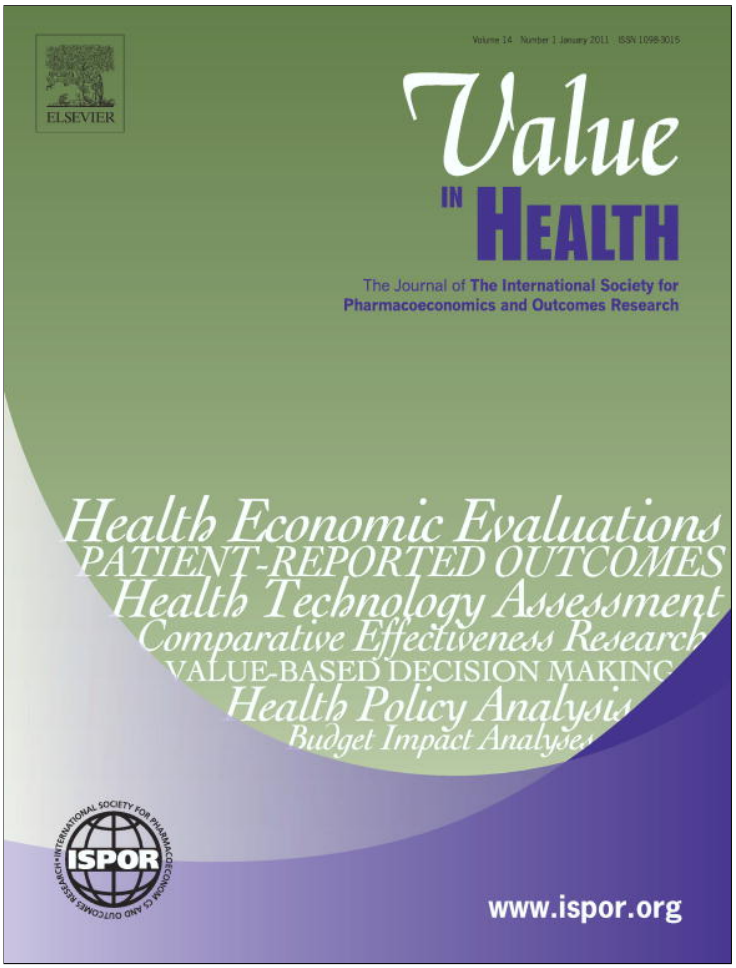

VALUE IN HEALTH 14 (2011) 3-14

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jval

Economic Evaluation

Can we account for selection bias? A comparison between bare metal and drug-eluting stents

Annemieke De Ridder, PhD[a,b,*], *Diana De Graeve, PhD*[b]

[a] *Hogeschool Universiteit Brussel, Brussel, Belgium*
[b] *University of Antwerp, Antwerpen, Belgium*

Keywords:
Coronary heart disease
Costs
Drug-eluting stent
Selection bias

A B S T R A C T

Objective: In this article we investigate the possibility to account for selection bias in observational data by using econometric techniques.

Methods: One-year costs of 15,237 patients who received a drug-eluting stent (DES) or a bare metal stent (BMS) in Belgium in 2004 were compared. The treatment choice between DES and BMS could be influenced by patient characteristics; therefore, cost estimates could be biased by overt and/or hidden selection bias. Overt bias was addressed by regression adjustment and propensity score matching. Hidden selection bias was dealt with by using an instrumental variable (IV) approach.

Results: Due to the higher purchase price DES patients incur higher (unadjusted) costs in the short-term; these costs are, however, compensated in the long-term due to less in-stent restenosis and hospitalizations. Analyses indicated that, for the diabetic population, the null hypothesis of similar average 1-year costs of patients receiving a BMS or DES cannot be rejected. For the non-diabetic patients a significant cost difference between BMS and DES patients was found. It cannot be ruled out that the treatment-effect model does not correct for all observable or unobservable characteristics and that the estimated treatment effect is biased, possibly due to weak instruments.

Conclusion: It is interesting and necessary to explore the use of econometric tools in cost and cost effectiveness analysis to investigate the effect of a technology in everyday practice and to take into account patient and disease characteristics and uncertainty. Further research is however necessary to investigate how we can fully correct for selection bias when using observational data.

Introduction

Coronary heart disease (CHD) is a major cause of death and morbidity in developed countries. It is caused by narrowing of the coronary arteries and is treated by coronary artery bypass grafting (CABG) or percutaneous coronary interventions (PCI), which include balloon angioplasty and stenting. Over the past decade the PCI technique to treat CHD has developed rapidly. Thanks to technical and pharmacological innovations, the effectiveness and safety of coronary stent devices have gradually improved. The latest generation of coronary stents in-

Funding: The authors have no financial relationships to disclose.
* *Address correspondence to:* Dr. Annemieke De Ridder, University of Antwerp, IOIW, Venusstraat 35, 2000 Antwerpen, Belgium.
E-mail: annemieke.deridder@ua.ac.be.
1098-3015/$36.00 – see front matter
doi:10.1016/j.jval.2010.10.014

cludes the so-called drug-eluting stents, which are stents coated with pharmaceutical agents that suppress neointimal hyperplasia [1,2]. Evidence (randomized controlled trials [RCTs] and registries) shows that the use of drug-eluting stents (DES) does not affect (cardiac) mortality or the occurrence of myocardial infarction (MI) but that, compared to classic bare metal stents (BMS), DES have proven to be successful for the prevention of restenosis after PCI. This could result in a major cost saving for the health care payer [3–6].

In Belgium, the use of BMS is still the standard procedure. This is said to be mainly due to the higher device price of DES compared to BMS. Current use of DES in Belgium is mainly driven by the one and only approved indication of reimbursement, "patients with treated diabetes" (i.e., patients who are medically treated with insulin or oral hypoglycemic agents). DES are also used in non-diabetics, but in these cases hospitals are only reimbursed at the level of BMS and are not allowed to charge the patient an out-of-pocket payment for the device.

