but does not asymptote to 1 because not all of the joint distribution involves health gains, however the function is strictly increasing; (2) if the ellipse is mainly in the second quadrant (less costly, more effective), as in case (1) CEAC does not start from zero and is not asymptotic to 1, however, the CEAC is not strictly an increasing function, due to the position of the joint distribution in the first and third quadrants (as the WTP increases the joint distribution in the first quadrant is included as cost-effective before the joint distribution in the third quadrant is excluded as no longer cost-effective, and the CEAC rises before falling); (3) and (4) are the inverse of case (1) and (2) respectively.

CEAC = Cost-Effectiveness Acceptability Curve; QALY = Quality-Adjusted Life-Year; WTP = Willingness To Pay

Threshold Analysis

Threshold analysis is a modification of one-way sensitivity analysis. Instead of a single variable being varied throughout its range of likely values, it is varied with the purpose of finding the threshold at which the decision alternatives have the same expected value. The threshold is also called the "break-even point".

7.5 Conclusions: Evaluating Pharmacoeconomic Studies

The established standard for assessing any published study is that sufficient information should be presented for a peer researcher to be able to repeat the study (and get the same answer). In the case of a cost-effectiveness analysis, it means that it should be completely described in the text and that all study variables and information sources should be reported.

Many published cost-effectiveness analyses are black boxes that are difficult for the average reader to assess, let alone reproduce; in such circumstances, the credibility of the researchers is important. Guidelines or checklists can be useful aids in assessing studies. Guidelines for evaluating pharmacoeconomic studies have been presented by university researchers [Anonymous, 1995], the Pharmaceutical Research and Manufacturers of America [Clemens, 1995], and state governments [Torrance, 1996]. In addition to these, criteria for assessing the conduct and reporting of various kinds of studies have been developed. The Users' Guides to the medical literature series published in *JAMA* includes articles on the assessment of economic analysis [Drummond, 1997], decision analysis [Richardson, 1995a; Richardson, 1995b; O'Brien, 1997; Naylor, 1996], health-related quality of life/outcomes research [Guyatt, 1997; Anonymous, 1995; Hartmaier, 1995], and systematic reviews [Oxman, 1994]. There are also published criteria for the reporting of clinical trials (CONSORT, 2010) [Schulz, 2010] and for writing manuscripts of different types [International Committee of Medical Journal Editors, 2022]. Recently, also several task force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) set up a series of Good Practice for Outcomes Research to consolidate the knowledge on economic appraisal and to provide recommendations to optimize the reporting of health economic evaluations [Husereau, 2022].

Questions

1. Tick the correct sentence

A. Delphi panels are a particular type of experimental studies
B. Cost-utility and cost-effectiveness analyses are synonyms
C. Performing an *ad hoc* study design is one of the methods used in pharmacoeconomics to find answers
D. None of the answers are correct

2. Tick the correct sentence

A. Cross-sectional studies are neither prospective nor retrospective
B. Experimental studies may be retrospective or prospective
C. Observational studies may experimental or cross-sectional
D. Longitudinal studies analyze a point in time

3. Tick all that apply

A. The U.S. state and federal governments make available basic demographic information
B. The Census Bureau publishes past and projected population demographic data
C. The National Center for Health Statistics also collects and makes available ongoing national survey data
D. The results of several studies can be combined in meta-analyses, systematic reviews, or network meta-analyses

4. Tick all that apply to meta-analyses

A. Meta-analyses may contain systematic reviews
B. A meta-analysis is the statistical pooling of data from several studies
C. Fixed- and random-effect models may be used to combine the results obtained by meta-analyses
D. Fixed-effect models are adopted when there are two sources of availability

5. Tick the correct answer

A. Cochrane Database of Systematic Reviews is an evidence-based minimum set of items used for reporting systematic reviews and meta-analyses

B. PROSPERO is an evidence-based minimum set of items used for reporting systematic reviews and meta-analyses
C. One of the most useful sources for meta-analyses of trials evaluating the efficacy of treatments is the Cochrane Database of Systematic Reviews
D. One of the most useful sources for meta-analyses of trials evaluating the efficacy of treatments is PRISMA

6. Tick all that apply to network meta-analyses

A. They allow to combine trials comparing different sets of treatments and form a network of evidence within a single analysis
B. Their validity is conditional to the following assumptions: homogeneity, transitivity, and consistency
C. Their relevance in pharmacoeconomics is related to the possibility to inform simulation modeling about treatment comparisons that have not been sufficiently studied in clinical trials
D. They consider only RCTs conducted to investigate the same direct comparison

7. Tick all that apply

A. In the USA health system, the claims are the reimbursement requests to Health Maintenance Organizations made by providers such as hospitals, pharmacies, and healthcare professionals
B. Diagnoses are encoded by NDCs
C. DRGs are used in the health systems all over the world
D. The exact nature of any procedure performed by a physician is documented in terms of a CPT code

