Technical Document
Uncertainty

Version 1.0

09.10.2008

Contact:

Institute for Quality and Efficiency in Health Care (IQWiG)
Dillenburger Straße 27
D-51105 Cologne
Tel.: +49 (0)221 / 35685-0
Fax: +49 (0)221 / 35685-1
E-mail: knb-methoden@iqwig.de
Table of Contents

Table of Contents .................................................................................................................... i
1   Introduction .......................................................................................................................... 1
2   A Short Review of the Literature on the Efficiency Frontier and the Empirical
    Developments ..................................................................................................................... 3
3   The Issue of Efficiency Measurement and its Degree of Uncertainty .............................. 5
    3.1   Dealing with Unbiasedness of the Estimates ............................................................. 5
    3.2   Constructing Confidence Intervals .............................................................................. 7
4   Efficiency Frontiers with Small Sample Data Sets Based on Secondary Data .......... 9
    4.1   Constructing an Efficiency Frontier ........................................................................... 9
    4.2   Dealing with Uncertainty ............................................................................................ 11
References .................................................................................................................................. 18
1 Introduction

According to what is stated in the main report, the efficiency frontier (EF) concept was chosen “as the framework for IQWiG health technology assessments because it meets the requirements imposed by German law while remaining consistent with basic international methodological standards. Specifically, the method allows consideration of the efficiency of resource use in a single therapeutic area, with retention of the benefits as established by IQWiG and avoidance of discrimination” [1].

In an EF context, any evaluation of a new product is conducted under the assumption that all the available technologies are efficient. Without this assumption, it would be possible for decision makers either to adjust the mix of technologies before considering new technologies or for the manufacturer to compare the new technology with an alternative technology that is not on the frontier. Subsequently, the new technology could be accepted for reimbursement without actually improving overall efficiency.

The EF approach is an accepted evaluation instrument recommended and applied internationally. It informs decision makers about the level of efficiency of a specific technology, under the premise that this decision model “provides unbiased estimates of expected cost and effects, and of decision uncertainty, in a timely fashion and within resource constraints as determined by the decision maker that commissions the model” [2]. Unbiasedness of the estimates and their degree of accuracy are among the most important aspects to consider and check if an efficiency frontier aims to be informative for decision makers. Unfortunately, under the term “uncertainty” we quite often confuse the uncertainty related to the estimate of the true value of a parameter with uncertainty related to the confidence intervals surrounding the estimate (which should provide the degree of precision of that estimate, rather than its reliability). In fact, it is possible to have “unbiased” estimates with large confidence intervals (which are therefore of no use for decision making) and “biased” estimates with narrow confidence intervals (also of no use for decision making, but for different reasons).

Decision makers should receive not only unbiased information in point estimates of the efficiency levels, but also on their confidence intervals. Confidence intervals are of interest in the sense that they convey information on estimate precision. Therefore, “if we report that a
firm is 80% efficient, we would like to be able to distinguish between cases where our uncertainty is small (for example when we can be relatively sure that the true efficiency level is between 78% and 82%) and cases when our uncertainty is large (for example when we can be relatively sure only that it is between 65% and 95%)” [3, p. 91].

It is then very important to be aware that assessing the reliability of point estimates (for example, the potential amount of bias in the estimate of an EF) reduces the level of uncertainty in a way that is different from providing information on the confidence interval of a point estimate (how to construct confidence interval around estimates). In the following sections, our main interest in this technical appendix concerns the first issue, namely how to provide information on the level of bias in the estimates, the second issue being less relevant in our context. In fact, given an unbiased estimate and an estimator that produces efficient and consistent estimates, the degree of accuracy is just a matter of sample size!

As we will see, there are different ways to compute or estimate EFs. In general, all these methods work under the assumption that researchers have a large number of observations to estimate model parameters. Unfortunately, this may not always be the case, especially regarding the tasks for which IQWiG is commissioned. Whenever the number of observations available is limited (we can imagine examples with even less than 10 observations), it is hard to use inference formulas to obtain confidence intervals. It is for this reason that in this appendix we first review the standard methodologies used for estimating or computing EFs with large data sets and then explore alternative approaches to “construct” EFs so as to obtain uncertainty measures based on small data sets.