The cost effectiveness of DES in Belgium was investigated in the Health Technology Assessment (HTA): drug-eluting stents in Belgium [7], results are published in Neyt et al. [8]. The cost effectiveness analysis (CEA) indicated that the DES are not cost effective compared to BMS. The CEA was performed using RCT data in a model, which is standard practice in HTA.

Evaluation bodies recommend using studies with a high certainty of results, especially RCTs). Clinical-trial data are then supplemented by a great deal of economic modeling. The requirement to use RCTs is justified for the demonstration of causality: randomization of patients ensures that differences in effect can solely be ascribed to a single determining factor, e.g., the different treatments. When study groups are not randomized, there may be systematic differences between the groups regarding known factors as well as unknown factors and this may bias the comparison. However, the use of RCT's invokes critique as well [9–12].

First, there may be problems with external validity. Most trials have stringent inclusion and exclusion criteria, restricting participation to a homogeneous, highly selective group of patients and, therefore, may not be widely generalizable [9-13]. In the case of PCI for example, most trials pertain to patients with single, previously untreated coronary lesions. In reality, stents are implanted in more complex scenarios. In models the relative risk-benefit ratio of DES versus BMS from (meta-analyses of) trials is usually extrapolated to such complex cases. Hard evidence that this is appropriate is lacking.

Second, in RCTs, there is a strict follow-up of patients. The outcomes measured may not reflect clinical practice [11–13]. In PCI trials, for example, there is an angiographic follow-up of patients 6 to 9 months after the index procedure to assess in-stent restenosis. This may lead to revascularizations that are not clinically driven and increases the absolute difference in the rates of clinical restenosis between BMS and DES. In order to cope with this, it is assumed that the relative benefit (e.g., the relative risk reduction) is not affected; but again there is no strong evidence that this is appropriate. The third critique on RCTs is an issue that appears to be rarely addressed. When populating decision models for cost-effectiveness analysis using evidence synthesis methods, the effect estimate (e.g., the relative risk) is usually constant across different "baseline risks." Usually an overall relative risk is applied in the decision model to the baseline rate in the specific population [14].

Fourth, there is the problem of publication bias: negative trial results are less often published than positive ones [12,13,15]. Further issues are more specific for clinical trials with medical devices as opposed to those with drugs. With devices, the clinical outcome can depend on the skill and experience of the surgeon and the setting in which he operates. Practitioners participating in RCTs are generally "enthusiastic volunteers" with strong motivation, and exceptional skills and experience, leading to improved outcomes. Learning curves should also be taken into account. In addition, devices frequently undergo product modification, with possible impact on effectiveness [16,17].

Given the disadvantages of using RCTs, we considered an observational study to compare costs of the two patient groups using regression analysis. Of course direct comparison of both patient groups is problematic because estimation of treatment effects can be prone to selection bias when the assignment to treatments is associated with the potential outcomes of treatment. The purpose of this article is to investigate whether it is possible to account for selection bias, i.e., whether we can mimic the random assignment of experimental design by using econometric techniques for this comparison.

Methods

Data

In order to investigate the cost of a PCI implant in Belgium, we compared the 1-year direct medical costs of patients who received a DES or BMS in Belgium in 2004. We will discuss costs from the viewpoint of the health care payer, which means that we take health insurance system (HIS) reimbursements as well as patient copayments into account (supplements and non-covered items are not taken into account). For this study, data of the Belgian Working Group of Interventional Cardiology (BWGIC) with clinical information of the PCI were linked with patient reimbursement data of the different sickness funds obtained from the Intermutualistisch Agentschap (IMA). The first PCI of the patient in 2004 is called the index PCI. Cost data from 1 year previous until 1 year past the index PCI date were collected. Those data included all costs generated by the patient and covered by the HIS; including the costs of the index-PCI and the hospitalizations, ambulatory follow-up costs, costs of complications or re-intervention, and also all other non-PCI-related costs of other illnesses, preventive activities, etc. Additionally, vital statistics were collected until 1 year after the index PCI. This way a total of 15,237 patients were included in the analysis, and the database contains all information on patients who underwent at least one PCI with stenting in 2004, who received only one type of stent during the index-PCI (BMS or DES), who had follow-up data on the consumption of pharmaceuticals, and who did not receive a stent in both 2003 and 2004 during one and the same hospitalization.

Stent costs in Belgium are reimbursed on a lump sum basis: whatever the number of stents implanted, a fixed amount per hospitalization is reimbursed to cover device and material costs. In Belgium, only diabetic patients receive a higher reimbursement for DES. Implantation of a DES in a patient with diabetes has a lump sum of €1000 higher in comparison to the lump sum for BMS. A cardiologist can also decide to implant a DES in a non-diabetic patient, but in this case the reimbursement is limited to the lump sum for BMS and the hospitals have to bear the additional device cost themselves. Even though we are performing a cost study from the health-care payer's perspective, the extra expenditures for the intervention under consideration should be included to be able to make correct comparisons between BMS and DES patients. Otherwise we miss the extra cost of DES in our analyses. Therefore, we have added an extra lump sum of €1000 if a DES was used for the initial PCI or repeat PCI in non-diabetic patients (no matter how many stents are implanted).

When calculating and comparing the costs of BMS and DES patients we have always made a distinction between diabetic and non-diabetic patients. Given the importance of diabetes for PCI intervention and the selection of stent type in Belgium, we believe it would be inappropriate to analyze costs without taking the presence of diabetes into account.

