

# Practical Guidelines for Economic Evaluations Alongside Equivalence Trials

Judith E. Bosmans, PhD,<sup>1,2</sup> Martine C. de Bruijne, PhD/MD,<sup>1</sup> Hein P. J. van Hout, PhD,<sup>3</sup> Marleen L. M. Hermens, PhD,<sup>3</sup> Herman J. Adèr, PhD,<sup>4</sup> Maurits W. van Tulder, PhD<sup>1,2</sup>

<sup>1</sup>Health Technology Assessment Unit, Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands;

<sup>2</sup>Institute for Health Sciences, Faculty of Earth and Life Sciences, Vrije Universiteit, Amsterdam, The Netherlands;

<sup>3</sup>Department of General Practice, Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands;

<sup>4</sup>Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

## ABSTRACT

An effective treatment already exists for many diseases. In these cases the effectiveness of a new treatment may be established by showing that the new treatment is as effective as (i.e., equivalent to) or at least as effective as (i.e., noninferior to) the old treatment. For an economic evaluation accompanying a clinical equivalence or noninferiority trial it is important to decide before the start of the study on the appropriate research question. In many cases the objective of the economic evaluation will be to show equivalence or noninferiority of the cost-effectiveness of the treatments. This has major implications for the design and

analysis of the economic evaluation. In this article we propose methods for the analysis of economic equivalence and noninferiority studies that are similar to the methods applied to clinical equivalence and noninferiority trials. Furthermore, cost-effectiveness planes prove to be a valuable tool in the interpretation of the results in an economic equivalence or noninferiority trial. The concepts described in the article are illustrated using the results from an economic noninferiority trial.

**Keywords:** cost-effectiveness analysis, depression, economic evaluation, research methodology.

## Introduction

For many diseases effective treatments already exist. In these cases the research question of a study investigating these different treatment options may not be which treatment option is the superior one, but whether the different options are approximately equal to each other. Thus, the effectiveness of a new treatment may be established by showing that the new treatment is as effective as (i.e., equivalent to) or at least as effective as (i.e., noninferior to) the old treatment that was already shown to be effective, while having advantages like less side effects or an easier dose regimen for example. For clinical analyses there is much literature on the design and analysis of such so-called equivalence trials. Nevertheless, there are also cases in which one wants to demonstrate that two treatments are not only clinically, but also economically equivalent. Therefore, when conducting an economic evaluation alongside an equivalence or noninferiority trial, it is important to determine the appropriate research question beforehand. When the objective is to evaluate whether one of

the two treatments in a clinical equivalence or noninferiority trial has significantly lower costs a cost minimization analysis is indicated. Nevertheless, a drawback of cost minimization analyses is that costs alone are considered and not both costs and effects. Moreover, it is often relevant for decision-makers to know whether the new treatment is as expensive as (i.e., equivalent to) or not more expensive than (i.e., noninferior to) the old treatment. For example, in a trial comparing antidepressant treatment with psychotherapy in depression, it is relevant for doctors to know that both treatments are equivalent or at least noninferior in both clinical and economic aspects. The choice of treatment can then be based on other aspects, like the patient's preference. This is also true for a trial comparing two or more medications, for example two antidepressants from different pharmaceutical classes. The objective of this article is to describe the consequences for the design and analysis of an economic evaluation assessing equivalence or noninferiority of costs and cost-effectiveness.

### Introduction to the Example Trial

Concepts described in this article are illustrated using the results from a noninferiority trial we performed. In this randomized trial we compared standardized usual care by the general practitioner (GP) with and without antidepressants in patients with minor depressions.

*Address correspondence to:* Judith E. Bosmans, Vrije Universiteit, Faculty of Earth and Life Sciences, Institute for Health Sciences, W&N Building, Room U435, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. E-mail: [judith.bosmans@falw.vu.nl](mailto:judith.bosmans@falw.vu.nl)

10.1111/j.1524-4733.2007.00245.x

The rationale for starting this trial was that antidepressants are often prescribed although there is insufficient evidence for their effectiveness in this group of patients. When antidepressants have no additional benefits in terms of health effects over standardized usual care alone in patients with minor depressions disadvantages of antidepressant use such as adverse effects, labeling, medicalization, dependency, and costs prevail. Thus, if standardized usual care alone is at least as effective as standardized usual care plus antidepressants the prescription of antidepressants in this group of patients does not seem justified. One would expect that addition of antidepressants to standardized usual care by the GP is more expensive than standardized usual care alone. Nevertheless, it is also plausible to assume that patients who do not receive antidepressants, make more use of other medication and other health-care services. Thus, from an economic perspective, if standardized usual care alone is not more expensive than standardized usual care plus antidepressants routine prescription of antidepressants does not seem justified. Therefore, the hypothesis in the economic evaluation was that standardized usual care alone is at least as effective as and not more expensive than (i.e., noninferior to) standardized usual care plus antidepressants.