The remaining part of the technical appendix is organized in 3 sections. In Section 2 we present a short review of the literature on efficiency frontiers and the relevant empirical developments. Section 3 broadly discusses measurement issues and the role of uncertainty in efficiency frontiers when a large amount of sample data exists. Section 4 concentrates instead on the derivation of an EF and the measurement of uncertainty in a context with small sample data sets.
2 A Short Review of the Literature on the Efficiency Frontier and the Empirical Developments

In order to better understand how we address the uncertainty issue in the present context, it is useful to review briefly the theoretical and empirical literature on the EF.

The measurement of economic efficiency dates back to the original work initiated by Farrell [4], who was the first researcher to recognize the close link between efficiency and the use of frontier functions, and who introduced a method to decompose the overall efficiency of a production unit into its technical and allocative components.

Since then, the literature on this topic has grown enormously. From the initial concept of the unit efficient isoquant developed by Farrell [4], several alternative ways of specifying the technological set of a producer, i.e., production, cost, revenue or profit functions have been developed. Distance functions have also been employed as measures for technical and allocative efficiency. Another important development has been to move from the simple one-input-one-output model hypothesized by Farrell to its multi-input-multi-output generalization, whose mathematical programming characterization was proposed by Charnes, Cooper and Rhodes [5].

While the theoretical concept of defining an efficiency frontier function against which to measure the current performance of productive units has remained almost unchanged since the initial work by Farrell, the empirical counterpart “research activity” has evolved in several directions, producing different techniques to either calculate or estimate efficient frontiers.

In a thorough review of the literature on efficiency frontiers, Murillo-Zamorano [6] provides a very clear and complete taxonomy of the different ways of classifying these techniques. According to the author, a first important classification refers to the distinction between parametric and non-parametric methods

that is, between techniques where the functional form of the efficient frontier is imposed a priori and those where no functional form is pre-established but one is calculated from the sample observations in an empirical way. The non-parametric approach has been traditionally assimilated into Data Envelopment Analysis (DEA); a mathematical programming model applied to observed data that provides a way for the construction of production frontiers as well as for the calculus of efficiency scores relatives [sic] to those
constructed frontiers. With respect to parametric approaches, these can be subdivided into deterministic and stochastic models. The first are also termed ‘full frontier’ models. They envelope all the observations, identifying the distance between the observed production and the maximum production, defined by the frontier and the available technology, as technical inefficiency. The deterministic specification, therefore, assumes that all deviations from the efficient frontier are under the control of the agent. However, there are some circumstances out of the agent’s control that can also determine the suboptimal performance of units. Regulatory-competitive environments, weather, luck, socio-economic and demographic factors, uncertainty, etc., should not properly be considered as technical efficiency. The deterministic approach does so, however. Moreover, any specification problem is also considered as inefficiency from the point of view of deterministic techniques. On the contrary, stochastic frontier procedures model both specification failures and uncontrollable factors independently of the technical inefficiency component, by introducing a double-sided random error into the specification of the frontier model. A further classification of frontier models can be made according to the tools used to solve them, namely the distinction between mathematical programming and econometric approaches. The deterministic frontier functions can be solved either by using mathematical programming or by means of econometric techniques. The stochastic specifications are estimated by means of econometric techniques only. Most of the literature related to the measurement of economic efficiency have [sic] based their [sic] analysis either on any of the above parametric or on non-parametric methods. The choice of estimation method has been an issue of debate, with some researchers preferring the parametric (e.g. Berger, 1993) and others the non-parametric (e.g. Seiford and Thrall, 1990) approach. The main disadvantage of non-parametric approaches is their deterministic nature. Data Envelopment Analysis, for instance, does not distinguish between technical inefficiency and statistical noise effects. On the other hand, parametric frontier functions require the definition of a specific functional form for the technology and for the inefficiency error term. The functional form requirement causes both specification and estimation problems [6, pp. 35,36].