Analysis

We have calculated total direct medical costs of the patients from the day of the PCI implant until 1 year after the procedure, for diabetic and non-diabetic patients separately. We have also calculated these costs after 1 and 3 months of follow-up. That way we can test the hypothesis that although DES is more expensive in the short-term, their extra cost is compensated in the long-term.

When we compare costs between the two treatments it is very important that the cost differences that are found reveal the causal effect of treatment. Because we work with observational data, the treatment choice between DES and BMS could be influenced by certain patient characteristics. This is called selection bias. There are two types of selection bias: overt selection bias and hidden selection bias. In our case, overt selection bias occurs when observed patient characteristics influencing costs differ for patients who receive a BMS or a DES; hidden bias occurs when costs are influenced by unobserved characteristics. When selection bias is present it is not correct to simply compare the costs of the two treatment groups because the estimates of the effects of the alternative treatments will be statistically biased. Several techniques have been developed to correct for overt and hidden bias when working with observational data. To address overt bias matching, traditional regression methods, propensity score matching, or a combination of those are used. Hidden bias is usually addressed by instrumental variable analysis [18–20]. Below we will look into these two types of bias more closely.

Overt bias

Overt bias is usually addressed by regression adjustment or propensity score matching or a combination of both methods. For regression adjustment typically an ordinary least squares (OLS) model is estimated where the outcome variable is regressed onto the treatment variable and all important observed patient characteristics. The coefficient of the treatment variable then represents the corrected average treatment effect. Another possibility is to perform propensity score matching. For this method a propensity score p(X) is estimated for each subject in the data set. This p(X) is the conditional probability of being assigned to treatment 1 versus treatment 2 given a vector of observed characteristics. P(X) is usually obtained from logistic or probit regression. When the p(X) is estimated for each subject we need to find two subjects with the same p(X); we can think of these subjects as if they were randomly assigned to each treatment group because they have the same probability of being in either group, given their characteristics. The bias of the confounding covariates is then reduced when the comparison of outcomes is performed using subjects of both groups who were as similar as possible [21–23]. To calculate the average treatment effect it is, however, not sufficient to simply estimate p(X). The probability of finding two subjects with exactly the same p(X) is very small, so in order to calculate the treatment effect we need a good method to match the subjects according to their p(X). Several methods exist; the most commonly used are nearest neighbor matching, radius matching, kernel matching, and stratification. Although not all of these methods provide the same results, we ought to look at them more closely. For stratified matching the subjects are divided into several strata such that in each stratum subjects on treatments 1 and 2 have on average the same p(X). Outcomes of these subjects are then compared to calculate the average treatment effect. An important disadvantage of this method is that you lose those observations for which a stratum contains only subjects on treatment 1 or 2. This can be overcome by working with nearest neighbor matching in which subjects with the closest p(X) are matched and compared; a weakness here is that some matches could be very poor. With radius matching, each subject on treatment 1 is matched with a subject on treatment 2 whose p(X) falls in a predefined neighborhood (radius) of the p(X) of the subject in group 1. Here we have to make a trade-off between a small radius for which it is possible that some treated units are not matched and the fact that the smaller the radius, the higher the quality of the matches. Finally, we can work with kernel matching for which all subjects from group 1 are matched with a weighted average of all subjects from group 2, with weights inversely proportional to the distance between the p(X) of the subjects from both groups. It is found that none of these methods is a priori superior and that the joint consideration of all results is a good way to evaluate the robustness of the results. As stated, we can also use a combination of propensity score matching and regression adjustment to adjust our observed data. The p(X) could be used as a predictor in a regression model along with covariates that could not be balanced [22,23].

After discussing these two methods to correct observational data for overt selection bias we can wonder whether one of these methods should be preferred to the other. This was tested by Drake et al. [24]. They performed a comparison of the propensity score matching method and prognostic models in estimating treatment effects from observational studies by performing several simulations. They found that

Table 1 – Comparison of baseline characteristics of patients with DES and BMS.

Patient characteristic	Diabetics				Non-diabetics			
	Number	BMS	DES	*P* value	Number	BMS	DES	*P* value
Male gender	2795	66.8	60.4	0.00	12,442	74.8	72.0	0.02
Mean age	2795	68.6	66.6	<0.0001	12,442	65.2	63.7	<0.0001
Flanders	2795	67.5	53.4	0.00	12,442	63.4	44.6	0.00
Walloon region		25.4	36.8			29.8	50.3	
Brussels/other		7.10	9.8			6.8	5.0	
Alive	2795	91.6	95.2	0.003	12,442	96.0	96.7	0.717
Death in Q1		4.8	2.3			2.2	1.7	
Death in Q2		0.7	0.8			0.5	0.5	
Death in Q3		1.7	1.1			0.7	0.5	
Death in Q4		1.3	0.6			0.6	0.5	
Renal dysfunction	2695	7.6	4.3	0.00	11,656	2.1	2.7	0.16
1-vessel disease	2795	41.6	38.8	0.12	12,442	46.3	45.8	0.82
2-vessel disease		27.9	32.3			31.7	31.5	
3-vessel disease		30.5	28.9			22.0	22.7	
Peripheral vascular disease	2654	18.7	15.0	0.04	11,678	10.4	11.2	0.38
AMI or failed thrombolysis	2745	14.5	8.6	<0.0001	12,114	16.3	7.6	<0.0001
Stable CHD or asymptomatic patients	2790	36.1	47.0	<0.0001	12,411	37.9	42.1	0.00
Glycoprotein IIb/IIIa inhibitors	2795	26.1	20.1	0.002	12,442	23.9	15.1	0.00
Direct stenting	2795	36.3	37.2	0.675	12,442	42.7	39.0	0.006
Left main	2795	1.5	1.6	0.78	12,442	1.3	3.6	<0.0001
Proximal LAD	2795	14.0	15.5	0.36	12,442	16.2	24.3	<0.0001
Prior PCI	2795	19.0	24.4	0.01	12,442	17.3	32.4	<0.0001
Small vessel	2795	14.4	15.2	0.62	12,442	14.7	15.0	0.77
Long lesion	2795	7.7	3.9	0.00	12,442	6.8	6.7	0.86
Single room	2649	6.0	6.9	0.69	11,960	6.4	12.3	<0.0001
2-person room		16.5	17.2			15.0	20.5	
Common room		77.5	76.0			78.6	67.2	
No. of stents	2795	1.31	1.17	0.00	12,442	1.29	1.22	0.00

