Slide 1

How to use the Cost-effectiveness ratio in everyday Health Care research

Herman J. Adèr

Vrije Universiteit

Faculty of Medicine

Department Clinical Epidemiology and Biostatistics

Amsterdam, The Netherlands

22 May, 2000

Most of what I am going to say can be found in an article:

> *Confidence intervals for the Incremental Cost-Effectiveness Ratio (Theoretical evaluation and empirical comparison of methods)*

that has now been (re)submitted to PharmacoEconomics. It describes various methods to compute confidence bounds for the ICER. To evaluate these methods, data from two cost-effectiveness studies that have been conducted at our department, were used. The first one concerns a study on the cost-effectiveness on two kinds of therapy (respectively, corticosteriod injections and physiotherapy) for the painful stiff shoulder. The second one compared epidural and intramuscular pain relief treatment in patients that had undergone elective abdominal surgery. In both cases Hindrik Vondeling has been essentially involved. In the presentation, we will concentrate on the second study. Reports on the first study can be found in the article.

The immediate cause for the article has been a study group that Hindrik and myself organized at the Biostatistics department of our University.

The present presentation is the first of two presentations: this one concentrates on the application of the Cost-Effectiveness ratio, the other presentation discusses applications of bootstrapping techniques in this field, and goes into the subject of evaluation.

I will emphasize the *methodological* issues involved: how can one properly use these techniques for proper decision making. I am *not* an economist (nor a health care expert, for that matter). My background is social science methodology and mathematics. As a defence, I can say for myself that I have much experience in statistical consultation both in the social science and medical research. Application of the ICER offers many interesting methodological problems (See Slide 2).

Slide 2

The cost-effectiveness ratio (CER) adds an extra dimension to the interpretation of research results since both clinical relevance of the effect parameters and cost considerations can be studied in connection. The ratio offers an instrument to quantify the relation between these two aspects. As such, it can be a useful tool in decision making.

I explicitly invite you to comment on my presentation, be it to ask for clarification of points you do not understand, be it to give comments inspired by your own background.

Slide 3

Outline

- What is the incremental cost-effectiveness ratio?
- The CE-plane
- A decision tree
- Confidence intervals
- The box method.
- Fieller's formula
- Conditions for Fieller's formula
- Application to the PTR study
- Summary

Slide 4

$$\hat{R} \stackrel{\text{def}}{=} \frac{m(C_1) - m(C_2)}{m(E_1) - m(E_2)} \qquad (1)$$

in which C_1, C_2 and E_1, E_2 are variables that measure costs and effects in the two (treatment) groups $\mathcal{G}_1$ and $\mathcal{G}_2$.

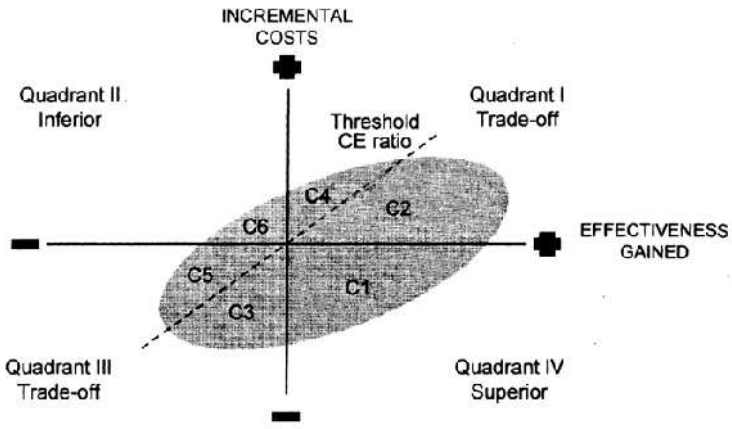
The function m in formula 1 indicates any *measure of central tendency*, usually the mean. Thus, Equation (1) assumes that both costs and effects are collected at an individual patient level, and another assumption is that this measure provides a good characterization of the costs or the effects in the group. Thus, when the costs are very skewed to the left, the mean may be a doubtful measure of central tendency. This is (one of the) reasons to work with log-transformed costs instead.
The other reason is that most techniques we will discuss have as an underlying assumption that both costs and effects have a normal (i.e. gaussian) distribution.
Another consequence is that it may be doubtfull to use *percentages* for the effects as is often appropriate in actual research, since they follow a *binomial* distribution ¡?¿ which is not usually comparible to the normal distribution.
But maybe the most important observation one may make when discussing Equation (1) that the ratio may have undesirable properties, when the means of the effects in the two groups are more or less equal. In that case, $\hat{R}$ may become very large indeed!