8. Tick all that apply

A. Classical statistics is based on hypothesis testing
B. In classical statistics, if the probability that the observations would arise under the null hypothesis is above an arbitrary threshold (most often 1 in 20, or 0.05), the null hypothesis is rejected
C. Bayesian analysis allows the calculation of conditional probabilities
D. Hypothesis testing is very useful in decision analysis, where we need to know the probability of a certain event occurring under a certain set of circumstances

9. Tick the correct answer

A. Systematic error is variability in the result caused by random or unpredictable variability in the factors determining the result
B. Random error is a bias in the result caused by nonrandom variability in these factors
C. The word "bias" requires a conscious or unconscious human motivation to alter the results
D. The word "error" does not imply that a preventable mistake has been made

10. Tick all that apply to randomization

A. Randomization may be applied to reduce selection bias in subjects entering the study
B. Simple randomization is recommended for RCTs enrolling less than 200 subjects
C. Blocked randomization and adaptive biased-coin randomization are subtypes of restricted randomization
D. The most used randomizations are covariate-adaptive randomization and response-adaptive randomization

11. Tick all that apply to propensity scoring

A. The propensity score is based on 4 assumptions
B. When these assumptions are satisfied, the treatment assignment is said to be strongly ignorable
C. The conditional independence assumption is crucial for correctly identifying the impact of the program
D. The propensity score is the probability that a subject will be assigned to group B rather than group A based on a composite of observable determining characteristics

12. Tick all that apply to sensitivity analyses

A. In simple sensitivity analysis, all the study variables are varied over the range of likely values
B. In two-way sensitivity analysis, the effects of varying two variables simultaneously are computed
C. Results of one-way analysis can be represented by a tornado diagram
D. If the variables are independent, a series of one-way sensitivities is informative

13. Tick all that apply to probabilistic analyses

A. Monte Carlo simulation allows to take into account five levels of uncertainty
B. Results of probabilistic analysis can be represented by a scatterplot
C. In CEACs, the willingness to pay for a unit benefit gained from a hypothetical decision-maker is placed on the y-axis and the probability that an intervention is cost-effective compared with the alternative, given the observed data, on the x-axis
D. 95% confidence ellipse represents the region in the incremental cost-effectiveness plane with a 95% probability of containing the mean incremental cost and incremental benefit

Answers

1. **C**
2. **A**
3. **A, B, C, D**
4. **B, C**
5. **C**
6. **A, B, C**
7. **A, D**
8. **A, C**
9. **D**
10. **A, C**
11. **B, C, D**
12. **B, C, D**
13. **B, D**

References

- Anonymous. Economic analysis of health care technology. A report on principles. Task Force on Principles for Economic Analysis of Health Care Technology. *Ann Intern Med* 1995; 123: 61-70. https://doi.org/10.7326/0003-4819-123-1-199507010-00011
- Briggs A. Handling uncertainty in economic evaluations and presenting the results. In: Drummond M, McGuire A (eds.). Economic evaluation in health care. *Oxford University Press*. New York, NY (USA): 2001; pp. 172-214
- Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331: 897-900. https://doi.org/10.1136/bmj.331.7521.897
- Chalmers I, Hetherington J, Elboume D, et al. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Kerrse MJN (eds). Effective care in pregnancy and childbirdth. Col 1: Pregnancy. *Oxford University Press*. New York, NY (USA): 1989; pp. 39-65
- Clemens K, Townsend R, Luscombe F, et al. Methodological and conduct principles for pharmacoeconornic research. Pharmaceutical Research and Manufacturers of America. *Pharmacoeconomics* 1995; 8: 69-74. https://doi.org/10.2165/00019053-199508020-00008
- Drummond MF, Richardson WS, O'Brien BJ, et al. Users' guides to the medical literature. XUI. How to use 'an article on economic analysis of clinical practice. A. Are the results of the study. Evidence based Medicine Working Group. *JAMA* 1997; 277: 1552-7. https://doi.org/10.1001/jama.277.19.1552
- Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions. *Health Economics* 2004; 13: 405-15. https://doi.org/10.1002/hec.903
- Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002; 359: 57-61. https://doi.org/10.1016/S0140-6736(02)07283-5
- Guyatt G, Naylor CD, Juniper E, et al. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. *JAMA* 1997; 277: 1232-7. https://doi.org/10.1001/jama.277.15.1232