As shown in the next sections, dealing with uncertainty when estimating efficiency frontiers depends heavily on the specific technique adopted.
3 The Issue of Efficiency Measurement and its Degree of Uncertainty

As noted above, a large body of literature has developed which addresses the issue of computing or estimating the level of efficiency and its level of precision since the seminal paper by Farrell [4].\(^1\) Our goal here is to present techniques available for dealing with the unbiasedness and precision of the estimate within the restricted set of models of efficiency frontiers that IQWiG accepts as tool to compare alternative medical interventions. Following the approach presented in the main document on the assessment of the relation of benefits to costs in the German statutory health care system, this implies that only single-input-single-output frontiers and their estimation techniques will be examined.

From an empirical point of view, the computation of EFs by means of mathematical programming or econometric techniques requires availability of information on input and output. The larger the set of information is, the better the outcome of the analysis will be.\(^2\) While using the econometric approach, the standard statistical properties of an optimal estimator need to be respected: a) unbiasedness, b) efficiency, and c) consistency.

3.1 Dealing with Unbiasedness of the Estimates

In computing or estimating an efficiency frontier, several confounding elements can intervene in the process, and the resulting levels of efficiency will then be affected. These confounding elements can affect the value of the estimates (thus providing “biased” estimates). In a modelling approach, confounding elements can arise due to 1) model choice, 2) missing variables, and 3) measurement errors in both the dependent and explanatory variables.

The model-choice problem refers to the particular functional form that the modeller has to specify as the formal approximation of the true relationship between dependent and independent variables. In the specific context of efficiency frontiers, enormous progress has been made with the adoption of the flexible functional forms that are second-order linear

\(^1\) It is not the main goal of this paper to review all such studies. Among the several publications available, the interested reader can refer, for example, to the work by Murillo-Zamorano [6] or Kim and Schmidt [3] and to the literature referenced there.

\(^2\) Concerning the econometric approach, standard econometric techniques, such as Ordinary Least Squares (OLS), have the limitation of estimating “an average relationship” rather than “the optimal relationship”. Adopting a frontier function approach has the more appealing characteristic that these parameters are closer to the theoretical concept of the optimal relationship between inputs and outputs of a production process.
approximations to unknown true functions. The side effect of such functional forms is the larger number of parameters to estimate. However, as long as the number of observations is large enough, only computational problems can limit their estimation.

The missing-variables problem refers to the impossibility for a researcher to observe all relevant information useful to reveal the true relationship between input and output. Whatever is observable should be introduced as an explanatory variable in the model. This is especially true when dealing with patient-level data representing a high heterogeneity in their responses to treatment because of personal characteristics. Whatever is observable by researchers in terms of heterogeneity should be included in the model to explain variation.

For example, the risk of developing cervical cancer may depend on family history. In principle, this can be observed and subsequent decisions, such as the decision of whether to screen, based on this observation. This contrasts with variation in the rate of disease progression which is unobservable at the time at which the decision to screen is made. Thus, one could not decide to screen only those patients whose cancer would develop at a fast rate. When estimating the cost-effectiveness of an intervention for a heterogeneous population, one can condition on the observed characteristics and separate the overall group into homogenous subpopulations within which patients have identical observed characteristics. The model can then be run separately for each homogenous group to generate estimates of cost-effectiveness conditional on each set of observed characteristics. Adoption decisions can then be made for each of these mutually exclusive and identifiable patient groups [2, p. 246].

Finally, the measurement-error problem refers to the capability of the researcher to measure all variables properly. In many cases it can be very difficult to measure some phenomena and, depending on the degree of approximation used, the level of bias may become relevant. Whenever an error in variable issues arises that cannot be reduced through measurement, researchers can characterize it using empirical distributions.

For example, the rate at which an individual’s cervical cancer develops will vary between patients. We can describe the distribution of the rate of cancer progression by counting the number of patients who progress at different rates. Nevertheless, further investigation would not reduce variation in the rate of progression. Another example is where, given a probability of an event occurring, such as death, the realization of that event can be
imagined as being governed by a lottery. So we may know with certainty that the probability of death is, for example, 5%, but we do not know which 5% of people will die [2, p. 245].