Slide 5

Dancing around infinity

In fact, this property is the main reason the literature on the Cost-Effectiveness ratio is so vast and is extending even now. Many of the problems can be overcome by using the bootstrapping techniques that will be the subject of the second talk. In the litterature, recently other approaches have also been proposed, like the use of *accepatability regions* and *response surfaces*. We will restrict ourselves to the ICER and its confidence bounds.

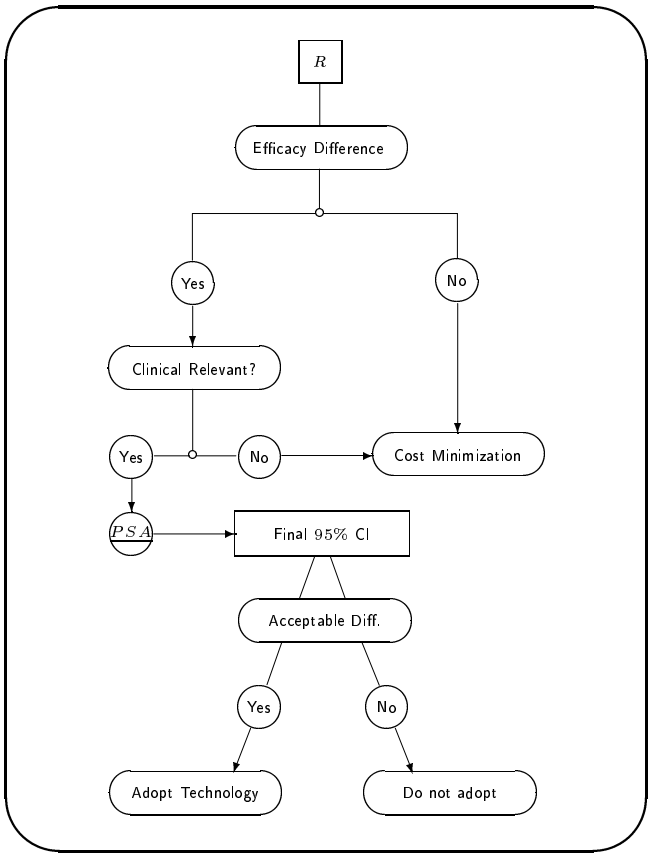
Slide 6



Slide 6 (after (Hunink, Bult, de Vries, & Weinstein, 1998)) is a good picture of the situation that concerns us here: On the x-axis effect differences between the groups to be compared are indicated. On the y-axis, costs differences between the groups are indicated, so that each point in the plane corresponds to a particular ICER. (One may imagine the value of the ICER to lay in a third dimension).

It is often argued that when $\hat{R}$ lays in the second or fourth quadrant, probabilistic considerations are inappropriate, since in the first case, costs increase while effectiveness decreases, and in the second case costs decrease while effectiveness is superior.
I don't think this is satisfactory! To explain this, note that in these cases $\hat{R}$ is only a point estimate of the value (R) in the patient population. The grey oval in Slide 6 indicates a region in which, even if the observed ratio would fall in the second or fourth quadrant, the true (population) ratio *could* lay in one of the other quadrants. This makes it necessary to perform a statistical analysis of the cost-effectiveness ratio, even if the interpretation of the observed ratio seems straightforward.

Slide 7



($\underline{PSA}$: Probabilistic sensitivity analysis.)
Sacristán, Day, Navarro, Ramos, and Hernández (1995) propose a stepwise procedure that can be represented as a decision tree (see Slide 7). The tree is traversed from top to bottom. One starts by evaluating the effect variables. Even in case there is a statistical significant difference between treatment groups, only the *clinical relevance* of the effect difference determines if cost minimization is sufficient. Note that to use the point estimate of the efficacy difference is not sufficient for a decision of clinical relevance. Instead, it should be based on the confidence interval of the difference: a wide interval may make a clinical relevant estimate questionable. Something similar is true in the case of cost minimization: it is usually interesting to calculate confidence bounds for the cost difference to get a realistic 'feel' of the variability of the costs parameter.
When effect differences are clinically relevant, we are in the situation that concerns us here: the CER becomes an interesting measure and we should decide if we can handle costs as a random variable.
$\underline{PSA}$ (*probabilistic sensitivity analysis*); includes a process that may be unsystematic or even

arbitrary: the researcher makes some plausible assumptions (or, *im*plausible assumptions if he seeks to explore the boundaries of the particular CE-plane he is interested in) and computes confidences bounds using one of formulas given below.