Patients who were diagnosed with minor or mild-major depression by their GP were eligible for the trial. Excluded were patients who were 17 years old or younger who currently received some form of depression treatment, who had comorbid psychiatric disorders or who were unable to complete the questionnaires. All patients received four consultations with their GP during which they received advice and support (standardized usual care). Half of the patients were randomly allocated to receive an antidepressant (standardized usual care plus antidepressants). In total, 181 patients were included in the study (standardized usual care: 96 patients, standardized usual care plus antidepressants: 85 patients). There were no differences in baseline sociodemographic and clinical characteristics between the two treatment groups. Complete cost data were available for 89 patients. Patients with complete follow-up were more likely to be previously treated for depression or to have Dutch origins and were less depressed at baseline. Results from an imputed analysis did not differ from the complete case analysis.

## Design and Analysis of Clinical Equivalence and Noninferiority Trials

### Design

The aim of a clinical equivalence trial is to show the therapeutic equivalence of two treatments [1]. Absence of a statistically significant difference between the treatments compared in a randomized controlled trial

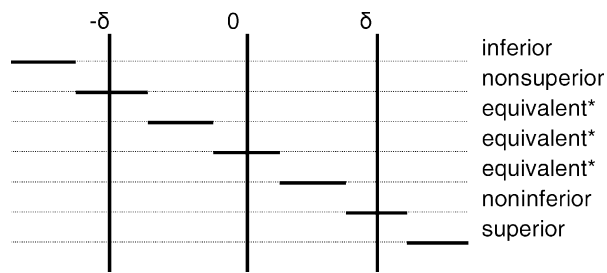
does not imply equivalence of the treatments [2]. Moreover, in trials with large sample sizes, it is possible to demonstrate a small, but statistically significant difference between treatment groups that is clinically unimportant.

In equivalence trials the hypothesis is that the new treatment and the control treatment are equally effective, as opposed to superiority trials where the hypothesis is that the new treatment is more effective than the control treatment. In an equivalence trial a clinically acceptable difference,  $\delta$ , has to be determined a priori. Because the aim of the study is to show that the new treatment is sufficiently similar to the old treatment  $\delta$  has to be selected so that any clinically unimportant difference is smaller than  $\delta$  [1,3].

### Analysis

If we have a predefined range of equivalence from  $-\delta$  to  $\delta$ , equivalence is demonstrated if the two-sided confidence interval around the observed difference between the two treatment groups lies entirely in this range of equivalence. Before the start of an equivalence study, a two-sided  $\alpha$  (i.e., the significance level) has to be defined. For measures where higher scores indicate improvement noninferiority is demonstrated if the lower limit of the one-sided confidence interval around the observed difference between the two treatment groups is equal to or greater than  $-\delta$  (noninferiority margin). In a noninferiority study a one-sided  $\alpha$  is used, which also has to be defined beforehand [1,4]. The European Agency for the Evaluation of Medicinal Products recommends using a one-sided  $\alpha$  of 2.5% or a two-sided  $\alpha$  of 5% [5]. After demonstrating equivalence or noninferiority, it is still possible to test for superiority of effects [4]. In traditional superiority trials statistical superiority is shown if the confidence interval around the difference in costs or effects does not contain zero. Clinically meaningful superiority for measures where higher scores indicate improvement can be shown if the lower limit of the confidence interval around the difference in effects is greater than  $\delta$ . This approach of checking the location of the confidence interval in relation to  $-\delta$  and  $\delta$  is demonstrated in Figure 1.