AMI, acute myocardial infarction; BMS, bare metal stent; CHD, coronary heart disease; DES, drug-eluting stent; LAD, left anterior descending; PCI, percutaneous coronary intervention.

when omitting a covariate (e.g., in the case of hidden bias) propensity score matching has comparable biases to those of prognostic modeling. Propensity scoring, however, seems preferable when considering model misspecification, particularly so because incorrect propensities score models have smaller biases. Polsky and Basu [18] have found similar results; they state that OLS is more efficient than propensity score matching if the model is correctly specified. If the model is incorrect, OLS may fail to remove or even increase overt bias whereas propensity scoring is fairly consistent in reducing overt bias.

It is obvious that we will have to take the possibility of overt selection bias into account when comparing the 1-year costs of BMS and DES patients. Patients who receive a BMS could be very different from patients who receive a DES. When we look at Table 1 it is obvious that there are many significant differences between both patient groups; e.g., patients who receive a DES are significantly younger, they suffer significantly less from renal dysfunction, they have a significantly less chance of being admitted with acute myocardial infarction and thrombolysis, etc. That is why it is very important to correct for these patient characteristics when comparing the costs of both groups of patients. We will first use traditional regression techniques (OLS) to correct for possibly confounding covariates. After that we will compare the results with those of propensity score matching.

Hidden bias

The second potential bias that can occur in working with observational data is hidden selection bias. This type of bias is more difficult to handle. Hidden selection bias means that unobserved characteristics are correlated with both the initial treatment choice between BMS and DES and the observed costs of both stents. Traditionally, hidden selection bias is dealt with by using an instrumental variable (IV) approach. In a regular IV approach one needs to identify instrumental variables for which two assumptions should hold: the instruments should be correlated with the treatment choice variable and the instruments should be independent of the outcomes variable. When these two assumptions are met, the instruments can effectively randomize subjects across the treatment arms. The analysis then proceeds in two steps: first the probability to receive a certain treatment (e.g., DES) is estimated by probit regression of the treatment dummy onto the chosen instruments. Then the predicted probability is included in the cost equation instead of the treatment dummy. That way the endogeneity is removed. This IV technique is, however, mainly used for continuous endogenous regressors; interpretation becomes difficult for endogenous dummies. When the endogenous regressor is a dummy variable, a treatment effect model is usually estimated. This can be done by using a Heckman two-step estimator, which also proceeds in two steps: first, a probit regression is used to estimate the

Table 2 – Average costs before and after index PCI in Euros (€).

	No.	Diabetics			No.	Non-diabetics		
		DES	BMS	Difference (P value)		DES	BMS	Difference (P value)
Total average medical cost 1 year before PCI	2795	8478.50	8531.90	−53.40 (0.992)	12,442	5889.3	5195.90	693.40 (0.0009)
Total average medical cost after PCI	2795	6632.74	6126.39	506.35 (0.0005)	12,442	6276.19	5896.90	379.29 (0.0001)
• 1 month		8246.10	7845.75	400.35 (0.0656)		7279.57	7056.06	223.51 (0.0841)
• 3 months		17,485.97	18,273.05	−787.08 (0.277)		13,221.52	13,946.35	−724.83 (0.040)
• 1 year								

BMS, bare metal stent; DES, drug-eluting stent; PCI, percutaneous coronary intervention.

probability to receive treatment (DES), the regressors contain valid instruments. Then the predicted probability is used to calculate an inverse Mills ratio which is included in the cost regression together with the other covariates and the treatment dummy. This cost regression can be estimated by OLS. The inverse Mills ratio (or risk variable) accounts directly for the part of the error term that is correlated with the treatment choice variable. When the risk variable has a significant impact on the outcome variable, this means that selection bias is present and that we should work with the results of the treatment effect mode [18,19,25]. A test of the significance of the coefficient (rho) of the risk variable is very important. When the risk variable is not a significant covariate this means that simple OLS estimates or propensity scoring can be used because only overt bias is present [20]. The treatment effect model can also be estimated by maximum likelihood estimation that produces more efficient estimates. A disadvantage of these methods is that it is often very difficult to find suitable instruments. The estimation of the models generally depends on arbitrary identifying restrictions for the selection equation [18,19].

The data set we work with is very elaborate; still it is possible that some important patient characteristics are not observed but are very important in determining whether a patient receives a DES or BMS. That is why it is essential here to test the possibility of hidden selection bias. When hidden bias is present we will analyze the results by estimating a treatment effect model.

Results

We analyzed the full 1-year cost since the index PCI at "day 0" from the viewpoint of the health care payer. This means that we take into account HIS reimbursements and patient co-payments (supplements and non-covered items are not considered). It is not possible to make a clear distinction between PCI-related costs and other costs. Table 2 shows the reimbursements and co-payments of these 1-year follow-up costs for the diabetic and non-diabetic patients, respectively, subdivided for patients who received a BMS or a DES. The 1-year costs before the PCI implant were also added to allow for comparison. The diabetics costs of the previous year are quite similar for patients with DES or BMS. For the non-diabetics this is not the case; costs are significantly different between BMS and DES patients. DES patients incur significantly higher costs in the year prior to their PCI implant. Costs of the previous year could be considered as a proxy for the health condition of the patient. Therefore this could be an indication that non-diabetics who receive a BMS were in better health during the year prior to the intervention than those who receive a DES.