Slide 8

A *confidence interval* consists of two boundary points between which we have a certain specified *level of confidence* that the population parameters lies (Kleinbaum, Kupper, & Muller, 1988, page 24)

For example, a 95% procent confidence interval for the parameter θ would consist of lower and upper limits determined so that, in repeated sets of samples of the same size, 95% of all such intervals would be expected to contain the parameter θ. Kleinbaum adds: 'Care must be taken when interpreting such a confidence interval; rather, θ is a fixed (unknown) constant and the random quantities are the lower and upper limits of the confidence interval which vary from sample to sample.

A straightforward way to arrive at crude confidence intervals is the so-called *Box method* O'Brien, Drummond, Labelle, and Willan (1994): The individual confidence intervals of cost difference and effect difference are used to arrive at a 'box' in the CE-plane within which the ratio varies. But this method does not take into account the bivariability of costs and effects!

An alternative approach is described in O'Brien et al. (1994). They assume both the mean costs and mean effects to be normally distributed and derive a formula for the variance of R under that assumption.

Slide 9

Variance according to OBrien

$$\begin{aligned}\mathrm{Var}(R) \approx & \left(\frac{\frac{\sigma^2_{C_1}}{n_1}+\frac{\sigma^2_{C_2}}{n_2}}{e^2}\right) + c^2\left(\frac{\frac{\sigma^2_{E_1}}{n_1}+\frac{\sigma^2_{E_2}}{n_2}}{e^4}\right) \\ & -2c\left(\frac{\frac{\rho_1\sigma^2_{C_1}\sigma^2_{E_1}}{n_1}+\frac{\rho_2\sigma^2_{C_2}\sigma^2_{E_2}}{n_2}}{e^3}\right) \qquad (2)\end{aligned}$$

In Slide 2, $\rho_.$ indicates the correlation between costs and effects in group $\mathcal{G}_.$. The concequence is, that, according to their method, confidence bounds around $\hat{R}$ are always *symmetric*. In many cases this is by no means realistic!

Slide 10

Fieller's formula

$$CI \quad : \quad \hat{R}\left(\frac{(1-z^2_\alpha \mathrm{rcov}(c,e)) \pm z_\alpha\sqrt{D}}{1-z^2_\alpha \mathrm{cv}^2(e)}\right) \qquad (3)$$

D: discriminant expression.

z_α: significance threshold of the standard normal distribution

$\mathrm{rcov}(c,e)$ is the 'relative' covariance

$\mathrm{cv}(.)$ is the coefficient of variation.

Fieller's formulas given in Slide 10 are most commmonly used to calculate confidence bounds. The discriminant expression D is:

$$\begin{aligned}D \quad = \quad & (\mathrm{cv}^2(e)+\mathrm{cv}^2(c)-2\mathrm{rcov}(c,e)) - \\ & z^2_\alpha(\mathrm{cv}^2(e)\mathrm{cv}^2(c)-\mathrm{rcov}^2(c,e))\end{aligned}$$

The definition of rcov and cv. is given in Slide 11.

Slide 11

Relative covariance and coefficient of variation

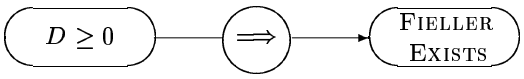
$$\mathrm{rcov}(c,e) = \frac{\mathrm{cov}(c,e)}{ce} \quad (4)$$