In superiority trials the intention-to-treat analysis is generally conservative, because inclusion of protocol violators and withdrawals tends to make the two treatment groups more similar. This means that there is a strong incentive in superiority trials to strictly adhere to the study protocol. Nevertheless, in an equivalence trial any blurring of the difference between the treatment groups increases the chance of finding equivalence, although in reality the trial may have had poor discriminatory power. Hence, in an equivalence trial the intention-to-treat analysis is no longer conservative. Because the per-protocol analysis includes only patients who properly followed the study protocol this



**Figure 1** Position of the confidence interval (black horizontal bar) in relation to  $-\delta$ , 0, and  $\delta$  with the appropriate conclusion for measures where a higher score indicates improvement. \*in these cases the conclusion noninferior or nonsuperior can also be drawn.  $\delta$  = clinically acceptable difference.

analysis is expected to enhance any difference between the treatment groups and thus to decrease the chance of declaring equivalence. For this reason it is recommended to carry out both an intention-to-treat and a per-protocol analysis in an equivalence trial, where the per-protocol analysis gives the most conservative estimate of the treatment effect. Nevertheless, there are circumstances when the protocol analysis is biased toward a conclusion of equivalence or noninferiority. This is the case if, for example, patients who do not respond to one of the two treatments drop out early. Therefore, it is important to examine the subgroup of patients excluded from the per protocol analysis carefully to make sure there is no selective drop out. Moreover, a higher withdrawal rate in one of the treatment groups because of lack of improvement is an indication in itself that the treatments are not equivalent or noninferior [1]. Because in general it will be more difficult to demonstrate equivalence in the per-protocol population, sample size calculations should be based on this population [1,3,5,6].

### Economic Evaluations Alongside Equivalence and Noninferiority Trials

#### Economic Equivalence Margin

In studies where the objective is to show equivalent or noninferior costs and cost-effectiveness, the approach is similar to the one applied in clinical equivalence or noninferiority studies. Before the start of the study, an acceptable difference in total costs,  $\theta$ , should be defined.  $\theta$  has to be selected so that any economically unimportant difference in costs is smaller than  $\theta$ .

One way to determine  $\theta$  may be by defining a relevant difference in resource use. Subsequently, this difference in resource use can be valued, preferably using standard costs. Finally, to get  $\theta$ , the different cost categories are added. When performing the study from a societal perspective, it is the opportunity cost of any difference in resource use that is important. Therefore, it is important to include all relevant direct and indirect costs in the determination of  $\theta$ .

Another approach to define  $\theta$  may be to equal  $\theta$  to a percentage of the available health-care budget. In all cases  $\theta$  should not be too small or too large in comparison with the expected total costs. For example, a cost difference of €25 will not be relevant when the expected total costs are approximately €5000, but will be relevant when the expected total costs are around €250. As  $\theta$  may vary depending on the context, setting, and location of the study and on the prevalence of the disorder under study, it is important to perform some kind of sensitivity analysis to investigate the uncertainty surrounding  $\theta$ . An approach to this is proposed in the section “Analysis” below.

In our example trial we used studies from literature and studies we previously performed to determine a relevant difference in resource use from a societal perspective. This resulted in a relevant difference of two visits to the GP, one outpatient visit, and 3 days of absenteeism from paid labor. Using Dutch guideline prices [7] that were adjusted for 2002 using consumer price index figures [8] and after rounding off upwards, this resulted in a mean cost difference of €500. The same approach can be used when the study is performed from a different perspective. Nevertheless, this will have consequences for the types of resource use that are included in the determination of  $\theta$ .

#### Analysis

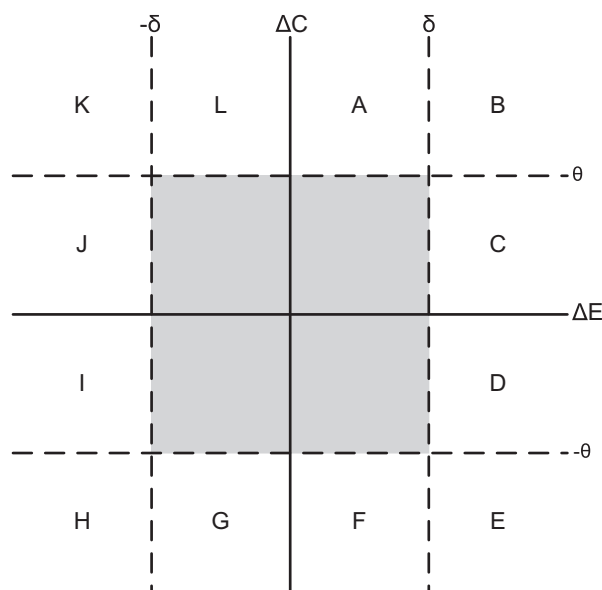
The first step in the analysis of an economic equivalence or superiority trial is the analysis of the difference in costs between the treatments. To show equivalence or noninferiority of costs, the confidence interval approach is used. Equivalence is demonstrated if the confidence interval around the difference in total costs lies entirely between the equivalence margins  $-\theta$  and  $\theta$ . A negative cost difference between the new intervention and the control intervention implies cost savings and a positive cost difference implies extra costs. Therefore, noninferiority of costs is shown if the upper limit of the confidence interval of the difference in total costs is smaller than  $\theta$ . Economically meaningful superiority is shown if the upper limit of the confidence interval around the difference in costs is smaller than  $-\theta$ .