In Figure 1 the evolution of costs over the year of follow-up is presented. It is obvious that, as expected, due to the higher device price, DES patients incur higher costs in the short-term. In the long-term, the DES group has lower costs on average. A possible explanation for this could be that it is due to fewer restenoses and hospitalizations in these patients. After 1 month the DES patients have significantly higher costs compared to BMS patients; after 3 months the DES patients are still more expensive, but the difference in costs is not significant anymore. After 1 year the cost difference reverses, BMS becomes the most expensive patient group (Table 2).

When we review the 1-year follow-up costs, it is obvious that the patients in our sample are very expensive. On average, in Belgium, HIS reimbursements in 2004 amounted to €1607 per individual [26] and our patients are 5 to 10 times more expensive. Total health care payer costs (reimbursements + copayments) amounted to €18,273 and €17,486 for BMS and DES, respectively, for the diabetic patients. For the non-diabetic patients the amounts were €13,946 and €13,222,

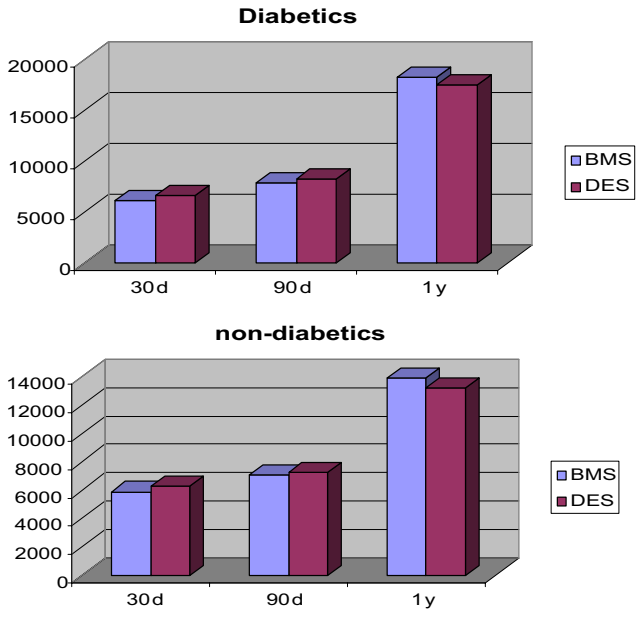

Fig. 1 – Evolution of costs (1 month, 3 months, 1 year).

respectively. It is striking that the non-diabetic patients who receive a DES have significantly lower costs than those who receive a BMS, and that this difference is opposite to our findings for the costs prior to the PCI implant. For the diabetic patients there are no significant cost differences between BMS and DES patients.

As stated in the Methods section, direct comparison of the costs of both stent types is inappropriate because our patients are not randomly allocated to the two treatments and the characteristics of the patients receiving DES or BMS are not similar. Taking into account the observed patient characteristics available in the database could partially correct for this observational bias (overt selection bias) but does not solve the problem of non-random allocation (hidden selection bias). We will first try to eliminate the overt selection bias by performing regression analysis and propensity score matching on the data. After that we will test for the existence of hidden selection bias and adjust our results if hidden bias is present.

The dependent variable in our OLS model is the 1-year direct medical cost of the patients. As independent variables, multiple patient characteristics are taken into account. First the PCI type (dummy DES=1) is included, next some demographic characteristics are incorporated: sex and age of the patient, the region where the patient lives and whether the patient survives the follow-up period. We further take into account a number of disease severity characteristics of the patients; whether the patient suffered from an acute infarction or failed thrombolysis when admitted, whether he or she suffers from stable or asymptomatic coronary artery disease, the number of diseased vessels of the patient, whether he or she suffers from renal dysfunction, from peripheral vascular disease, whether thrombocyte aggregation blockers are used during the hospitalization, and whether the lesion is left main or proximal left anterior descending; we further correct for the fact whether direct stenting is applied and for the total costs (reimbursements + copayments) of 1 year before the hospitalization. The latter variable is introduced as a proxy for the health status of the patient (other than vascular) for which we do not have other indicators. In previous analyses, other explaining covariates were included, such as the hospital in which the stent was placed, the experience of the hospital, the experience of the cardiologist, etc. These did not add to the power of the model or the inclusion of the variables invalidated the results of the treatment effect model; they were not included in the final analysis.

Several models were fit to identify the appropriate model in terms of statistical assumptions. Concerning model specification the OLS model performed best. Using the Ramsey RESET test, specification could not be rejected at the 0.01 level. Because the costs were skewed to the right, we also estimated the model with the logarithm of costs but that gave a worse specification.

The results of the OLS regressions for the diabetic and the non-diabetic sub-samples are summarized in Table 3.

A second possibility to eliminate overt selection bias is to perform propensity score matching. For this method subjects from both treatment groups are matched as well as possible according to their propensity score. The costs of the matched subjects are then compared rather than comparing the average costs of all subjects in both groups. The propensity score was estimated by performing a probit regression. As stated in the methods section several methods exist to match the subjects. We have used the most common methods, including nearest-neighbor matching, radius matching, kernel matching, and stratification. We wanted to use the same covariates as for the regression approach in order to be able to compare the results. We could not keep all covariates in the analysis because not all of them satisfied the balancing property. That is why the dummy variables representing whether the patient survives the follow-up period were dropped for the non-diabetics. We have restricted the estimations by using the common support option. This restriction implies that the test of the balancing property is performed only on the observations whose p(x) belongs to the intersection of the supports of the p(x) of groups with BMS and DES. Imposing this restriction may reduce the number of observations but improves the quality of the matches [23]. The results can be found in Table 4.