- Hartmaier SL, Santanello NC, Epstein RS, et al. Development of a brief 24-hour migraine-specific quality of life questionnaire. *Headache* 1995; 35: 320-9. https://doi.org/10.1111/j.1526-4610.1995.hed3506320.x
- Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at http://handbook.cochrane.org (last accessed August 2022)
- Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. *Value Health* 2022; 25: 3-9. https://doi.org/10.1016/j.jval.2021.11.1351
- International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated May 2022. Available at https://www.icmje.org/icmje-recommendations.pdf (last accessed September 2022)
- Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol* 1999; 150: 327-33. https://doi.org/10.1093/oxfordjournals.aje.a010011
- Johnson RE, Hornbrook MC, Hooker RS, et al. Analysis of the costs of NSAID-associated gastropathy. Experience in a US health maintenance organisation. *Pharmacoeconomics* 1997; 12: 76-88. https://doi.org/10.2165/00019053-199712010-00008
- Lachin JM. Statistical properties of randomization in clinical trials. *Control Clin Trials* 1988; 9: 289-311. https://doi.org/10.1016/0197-2456(88)90045-1
- Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of Health Services. The Evidence-Based Medicine Working Group. *JAMA* 1996; 275: 554-8. https://doi.org/10.1001/jama.275.7.554
- Nixon RM, Wonderling D, Grieve RD. Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. *Health Econ* 2010; 19: 316-33. https://doi.org/10.1002/hec.1477
- O'Brien BJ, Heyland D, Richardson WS, et al. Users guides to the medical literature. XIII. How to use an article on economic analysts of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine

Working Group. *JAMA* 1997; 277: 1802-6. https://doi.org/10.1001/jama.1997.03540460066034
- Oxman AD, Cook OJ, Guyatt GH. Users' guides to the medical literature. Vl. How to use an overview. Evidence-Based Medicine Working Group. *JAMA* 1994; 272: 1367-71. https://doi.org/10.1001/jama.272.17.1367
- Poret AW, Neslusan C, Ricci J, et al. Retrospective analysis of the health care costs of bupropion sustained release in comparison with other antidepressants. *Value Health* 2001; 4: 362-9. https://doi.org/10.1046/j.1524-4733.2001.45068.x
- Pradelli L, Calandriello M, Di Virgilio R, et al. Comparative pharmacoeconomic assessment of apixaban vs. standard of care for the prevention of stroke in Italian atrial fibrillation patients. *Farmeconomia. Health economics and therapeutic pathways* 2014; 15(Suppl 1): 15-25. https://doi.org/10.7175/fe.v15i1s.974
- Pradelli L, Eandi M, Povero M, et al. Cost-effectiveness of supplemental n-3 in total parenteral nutrition therapy in the Italian, French, German and UK context: a discrete event simulation model. *Value Health* 2012; 15: A539. https://doi.org/10.1016/j.jval.2012.08.1892 [a]
- Pradelli L, Mayer K, Klek S, et al. ω-3 Fatty-Acid Enriched Parenteral Nutrition in Hospitalized Patients: Systematic Review With Meta-Analysis and Trial Sequential Analysis. *JPEN J Parenter Enteral Nutr*. 2020; 44: 44-57. https://doi.org/10.1002/jpen.1672
- Pradelli L, Povero M, Myers G. Perioperative therapies reducing homologous transfusions during Cardiovascular Surgery: A Network Meta-Analysis of Randomized Controlled Trials. *Value Health* 2016; 19: A705. https://doi.org/10.1016/j.jval.2016.09.2059
- PROSPERO. International prospective register of systematic reviews, available at www.crd.york.ac.uk/prospero/), PRISMA. Transparent reporting of systematic reviews and meta-analyses, available at www.prisma-statement.org/ (last accessed September 2022)
- Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1995; 273: 1292-5. https://doi.org/10.1001/jama.273.16.1292 [a]
- Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? Evidence Based Medicine Working Group. *JAMA* 1995; 273: 1610-3. https://doi.org/10.1001/jama.273.20.1610 [b]

- Robinson S. Simulation. The practice of model development and use. London: Wiley, 2004
- Roter DL, Hall JA, Merisca R, et al. Effectiveness of interventions to improve patient compliance: a meta-analysis. *Med Care* 1998; 36: 138-61. https://doi.org/10.1097/00005650-199808000-00004
- Sands BE, Sandborn WJ, Panaccione R, et al.; UNIFI Study Group. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* 2019; 381: 1201-14. https://doi.org/10.1056/NEJMoa1900750
- Salanti G. Assumptions of NMA. CMIMG meeting – Stream 2. Bristol, 16-17 July 2013. Available at https://methods.cochrane.org/cmi/sites/methods.cochrane.org.cmi/files/uploads/NMA_Assumptions.pdf (last accessed August 2022)
- Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *J Pharmacol Pharmacother* 2010; 1: 100-7. https://doi.org/10.4103/0976-500X.72352
- Souter MJ, Signorini DF. Meta-analysis: greater than the sum of its parts? *Br J Anaesth* 1997; 79: 420-1
- Torrance GW, Blaker D, Detsky A, et al. Canadian guidelines for economic evaluation of pharmaceuticals. Canadian Collaborative Workshop for Pharmacoeconomics. *Pharmacoeconomics* 1996; 9: 535-59. https://doi.org/10.2165/00019053-199609060-00008