$$\mathrm{cv}(.) = \frac{s_.}{\mathrm{m}_1(.) - \mathrm{m}_2(.)} \quad (5)$$

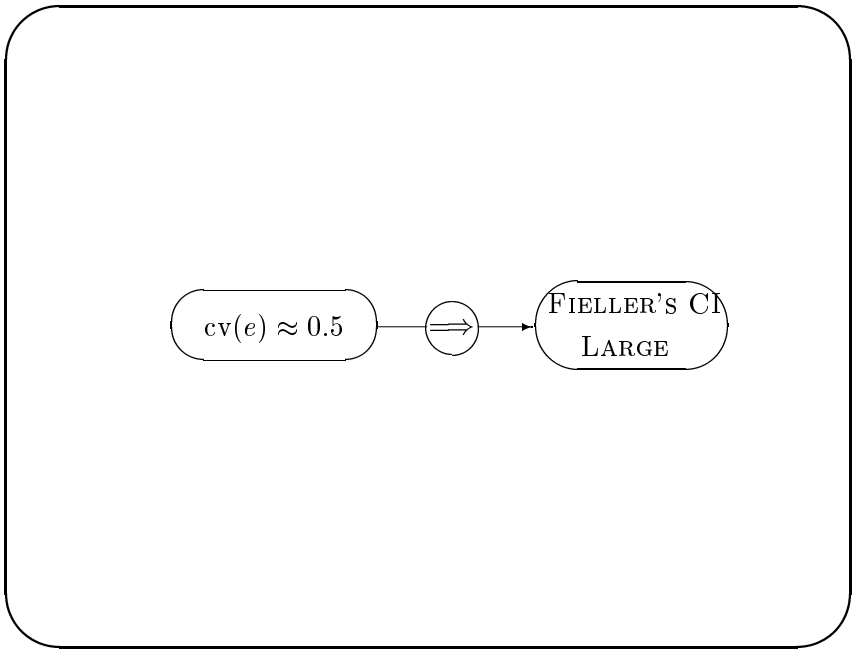
$s_.$: sample standard deviation.

It is realively easy to derive conditions under which these formulas hold.