There is a possibility that our cost estimates are not only biased by observed patient characteristics but also by unobserved covariates. To investigate the possibility of hidden selection bias treatment, effect models were estimated (using maximum likelihood) for both the diabetic and non-diabetic patients. When estimating such a model we first need to find good instruments to predict the probability that the patient will receive a BMS or a DES. The instruments that were used take into consideration whether the patient has had a previous PCI; whether the patient has small vessels or long lesions; whether the patient stays in a single, double, or common room; and the number of stents the patient needs. In order to have valid instruments these variables must meet two assumptions. First, the instruments need to have enough explanatory power to explain the treatment choice; which means that the instruments should be significantly correlated with the treatment dummy. This is tested by calculating the Shea partial R^2 and comparing the F-value with the critical values reported by Stock and Yogo [27]. Second, the instruments should not be correlated with the error term in the cost equation. This condition is tested by calculating the Hansen J statistic. For the diabetics, a Shea partial R^2 of 0.0024 (P = 0.000) is found. When we look at the critical values of Stock and Yogo it can be concluded that the F-value is just above a 30% relative bias. Our only concern is that by the second Stock and Yogo critical value it is found that F is just below a 25% distortion of the Wald test size. Because the two former tests were positive, we conclude here that our instruments are strong enough to proceed. For the diabetics a Hansen J of 4.402 (P = 0.354) is found, which means that the second condition is also met. Also, for the non-diabetics, the assumptions were fulfilled: a Shea partial R^2 of 0.0262 (P = 0.000) is found and the Stock and Yogo critical values are also convincing; we find an F-value above a 5% relative bias and a 10% distortion of Wald test size. The Hansen J statistic amounts to 10.622 (P = 0.0594). Now that the validity of the instruments is confirmed, the treatment effects model can be estimated. Results can be found in Table 5.

We first look at the coefficient for the risk variable (rho) and at the likelihood ratio test with null hypothesis rho = 0 (independent equations). For the diabetics it is found that rho is not significantly different from zero. This means that, for the dia-

Table 3 – OLS results.

Dependent variable: 1-year medical costs (in €)	OLS	
Independent variables	Diabetics (n=2564)	Non-diabetics (n=11,200)
Choice of PCI (DES)	−333.70	−460.48
Demographic characteristics		
Age	35.22	33.60†
Male gender	−1220.78†	−795.65†
Region Flanders vs. Walloon region	−396.28	906.73†
Brussels+abroad vs. Walloon region	1232.95	1986.53†
Death in quarter 1 vs. alive	−11,447.52‡	−4423.96†
Death in quarter 2 vs. alive	6602.50†	4992.59†
Death in quarter 3 vs. alive	8864.09‡	11,262.24†
Death in quarter 4 vs. alive	8558.84‡	16,351.45†
Disease severity		
Acute infarct/thrombolysis	5400.10‡	5037.45†
Stable/asymptomatic coronary artery disease	−1338.98†	−1348.50†
	173.35	1091.26†
Number of diseased vessels		
2 vs 1	2062.41‡	2285.28†
3 vs 1	8764.96‡	6478.19†
Renal dysfunction	−39.59	1523.19†
Peripheral vascular disease	2221.88‡	1032.28†
Glycoprotein IIb/IIIa inhibitors	−1991.04	1590.26
Left main	1151.73	1621.52†
Proximal LAD	−597.48	
Number of stents		
Other	−1053.39*	−421.00
Direct stenting	0.65‡	0.61†
Costs of previous year	10,394.41‡	6488.00†
Constant term	0.32	0.20
R^2	58.04‡	143.46†
F		

BMS, bare metal stent; DES, drug-eluting stent; LAD, left anterior descending; OLS, ordinary least squares; PCI, percutaneous coronary intervention.

* Significant at the 10% level.

† Significant at the 5% level.

‡ Significant at the 1% level.

betic population, the cost data are not biased by unobserved characteristics and that we can look at the results of the simple OLS regression or propensity score matching to find the average treatment effect of DES. For the non-diabetic patients rho is significantly different from zero; cost estimates are biased by hidden selection bias and we will look at the treatment effects model for the average treatment effect of DES.

We first discuss the results of the costs of the diabetics. Because no hidden selection bias was found, we focused on the results of the OLS model and propensity score matching. The most important result of both analyses was that the null hypothesis of similar average 1-year costs of patients receiving a BMS or DES cannot be rejected. The average treatment effects found by both methods are quite divergent. OLS found

Table 4 – Results of propensity score matching.

	Diabetics				Non-diabetics			
	Number	ATE	t-stat	*P* value	Number	ATE	t-stat	*P* value
Nearest neighbor	n BMS=434 n DES=2015	−1428.67	−1.208	0.113	n BMS=1165 n DES=1361	247.72	0.534	0.25<P<0.40
Radius	n BMS=547 n DES=2015	−1093.33	−1.291	0.098	n BMS=9822 n DES=1361	−472.16	−1.454	0.05<P<0.10
Kernel	n BMS=547 n DES=2015	−989.29	−1.043	0.149	n BMS=9822 n DES=1361	−395.63	−1.327	0.05<P<0.10
Stratification	n BMS=547 n DES=2015	−938.71	−1.044	0.149	n BMS=9823 n DES=1360	−287.31	−0.838	0.15<P<0.25

ATE, average treatment effect, difference between average 1-year medical cost DES and average 1-year medical cost BMS; BMS, bare metal stent; DES, drug-eluting stent.

Table 5 – Results of the treatment effect model.