In our example trial mean (SD) total costs in the standardized usual care group and in the standardized usual care plus antidepressants group were €4668 (5654) and €5418 (6003), respectively. The mean difference in total costs was -€751 and the 95% confidence interval around this difference ranged from -€3601 to €1522. The upper limit of the 95% confidence interval was greater than €500 ( $\theta$ ), so noninferiority for costs could not be shown.

#### Cost-Effectiveness Plane

The second and most important step in the analysis of economic equivalence and noninferiority trials is the joint analysis of costs and effects in a cost-effectiveness

Quadrant	Costs	Effects
A	More expensive	Equivalent
B	More expensive	Superior
C	Equivalent	Superior
D	Equivalent	Superior
E	Cheaper	Superior
F	Cheaper	Equivalent
G	Cheaper	Equivalent
H	Cheaper	Inferior
I	Equivalent	Inferior
J	Equivalent	Inferior
K	More expensive	Inferior
L	More expensive	Equivalent
	Equivalent	Equivalent



**Figure 2** Presentation of equivalence margins in the cost-effectiveness plane.  $\Delta C$  = incremental costs;  $\Delta E$  = incremental effects;  $\delta$  = equivalence margins for effects;  $\theta$  = equivalence margin for costs.

analysis. Bootstrapping can be used to evaluate the uncertainty surrounding the cost-effectiveness ratios. To visualize this uncertainty, the incremental costs and effects in all bootstrap samples are plotted on a cost-effectiveness plane. Subsequently, the equivalence or noninferiority margins,  $\delta$  (clinical) and  $\theta$  (economic), can be presented in the cost-effectiveness plane (Figs 2 and 3). For the sake of clarity, the margins as drawn in Figures 2 and 3 have no relation with a confidence interval around the Incremental Cost-Effectiveness Ratio (ICER). Position of a cost-effect pair in the area within the equivalence margins or noninferiority margins indicates that the differences in costs and/or effects for this pair are considered economically and/or clinically unimportant, because these differences are smaller than the predefined margin. Equivalence with regard to cost-effectiveness can be demonstrated if 100- $\alpha$ % of the cost-effect pairs lies in the area within

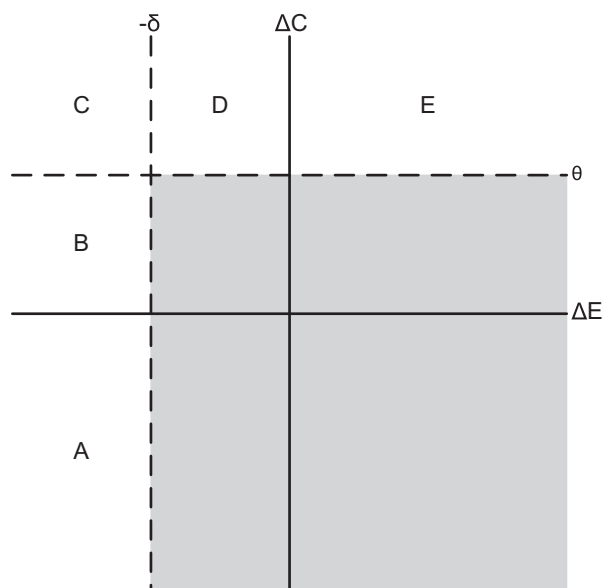
the equivalence margins (shaded area, Fig. 2). Noninferiority with regard to cost-effectiveness can be demonstrated if 100- $\alpha$ % of the cost-effect pairs lies right of the noninferiority margin for effects and below the noninferiority margin for total costs (shaded area, Fig. 3). When using this method, it is possible that the effects of the treatments are considered equivalent or noninferior, but not the costs. In these cases the conclusion of the study is that equivalence or noninferiority with regard to cost-effectiveness cannot be shown.

Figure 4 shows the cost-effectiveness plane of our example trial for the improvement in quality-adjusted life-years (QALYs) based on the EuroQol, including the noninferiority margins. The noninferiority margin for the improvement in QALYs was -0.03 and for costs, as mentioned above, €500. We used a one-sided  $\alpha$  of 2.5%, so 97.5% of the cost-effect pairs had to be situated in the noninferiority area to show noninferiority. As indicated in the figure, 61% of the cost-effect pairs was situated in the noninferiority area, so noninferiority with regard to cost-effectiveness for QALYs could not be shown.