3.2 Constructing Confidence Intervals

Under the assumption that our estimates are unbiased, constructing confidence intervals is a fairly simple task. Murillo-Zamorano [6] presents a quite complete and updated review of methods to compute confidence intervals under both mathematical programming and econometric approaches. In the first case, although non-parametric techniques are deterministic in nature, which has prompted researchers to describe them as non-statistical methods, recent developments have shown that it is possible to define a statistical model allowing for the determination of statistical properties of the non-parametric frontier estimators. These new methods are presented in Section 2.3 of the article by Murillo-Zamorano [6].

Concerning the econometric approach, we should distinguish between the deterministic and the stochastic frontier approach. The deterministic econometric approach uses the same theoretical apparatus developed for the mathematical programming approach, but allows the estimation rather than the ‘calculation’ of the parameters of the frontier functions. More important, statistical inference based on those estimates is possible and easy to derive. Several techniques, such as Modified Ordinary Least Squares (e.g., [7]), Corrected Ordinary Least Squares (e.g., [8]) and Maximum Likelihood Estimation (e.g., [9]), have been developed in the econometric literature in order to estimate these deterministic full-frontier models. In any case, neither programming models nor deterministic econometric approaches provide accurate measures of the input-output relationship, given that both are unable to take into account random shocks.

Aigner, Lovell and Schmidt (1977), Meeusen and van den Broeck (1977) and Battese and Corra (1977) simultaneously developed a Stochastic Frontier Model (SFM) that, besides incorporating the efficiency term into the analysis (as do the deterministic approaches) also captures the effects of exogenous shocks beyond the control of the analysed units. Moreover, this type of model also covers errors in the observations and in the measurement of outputs” [6, p. 48].
Section 3.1 in the paper by Murillo-Zamorano [6] discusses the main results in terms of inference when dealing with cross-sectional data.
4 Efficiency Frontiers with Small Sample Data Sets Based on Secondary Data

4.1 Constructing an Efficiency Frontier

When one is forced to work with few observations, the construction of an EF is hard to pursue on the basis of mathematical programming and econometric techniques. In all these cases, the EF can be empirically constructed using the concept of Incremental Cost-effectiveness Ratio (ICER). Following Briggs [10] and Briggs et al. [11], an EF can be obtained by simply inverting the axes of a cost-effectiveness plane. A cost-effectiveness plane can be represented as in Figure 4-1. It simply shows the difference (treatment minus control) in effectiveness ($\Delta E$) per patient against the difference in cost ($\Delta C$) per patient. By adopting the convention that the horizontal axis represents the effectiveness difference, the slope of the line joining any point on the plane to the origin is equal to the incremental cost-effectiveness ratio (ICER), $\Delta C/\Delta E$, the statistic of interest in cost-effectiveness studies. Several interventions could be accommodated on the same plane, each of them with its own ICER. An example of a multi-intervention graph is presented in Figure 4-2.

![Figure 4-1](image-url)

**Figure 4-1** Cost-effectiveness plane with the location of results for the example of the early endoscopy trial. ICER = incremental cost-effectiveness ratio. NW = north-west; NE = north-east; SW = south-west; SE = south-east (adapted from [10]).
In Figure 4-1, the value of ICER, equal to €1700, is obtained by considering that early endoscopy for dyspeptic patients cost an additional €80 per patient and resulted in an additional 4.7% of patients free of dyspeptic symptoms at 12 months. An EF is obtained from a cost-effectiveness plane by simply inverting the axes. Figure 4-2 is a clear example of how to construct an EF with only 6 observations!

![Baseline cost-effectiveness results on the cost-effectiveness plane presented as “efficient frontier”. PPI = proton pump inhibitor; H2RA = H2-receptor antagonist; PA = prokinetic agent; GORD = gastro-oesophageal reflux disease; GFW = GORD-free week (adapted from [11]).](image)

**Figure 4-2** Baseline cost-effectiveness results on the cost-effectiveness plane presented as “efficient frontier”. PPI = proton pump inhibitor; H2RA = H2-receptor antagonist; PA = prokinetic agent; GORD = gastro-oesophageal reflux disease; GFW = GORD-free week (adapted from [11]).