Independent variable: 1-year medical costs (in €)	Treatment effect model	
Dependent variables	Diabetics (n=2430)	Non-diabetics (n=10,721)
Choice of PCI (DES)	18.32	−346.81†
Demographic characteristics		
Age	27.80	27.29†
Male gender	−1159.96	−849.26†
Region Flanders vs. Walloon region	−330.82	626.50*
Brussels + abroad vs. Walloon region	1396.56	1933.43†
Death in quarter 1 vs. alive	−11,510.31†	−3806.91†
Death in quarter 2 vs. alive	5908.25	4746.67†
Death in quarter 3 vs. alive	9212.10†	11,173.07†
Death in quarter 4 vs. alive	9004.74†	15,780.73†
Disease severity		
Acute infarct/thrombolysis	5405.55	4873.09†
Stable/asymptomatic coronary artery disease	−1326.72*	−1371.13†
Number of diseased vessels		
2 vs. 1	−4.90	1142.01†
3 vs. 1	2086.20†	2382.15†
Renal dysfunction	8374.14†	6811.40†
Peripheral vascular disease	−160.01	1508.24†
Glycoprotein IIb/IIIa inhibitors	2248.88†	925.33†
Left main	−1105.68	2284.83*
Proximal LAD	1038.94	1773.14†
No. of stents	−650.07	
Other		
Direct stenting	−1072.71	−508.41*
Costs of previous year	0.66†	0.61†
Constant term	10,710.55*	7433.09†
Risk variable	0.0024	0.1209†
LR test of indep eq. (rho=0)	$\chi^2(1)=0.00$ ($P = 0.986$)	$\chi2(1)=12.47$ ($P = 0.0004$)

DES, drug-eluting stent; LAD, left anterior descending; PCI, percutaneous coronary intervention.
* Significant at the 10% level.
† Significant at the 5% level.

that DES patients have on average €334 ($P=0.599$) less costs than BMS patients, the propensity score matching method estimates the difference between €1429 and €939. As stated in the Methods section, OLS results are quite sensitive to model misspecification and an incorrect propensity score model has smaller biases. We have tested the model specification of our OLS model and found that correct specification cannot be rejected at the 0.01 level (RESET test F=5.05, $P = 0.025$). Because the costs were skewed to the right we also estimated the model with the logarithm of costs but that gave a worse specification (RESET test F=59.24, $P = 0.000$). Because of the sensitivity of the OLS results to the model specification, we were inclined to rely more on the results of the matching. However, the most important result was that, for the diabetic patients, the methods had no significant difference in costs found between BMS and DES patients.

When we look at the other covariates in the OLS model that influence average 1-year costs we can see that most effects are as expected. Male patients have €1221 fewer costs than female patients. Diabetic patients who die in the first quarter incur lower costs than patients who survive the first year after PCI implant; patients who die in the second, third, or fourth quarter are significantly more expensive. (The fact that patients who die in the first quarter have lower costs can be explained by the fact that the follow-up period for them is far less than 1 year. For patients dying in Q2 or later, the shorter follow-up period is obviously dominated by higher costs related to the death of the patient.) Patients who had a PCI after an acute infarction or after thrombolysis have significantly higher costs; patients with stable or asymptomatic disease incur lower costs. It is also found that patients with three-vessel disease compared to one-vessel disease have significantly higher costs (€2062). Patients with renal failure or those who need thrombocyte aggregation blockers during their hospitalization also generate more costs (€8765 and €2222, respectively). Finally, we found that a patient's medical cost in the previous year was a good, independent predictor for future costs.

For the non-diabetic patients we focused on the results of the treatment effect model and found a significant hidden selection bias. We also found a significant difference in costs between BMS and DES patients; compared to patients who received a BMS, patients who received a DES had, on average, €347 less costs during the year after the PCI implanted. Average treatment effects found by the OLS model and propensity score matching do not deviate strongly from this (ATE OLS: €460; ATE prop. scoring €287/€472). When we also correct for hidden selection bias, the difference in costs between BMS and DES decreases and it is significant. This means that selection indicates

that patients who receive a DES tend to be slightly less costly or in better health than those who receive a BMS.

The patient characteristics that explain the costs revealed a lot of similarities with the results of the diabetics. For the non-diabetics, we also found that male patients were less expensive. Costs of patients who die in the first quarter are lower whereas those of patients who die in the other quarters are higher, patients who had an acute infarction or failed thrombolysis when admitted, patients who had more than one diseased vessel, renal failure and those who used thrombocyte aggregation blockers have significantly higher costs. Patients with stable or asymptomatic disease incur lower costs. The costs of the previous year are again a good independent predictor for future costs. For the non-diabetics, in addition to these significant variables, other determinants have a significant impact as well. Elder patients are more expensive; patients living in Flanders or Brussels have higher costs than patients who live in the Walloon region, patients suffering from peripheral vascular disease are €1508 more expensive. Patients for whom the lesion is left main or proximal left anterior descending generate an extra cost; direct stenting provides a cost saving of €508.

Discussion

In this article we compare the 1-year direct medical costs (reimbursements + copayments) of patients who receive a BMS or DES in Belgium in a non-experimental setting using observational data. We use established econometric techniques to account for possible overt and hidden selection bias. This is not standard practice for technology evaluations. The classic framework for evaluation in health care is the randomized experiment. RCT is seen as the gold standard and it should ensure that subjects being compared differ only in their exposure to the intervention being considered [28-30]. As demonstrated in the introduction, the use of RCTs is not without critique.