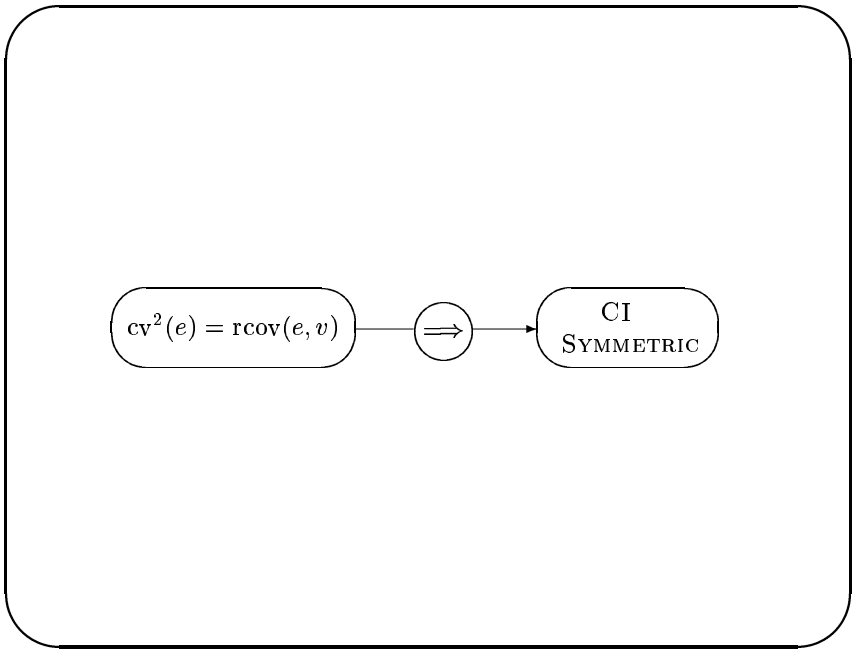
Slide 12



Slide 13



Slide 14

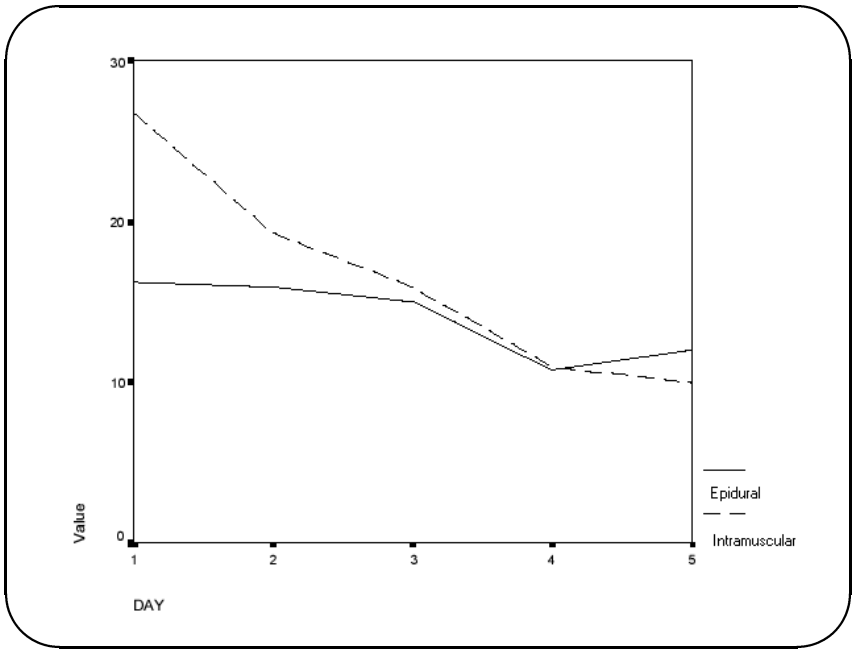


Slide 15

Parameter	Mean	Sd	Skewness	Kurtosis
Effects:				
VAS	16.28	11.44	0.81	0.32
DVas (Dayl - Day5)	9.48	18.42	0.08	0.51
SAT	7.46	1.54	1.46	1.98
Costs:				
Morphine Costs	265.39	92.7	0.57	0.35
Hospitalization Costs	3375	1207	3.07	14.31
Total Costs	3638	1222	2.97	13.48
ln(Total Costs)	8.16	.26	1.53	2.92

Pain relief treatment (PRT). We will now discuss the *Pain relief treatment* study on which the evaluations were based.
Epidural and intramuscular pain relief treatment are compared with patients that had undergone elective abdominal surgery (Ziekenfondsraad, 1995). Two hundred eighteen (218) patients were included and stratified according to place of operation: either below ($n_1 = 185$) or above the umbilicus ($n_2 = 33$). Here we use data of the first group only, since we considered the other group to be too small. In both randomization groups, treatment were continued until four days after surgery. Measurements were collected, per patient, every four hours until the fifth day after operation. After fourteen days patients were presented with a questionnaire aimed to assess satisfaction with various aspects of the treatment.
Although more data were collected, we use *morphine costs* and *hospitalization costs* as costs variables, since they were available for each patient at all measurements. As can be derived from Slide 15, the Hospitalization cost has distribution characteristics which make log-tranformation necessary if it is to be used in the CER calculations.
The primary effect variable in the study was a subjective pain measurement administered on a Visual Analogue Scale (VAS). The VAS can range from 0 to 100, where the highest score indicates most pain.
Satisfaction about the treatment as a whole (SAT) is another effect variable we will use. Table 15 gives the distribution statistics of the variables.

Slide 16



Data collected every four hours were averaged over days and daily averages were used to assess trend over time. Slide 16 gives an impression of the development over time of the VAS-score during four post-operative days. Note that in the epidural group the VAS-score immediately increases after the fourth day when treatment is discontinued. The effect of pain treatment is stable in the other treatment group.

This raises an interesting problem that is not easily overcome: we are much more interested in *trend* differences between groups than in the mean VAS differences. As it turned out, the CER can not account for trend in a straightforward way. As a proxy we use the difference between mean VAS scores on the first and the fifth day. This new variable is indicated by **DVas**.

Another approach would have been to compare VAS slopes between the epidural and intramuscular group.

The satisfaction scale (SAT) was based on a small questionnaire administered fourteen days after treatment. It is a simple instrument that gives an impression of the patient's satisfaction with the treatment. It ranges from 6 to 13, the lowest score indicating the highest satisfaction. Like the VAS, it is an measure that has been scored by the patient. There are no significant differences in satisfaction between the two groups (See Table 17). Although it may not be relevant to assess overall cost-effectiveness in this trial, it is included since it nicely illustrates the behaviour of the cost-effectivess ratio in is the neighborhood of zero.