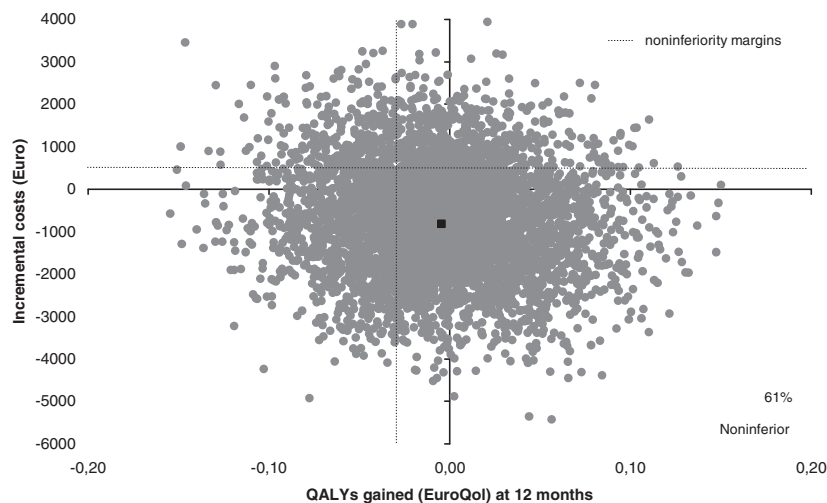
**ICER**

The traditional ICER may not be the best way to present the relation between costs and effects in an

Quadrant	Costs	Effects
A	Noninferior	Inferior
B	Noninferior	Inferior
C	More expensive	Inferior
D	More expensive	Noninferior
E	More expensive	Noninferior
	Noninferior	Noninferior



**Figure 3** Presentation of noninferiority margins in the cost-effectiveness plane.  $\Delta C$  = incremental costs;  $\Delta E$  = incremental effects;  $\delta$  = noninferiority margin for effects;  $\theta$  = noninferiority margin for costs.



**Figure 4** Cost-effectiveness plane for the improvement in QALYs based on the EuroQol for standardized usual care in comparison with standardized usual care plus antidepressants, including the noninferiority margins. QALYs, quality-adjusted life-years.

equivalence or noninferiority trial. The ICER is a continuous measure and corresponds to the slope of the line through the origin of the cost-effectiveness plane and through the point in the cost-effectiveness plane that is defined by the mean difference in costs and the mean difference in effects between the two treatment groups (point estimate of the ICER). In Figure 5a two example ICERs (lines A: ICER = 20 and B: ICER = 2) and two point estimates (points 1:  $\Delta C = 400$ ,  $\Delta E = 20$  and 2:  $\Delta C = 100$ ,  $\Delta E = 5$ ) are drawn. For every point estimate on a line through the origin the ICER is equal. In superiority research a judgment has to be made for every point on the line whether the effects are worth the costs. For example, for point 1 in Figure 5a: am I willing to pay €400 to gain 20 effect units? Nevertheless, in equivalence and noninferiority research the acceptable ICERs are limited by the equivalence or noninferiority margins. For example, the point estimates 1 and 2 on line A have the same ICER (20). Nevertheless, point estimate 1 is not relevant in an equivalence study, because the incremental costs are €400 and therefore greater than the equivalence margin for costs of €250, so this point estimate can never be associated with two equivalent treatments and line A is limited by the equivalence margins for costs. Line B on the other hand is limited by the equivalence margins for effects. The ICER drawn in Figure 5b is on one side limited by the noninferiority margin for costs and on the other side by the noninferiority margin for effects.