Similar examples could be constructed using data from other sectors. Figure 4-3 reports an example taken from Delea [12] based on lipid-lowering drugs.
Efficiency Frontiers with Small Sample Data Sets Based on Secondary

Figure 4-3 Absolute efficiency frontier given by the staircase for lipid-lowering therapies: effectiveness measured using reduction rate in LDL-C level. LDL-C, low density lipoprotein cholesterol; N = niacin (2.5 mg); CO = colestipol (4, 10, 15, 16, 20 mg); CH = cholestyramine (24 mg); F = fluvastatin (20, 40 mg); S = simvastatin (5, 10, 20, 40 mg); P = pravastatin (10, 20, 40, 80 mg); A = atorvastatin (10, 20, 40, 80 mg); L = lovastatin (20, 40, 80 mg) (adapted from [12]).

4.2 Dealing with Uncertainty

The above-mentioned examples present cost and benefit values for each intervention (or incremental cost and benefit values). The EF frontier is constructed based on these values. Information for decision makers derived from these data is valuable as long as it is reliable (or unbiased). The uncertainty surrounding the reliability of these estimates is what has encouraged the analysis of the role of uncertainty in economic evaluation. Briggs [10] reviews these statistical approaches to the treatment of uncertainty, both for primary studies where patient-level cost and effect information is available and in secondary studies that typically use decision-analytical methods to synthesize information from the published literature.

3 For more details on the results presented in this section the interested readers can refer to [10,11,13].
With the increasing use of the clinical trial as a vehicle for economic evaluation, there has been increasing interest in the use of statistical methods for handling uncertainty in patient-level data on both costs and effects. However, even for studies based on secondary data, the use of statistical methods for specifying parameter distributions in so-called probabilistic analyses is becoming increasingly common and is now the method recommended by a number of good-practice guides and regulatory agencies [10, p. 551].

Among the most commonly used statistical methods to deal with uncertainty, Briggs [10] and Briggs et al. [11] suggest using the “acceptability curves” approach, as it presents much more information on uncertainty than do confidence intervals. For a better understanding of this concept, we need first to review some basic concepts related to Fieller confidence limits and to bootstrap confidence intervals. In doing so we follow Briggs [10].

![Figure 4-4](image)

**Figure 4-4** Confidence intervals for cost-effectiveness ratios for the example of the early endoscopy. (a) Parametric Fieller’s theorem, and (b) non-parametric bootstrapping. PFD = patients free of dyspepsia (adapted from [10]).

**Fieller Confidence Limits**

Fieller’s approach is based on the assumption that cost and effect differences follow a joint normal distribution, rather than the cost-effectiveness ratio itself. Figure 4-4(a) shows the assumption of joint normality on the cost-effectiveness plane for the ICER estimates reported
in Figure 4-1: three ellipses of equal density are plotted covering 5%, 50% and 95% of the integrated joint density. Also plotted are the estimated confidence limits using Fieller’s theorem (€300 to -€1100), represented by the slopes of the lines on the plane passing through the origin. Note that the ‘wedge’ defined by the confidence limits falls inside the 95% ellipse; this is because Fieller’s approach automatically adjusts to ensure that 95% of the integrated joint density falls within the wedge.

**Bootstrap Confidence Intervals**

This is a re-sampling procedure that estimates an empirical sampling distribution for the statistic of interest rather than relying on parametric assumptions. Obtaining a large number of samples generates a vector of bootstrap replicates of the statistic of interest, which is the empirical estimate of the statistic’s sampling distribution. One thousand bootstrapped effect and cost differences for the ICER estimate reported in Figure 4-1 are plotted on the cost-effectiveness plane in Figure 4-4(b). Confidence limits can be obtained by selecting the 2.5th and 97.5th percentiles of the bootstrapped replications ordered from the most favourable to the least favourable cost-effectiveness ratio; this effectively ensures that 95% of the estimated joint density falls within the wedge on the cost-effectiveness plane defined by the confidence limits. As is clearly apparent from Figure 4-4(b), the bootstrap estimate of the joint density and the bootstrap confidence limits (€300 to -€1200) are very similar to those generated by Fieller’s theorem.