Slide 17

Parameter	Group	Mean(Sd.)	No.	T-test	Prob.	Cl (diff.)
PPR trial						
DVas	Epi	1.56 (16.21)	72			
	Intra	15.95 (17.64)	88	-5.32	< 0.01	(-19.73,-0.05)
SAT	Epi	7.65 (1.66)	65			
	Intra	7.25 (1.37)	80	1.57	0.12	(-.10, .89)
Total Costs	Epi	3844 (1489)	70			
	Intra	3472 (932)	87	1.91	.07	(-32.83,776.14)
ln(Total Costs)	Epi	8.20 (.30)	70			
	Intra	8.12 (.23)	87	1.89	.06	(-.004,.1626)

Slide 17 gives t-tests for the differences in means between groups for the PPR trial. DVas differs significantly between groups while Total Costs and SAT do not.
For SAT the mean difference is 0.40 and the 95% confidence interval is $(-.10, 0.90)$. That is, zero is included.

Slide 18

Study	Groups	Parameters	D	$\mathrm{rcov}(c,e)$	$\mathrm{cv}(e)$	$\mathrm{cv}^2(e)$
PRT	Epi vs Intra	LnTC/DVas	0.234	0.015	0.188	0.035
		LnTC/ SAT	0.332	-0.078	0.645	0.416

Slide 18 shows that Fieller's formula will exist in all cases, since Fieller's determinant is never less than or equal to zero. In case of the SAT, Fieller's Cl will be far from symmetric, since $(\mathrm{cv}^2(e) - \mathrm{rcov}(c,e) = 0.416 + 0.078 >> 0)$, with probable bad coverage of O'Briens interval, since it is symmetric by definition.
Since in neither case $|\mathrm{cv}(e)| \approx 0.5$, no degenerate confidence intervals are to be expected.
Finally, we give the resulting confidence intervals in Slide 19.

Slide 19

				Confidence interval		
Study	Groups	Costs	Effects	lwb.	CER	upb.
Fieller						
PRT	Epi vs Intra	lnTC	DVas	-0.011	-0.006	0.000
			SAT	-0.040	0.165	0.491
O'Brien						
PRT	Epi vs Intra	lnTC	DVas	-0.012	-0.006	0.000
			SAT	-0.167	0.165	0.498

Compared to the Fieller intervals, O'Brian's intervals are wider.

PTR. For DVas we find a unbiased estimate of $-.006$, and a confidence interval of $(-.011, 0.0)$. Transforming back to Dutch guilders, this would mean that comparing Epidural versus Intramuscular treatment, $Dfl.\ 0.99(.98, 1.00)$ has to be paid per one point decrease on the DVas scale between groups. Again, the enormous cost differences suggested by the crude calculations should be put into this perspective.
For the SAT, the point estimate would be 0.158, while the confidence interval would be (-.05, 0.397), based on Fieller's formula (the ABC interval is shifted). In this case, zero is included in the confidence interval and thus, from a cost-effectiveness point of view there is no difference between treatments in terms of satisfaction.

Slide 20

Summary, or: What have we learned?

- Calculation of the ICER may be worthwhile even if no clinical relevant effect difference can be demonstrated.
- Application of the ICER is not straightforward: there are several restrictions.
- The ratio is based on the *means* of costs and effects: trend cannot be analyzed.
- Between the requirements to furfill, normality of costs and effects and an effect coefficient of variation below 0.5 are the most important.

References

Hunink, M. G., Bult, J. R., de Vries, J., & Weinstein, M. C. (1998). Uncertainty in Decision Models Analyzing Cost-Effectiveness: The Joint Distribution of Incremental Costs and Effectiveness Evaluated with a Nonparametric Bootstrap Method. *Medical Decision Making*, *18*(3), 337–346.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariable Methods.* Boston: PWS-KENT Publishing Company.

O'Brien, B. J., Drummond, M. F., Labelle, R. J., & Willan, A. (1994). In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care*, *32*, 150–163.

Sacristán, J. A., Day, S. J., Navarro, O., Ramos, J., & Hernández, J. M. (1995). Use of confidence intervals and sample size calculations in health economic studies. *The Annals of Pharmacotherapy*, *29*, 719–725.

Ziekenfondsraad. (1995). *Evaluatie post-operatieve Pijnbestrijding* (OG92/009). Vrije Universiteit, Amsterdam: Ziekenfondsraad. (Project-groep post-operatieve Pijnbestrijding)