#### Presentation

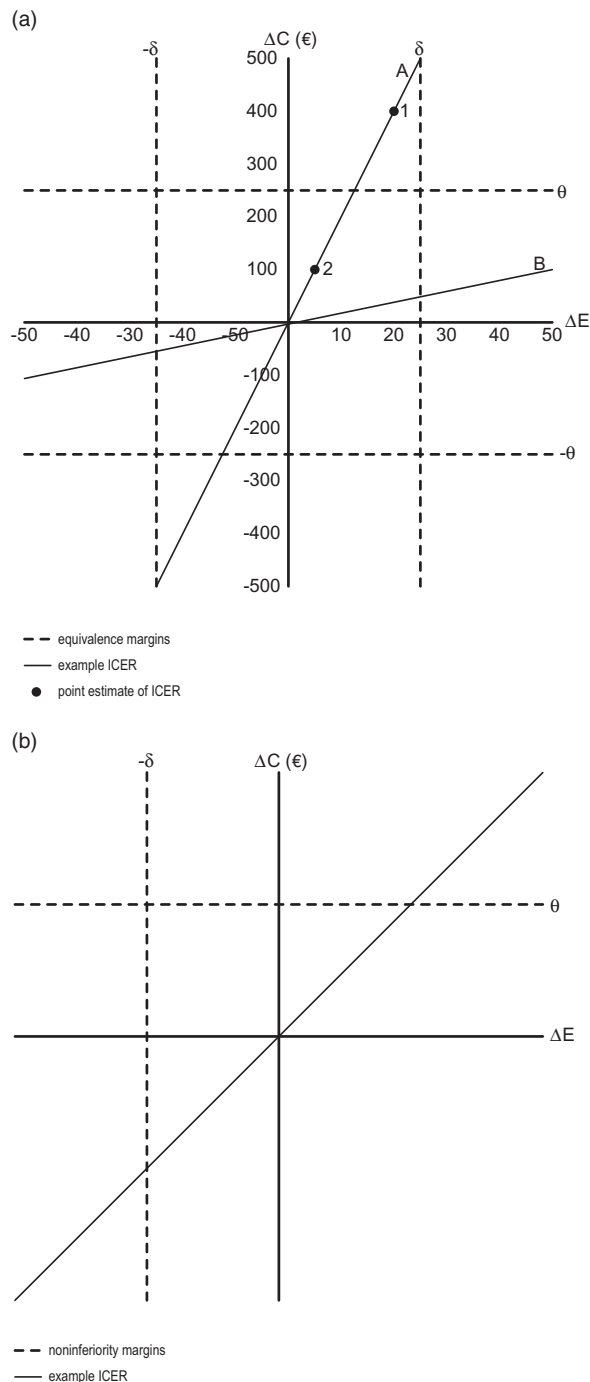
When presenting the results of an economic equivalence or noninferiority trial, it is important to show the confidence intervals around the differences in costs and effects. Costs, and therefore  $\theta$ , may vary in different settings or countries. Presentation of the confidence interval around the difference in total costs allows the reader to draw conclusions for a  $\theta$  that reflects his or

her specific circumstances. To determine what  $\theta$  is relevant to the reader, the reader has to consider both the resource use patterns and the unit costs used in his or her situation.

Considering the uncertainty regarding  $\theta$ , we recommend researchers to perform a sensitivity analysis in which  $\theta$  is varied. This exploration of the influence of  $\theta$  on the results is also important when no acceptable difference in costs is defined before the start of the study. The probability that the reference treatment is equivalent or noninferior to the other treatment for various values of  $\theta$ , although  $\delta$  is kept constant, can be drawn in a figure. This results in an equivalence or noninferiority curve (Fig. 6). It is also possible to vary  $\delta$  although  $\theta$  is kept constant. Nevertheless, generally there will be less uncertainty surrounding  $\delta$ . Moreover, a study will generally be sufficiently powered to show equivalence for a certain  $\delta$ .

The curve in Figure 6 at first sight resembles an acceptability curve. Because acceptability curves focus on the ICER as a continuous measure in relation to the amount a policymaker is willing to pay to gain one unit of effect, acceptability curves have no meaningful interpretation in economic equivalence and noninferiority research. In economic equivalence and noninferiority research, two separate and independent margins for costs and effects are defined, and we are interested in the proportion of cost-effect pairs that lies in the equivalence or noninferiority area. If chosen well the cost margin in itself incorporates some consideration of the amount of money a policymaker is willing to pay, or, in other words, considers (un)important.

It can be seen from Figure 6 that when  $\theta$  increases, so does the probability to demonstrate noninferiority. Given a particular noninferiority margin for effects ( $\delta$ ), the noninferiority curve in Figure 6 cuts the vertical axis at the probability of noninferior costs if standardized usual care alone is expected to have statistically significantly lower costs than standardized usual care



**Figure 5** (a) Examples of ICERs drawn in the cost-effectiveness plane for an economic equivalence study. (b) Examples of ICERs drawn in the cost-effectiveness plane for an economic noninferiority study. ICER, Incremental Cost-Effectiveness Ratio.

plus antidepressants. In other words, the probability of noninferior costs of standardized usual care alone in comparison with standardized usual care plus antidepressants if  $\theta = \text{€}0$ . The curve is asymptoting to the probability of noninferior effects given a particular  $\delta$ . An equivalence curve cuts the vertical axis at the prob-