In terms of Figures 4-2 and 4-3, these procedures should be replicated for each point in the graph. This will produce a graph like the one reported in Figure 4-5. (We will discuss the meaning of these figures in more detail later on.)
Figure 4-5  Ten thousand Monte Carlo simulations of the six management strategies in the example for gastro-oesophageal reflux disease (GORD) mapped on the cost-effectiveness plane. 1 = intermittent proton pump inhibitor (PPI); 2 = maintenance PPI; 3 = maintenance H$_2$-receptor antagonist (H$_2$RA); 4 = step-down maintenance prokinetic agent; 5 = step-down maintenance H$_2$RA; 6 = step-down maintenance PPI (adapted from [10]).

The Acceptability Curve

Fieller confidence limits and bootstrap confidence intervals work fine until we have to compare ICERs of the same sign but lying on both sides of the efficiency frontier. In this case, these approaches fail to compare ICERs properly. As suggested by Briggs [10], a solution to this problem can be found. In fact, in terms of the bootstrap replications on the cost-effectiveness plane in Figure 4-4(b), we could summarize uncertainty by considering how many of the bootstrap replications fall below and to the right of a line with a slope equal to a pre-specified ceiling ratio (R$_c$), lending support to the cost-effectiveness of the intervention.

Alternatively, if we are happy with an assumption of joint normality in the distribution of costs and effects, we can consider the proportion of the parametric joint density that falls on the cost-effective surface of the cost-effectiveness plane. Of course, the appropriate
value of $R_c$ is itself unknown. However, it can be varied in order to show how the evidence in favour of cost-effectiveness of the intervention varies with $R_c$. The resulting curve for the early endoscopy example given in Figure 4-1 and based on the joint normal assumption shown in Figure 4-4(a) is presented in Figure 4-6 and has been termed a “cost-effectiveness acceptability curve” as it directly summarizes the evidence in support of the intervention being cost-effective for all potential values of the decision rule [10, pp. 555-556].

![Cost-effectiveness acceptability curve](image)

**Figure 4-6** Cost-effectiveness acceptability curve for the example of the early endoscopy (adapted from [11]).

Several of these curves could be obtained for each intervention reported in Figure 4-5. This leads to a graph like the one reported in Figure 4-7. As expected, strategy 4 does not feature in Figure 4-7, indicating that it is never a contender for cost-effectiveness (it lies below the EF). Strategy 6 does feature there, although it never achieves more than 13% of simulations suggesting it is cost-effective, even at the most favourable ceiling ratio (about €260 per day free of GORD symptoms). In practice, decision makers will probably want to choose between strategies 3, 1, 5 and 2 and their choice will most likely be governed by which strategy is presently considered standard practice, the ceiling ratio and the attitude towards risk.
Efficiency Frontiers with Small Sample Data Sets Based on Secondary

**Figure 4-7** Multiple acceptability curves for the six management strategies in the example for gastro-oesophageal reflux disease (GORD) showing the proportion of times each strategy was considered cost-effective as a function of the ceiling ratio; a log scale is employed to better illustrate the low values; see legend to Fig. 4-5 for brief details of the strategies 1 to 6 (adapted from [10]).

The piece of information still missing in this analysis for decision makers is the ceiling ratio, or a threshold for the willingness to pay. These thresholds are exogenous to the analysis on uncertainty and should be provided in advance. Having this information, Figure 4-7 may appear as in Figure 4-8. In this case, having a ceiling ratio of 10 implies that strategy “3” is going to be selected, while in the case of a ceiling ratio above € 264, strategy “2” may be selected.
Figure 4-8  Multiple acceptability curves for the six management strategies with thresholds for the willingness to pay (see legend to Fig. 4-5 for brief details of the strategies 1 to 6) (adapted from [11]).
References