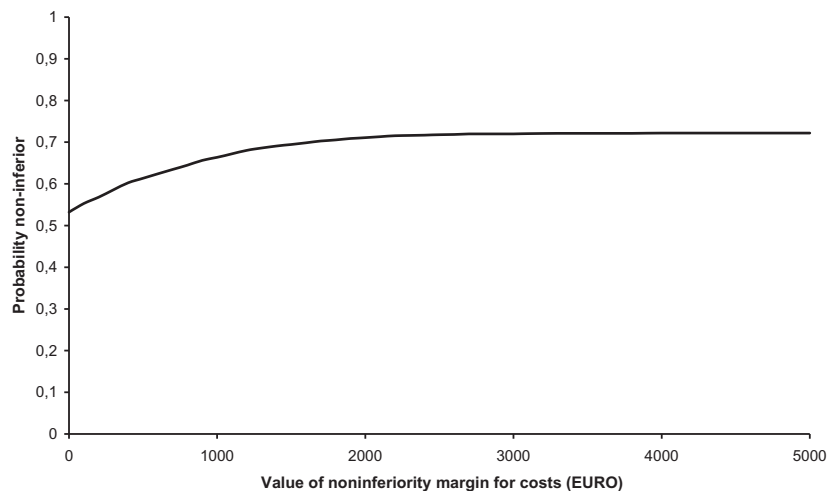
ability of equivalent costs if  $\theta = \text{€}0$  and asymptotes to the probability of equivalent effects given a particular  $\delta$ .

**Discussion**

In this article we describe how methods for equivalence and noninferiority trials can be applied to economic evaluations in this type of research. Conclusions are drawn using the position of a confidence interval in relation to a predefined acceptable difference in costs and effects. Cost-effectiveness planes make the evaluation and interpretation of the results for both costs and effects in an economic equivalence or noninferiority trial simpler, as the proportion of cost-effect pairs in the equivalence or noninferiority area can be estimated at a glance. Equivalence or noninferiority curves indicate the probability of equivalence or noninferiority for a range of equivalence or noninferiority margins. These curves show the uncertainty surrounding these margins and are a useful tool when no acceptable difference in costs has been specified before the start of the study.

We have already stated in the main section that after testing for equivalence or noninferiority, it is still possible to test for clinically or economically meaningful superiority, using the confidence interval approach. This approach may also be relevant in traditional superiority research. Evidence of a statistically significant difference does not imply that the difference found is clinically important. By defining a clinically important difference, basing sample sizes on this difference, and using the confidence interval approach described here to test for superiority, a decision can be given on the clinical importance of the results found in the study.

We agree with Briggs and O'Brien [9] that in economic evaluations the focus should be on cost-effectiveness analyses and that the use of cost-minimization analysis usually is not appropriate. They point out that a cost-minimization analysis is only appropriate in studies that have been designed to show therapeutic equivalence. In this article we argue that in economic evaluations accompanying clinical equivalence or noninferiority studies, it is important to decide before the start of the study on the appropriate economic research question. In many cases it may be useful to have the disposal of two clinically equivalent treatments that are either economically equivalent (i.e., new treatment as expensive as old treatment) or economically noninferior (i.e., new treatment not more expensive than old treatment) to each other. In these situations it is important that the equivalence margins for costs and effects are defined a priori. Analysis is then based on the relationship between costs and effects. Which one of two equivalent or noninferior treatments is eventually chosen depends on other factors than costs and effects. Side effects can play not



**Figure 6** Noninferiority curve given a noninferiority margin of  $-0.03$  for improvement in QALYs based on the EuroQol for standardized usual care in comparison with standardized usual care plus antidepressants. In this figure, the probability is drawn that counseling alone is noninferior to counseling plus antidepressants for various values of  $\theta$  although  $\delta$  is kept constant. QALYs, quality-adjusted life-years.

only an important role in this decision but also preferences of patients and health-care professionals, as well as the availability and burden of the treatments. This means that it is particularly important in equivalence research to record factors that may influence this decision process, during the study.

A problem with economic equivalence studies is that to provide useful information for health policymakers economic evaluations usually have a pragmatic study design, so that they resemble routine clinical practice as much as possible. This results in a diminished contrast between the treatment groups, which conflicts with the requirements for an equivalence trial, where diminishing of the contrast increases the chance of finding equivalence. The most common reason to perform an equivalence or noninferiority trial is that there already exists an effective treatment and that it is considered unethical to withhold this effective treatment from patients. When there is already some evidence for the efficacy of the new treatment (which must be the case because this will concern phase III trials), an economic equivalence or noninferiority trial is preferred over a superiority trial. When performing an (economic) equivalence or noninferiority trial care has to be taken that the study protocol is followed strictly.

To calculate sample sizes for equivalence and noninferiority trials, researchers can use standard sample size calculations [10,11]. Nevertheless, the resulting sample sizes usually need to be larger than in superiority trials, because the chosen clinically and economically important differences should generally be smaller than in a superiority trial [1]. In economic equivalence or noninferiority trials even larger sample sizes are needed, because the distribution of cost data typically is heavily skewed [12]. Often it will be not possible for financial, logistic or ethical reasons to include as many patients as are needed for the trial to have sufficient power for economic outcomes. Therefore, a common problem in economic evaluations is that they are

underpowered. The choice of  $\theta$  is inextricably bound up with the power of the study. Nevertheless, the choice of  $\theta$  is difficult; if  $\theta$  is chosen too small, this will lead to unnecessarily large trials. The other way round, if  $\theta$  is chosen too large, this will lead to a conclusion of equivalence for substantially different treatments. Some advice on how to determine  $\theta$  is given in this article. To investigate the uncertainty surrounding  $\theta$ , we propose to draw an equivalence or noninferiority curve. In this curve the probability of equivalent or noninferior cost-effectiveness is drawn for varying values of  $\theta$ , although  $\delta$  is kept constant.

## Conclusion

The availability of effective treatments for many diseases means that equivalence and noninferiority trials are becoming increasingly important in health-care research [13–15]. These trials may influence health policy and should be accompanied by an economic evaluation. It is important to decide before the start of the study on the appropriate research question for the economic evaluation. In many cases the objective of the economic evaluation will be to show equivalent or noninferior cost-effectiveness of the treatments. In this article we showed how cost-effectiveness techniques can be used to show noninferior or equivalent costs and cost-effectiveness. This has major implications for the design, analysis and report of the economic evaluation. The most important consequence of such a research question is that acceptable differences in costs and effects must be defined a priori. Analysis of effects and costs is based on the position of confidence intervals in relation to these acceptable differences. Cost-effectiveness planes prove to be a valuable tool in the interpretation of the results for both costs and effects in an economic equivalence or noninferiority trial. Equivalence or noninferiority curves show the uncertainty surrounding the equivalence or noninferiority

margins and are also useful when no acceptable difference in costs has been specified in advance of the study.

Source of financial support: no funding was acquired to assist in the preparation of this manuscript. Conflicts of interest: none declared.

## References

- 1 Jones B, Jarvis P, Lewis JA, et al. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36–9.
- 2 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
- 3 Ebbutt AF, Frith L. Practical issues in equivalence trials. *Stat Med* 1998;17:1691–701.
- 4 Rohmel J. Therapeutic equivalence investigations: statistical considerations. *Stat Med* 1998;17:1703–14.
- 5 EMEA. Points to Consider on Switching Between Superiority and Non-Inferiority. London: EMEA, 2000; 7–27.
- 6 Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. *Drug Info J* 1999; 33:109–18.
- 7 Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for costing in economic evaluations. *Pharmacoeconomics* 2002;20:443–54.
- 8 Statistics Netherlands/Central Bureau voor de Statistiek, 2002.
- 9 Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ* 2001;10:179–84.
- 10 Whitley E, Ball J. Statistics review 4: sample size calculations. *Crit Care* 2002;6:335–41.
- 11 Al MJ, Van Hout BA, Michel BC, et al. Sample size calculation in economic evaluations. *Health Econ* 1998;7:327–35.
- 12 Briggs A. Economic evaluation and clinical trials: size matters. *BMJ* 2000;321:1362–3.
- 13 Szegedi A, Kohnen R, Dienel A, et al. Acute treatment of moderate to severe depression with hypericum extract WS 5570 (St John's wort): randomised controlled double blind non-inferiority trial versus paroxetine. *BMJ* 2005;330:503–8.
- 14 Kinley H, Czoski-Murray C, George S, et al. Effectiveness of appropriately trained nurses in preoperative assessment: randomised controlled equivalence/non-inferiority trial. *BMJ* 2002;325:1323–7.
- 15 Addington DE, Pantelis C, Dineen M, et al. Efficacy and tolerability of ziprasidone versus risperidone in patients with acute exacerbation of schizophrenia or schizoaffective disorder: an 8-week, double-blind, multicenter trial. *J Clin Psychiatry* 2004;65:1624–33.