Given the disadvantages of using RCTs, we considered an observational study to compare DES and BMS. Direct comparison of costs between BMS and DES patients is improper because of the likelihood of selection bias, both overt and hidden. The purpose of this study was to investigate whether it is possible to account for selection bias by using regression analyses for this comparison. It is important to use the appropriate method for this. When costs are only biased by observable patient characteristics we can correct the estimates by performing OLS regressions or by performing propensity score matching. The choice between these two methods depends on the model specification. When model specification is correct, OLS seems to be the most appropriate technique; when we cannot be sure about it, the model specification propensity scoring result in less biased estimates. Cost estimates can also be biased by unobservable patient characteristics or hidden selection bias. In this case we can try to correct the estimates by estimating a treatment effect model.

For patients with diabetes, an unadjusted cost difference of €787 between receiving a DES and BMS ($P = 0.277$) was found; for the non-diabetic patients, an unadjusted, significant cost difference of €725 was found between DES and BMS. We then investigated the possibility of overt and hidden selection bias. For the diabetics it was found that costs were only biased by observable patient characteristics. When we correct the cost estimates for this by using OLS and propensity score matching, it is confirmed that costs of patients who receive a DES are not significantly different from costs of patients who receive a BMS. For non-diabetics the presence of hidden selection bias can be demonstrated. Therefore, a treatment effect model was estimated. It is found that patients who receive a DES incur significantly less costs than patients who receive a BMS: the difference amounts to €347 on average during the year after the PCI implant. When interpreting these results, we have to keep in mind that the cost of antiplatelet therapy was not taken into account. In order to prevent thrombosis, patients must take a second antiplatelet drug (a thienopyridine derivative, either clopidogrel or ticlopidine) in addition to aspirin. Following BMS, dual antiplatelet therapy is mandatory during the first month, whereas 3 to 6 months of dual antiplatelet therapy is advised after DES implantation. From 2006 onward, reports of an increased risk of late stent thrombosis occurring in DES have prompted cardiologists to extend this period up to 12 months, particularly in patients with a low bleeding risk. Antiplatelet drugs are not reimbursed for all patients and could therefore not be abstracted from the Belgian database. Only the antiplatelet costs of reimbursed patients could be taken into account in our calculations. Adding the costs for patients who are not covered for antiplatelet drugs would reduce the cost difference between DES and BMS and may even alter conclusions.

As stated in the introduction, the cost effectiveness of DES was investigated in Neyt et al. [7]. Neyt et al. is a "classic" evaluation model taking cost data and baseline risks from the same databases used in this article and applying a relative risk improvement of 0.34 for DES on the basis of a published meta-analysis of RCTs. They therefore applied both the strengths of observational data and data derived from meta-analysis of randomized trials. They state that due to different underlying characteristics of patients receiving a DES or BMS, no direct comparison is possible. Therefore they set up the situation "as it was" for both the BMS and DES subgroups. Then, they applied the relative improvement of applying DES on the BMS subgroups to model the costs. Similarly, but in the opposite direction, they apply the relative deterioration on the DES subgroups to reflect the situation if they would have been treated with BMS. (Neyt et al. [7] also model quality adjusted life year improvement, but this is disregarded in this comparison, because it is not part of our calculations.) In their model, Neyt et al. [7] calculate that the incremental cost from switching from BMS to DES is positive. In the base-case scenario (with an additional expenditure for DES of €1000, described in these analyses) the mean incremental cost of switching a diabetic patient from BMS to DES is between €793 and €999 (according to the subgroup) and for a non-diabetic patient between €996 and €1061 (according to the subgroup). Similarly, a mean cost saving is found when switching a DES patient to BMS of, respectively, €205 to €769 for non-diabetics and €333 to €863 for diabetics. These results are confirmed by a recent HTA concerning stents [31].

The results are in contrast with those of this article. How should this be interpreted?

As described above, the use of effectiveness information from trials is not entirely without problems. More specifically in this context, it is unclear whether it is correct to assume that the relative risk reduction (TLR) of 0.34 from one meta-analysis of trials can be used for everyday practice in Belgium in 2004 and for all risk groups. The most recent HTA concerning stents used a relative risk reduction of 0.43 (TVR) and 0.24 (TLR) [31]. Other inaccuracies are that Neyt et al. [7] take into account the costs of the hospital stay in which the revascularization takes place; costs of hospitalizations to prepare or diagnose the intervention and of outpatient consultations are not taken into account. Follow-up costs are considered when there is an incremental cost when DES and BMS are compared. For example, they take into account the extra costs DES patients have because they take antiplatelet drugs for a longer period. In addition, for the calculation of baseline risks of restenosis (which is different from staging or from disease progression), a proxy variable based on expert information from cardiologists had to be used. True real-world results could deviate from the calculated figures in the model and be more positive for DES. Hill et al. [32] state:

> The data needed to assess costs needs to include not only revascularization of the target lesion, but any revascularization experienced carried out. (. . .) We do not believe measures of restenosis are of direct relevance, we consider all revascularizations together since it is difficult from routine data sources to distinguish the precise location and nature of an intervention to allow separate analysis and costing. From the viewpoint of the NHS it is the overall cost of all such treatments that matters. (p. 148)

In this article we take into account all 1-year direct medical costs (both inpatient and outpatient) concerning all revascularizations (no distinction is made between restenosis, staging, and disease progression) and even other non-related medical costs.

The methodology used in this article has several advantages. Because we use observational data in regression analysis we can gain insight in the effect of a technology in everyday practice (i.e., outside the experimental setting) and take into account patient and disease characteristics. An additional strength of this methodology is that uncertainty is being considered automatically because *P* values and confidence intervals are calculated for every parameter in the model. At the same time, it is very important to keep in mind that we cannot be sure that the corrections for selection bias in the regression models used in this article are complete; that all observed and unobserved factors that explain the cost differences are taken into account. Biased results could be caused by the use of invalid instruments. Even though the instruments used in our models pass all validity tests available, they could still be too weak to explain the choice of DES. For example, the hospital where the stent was placed was important to explain the choice of stent; when we add hospital dummies to our probit regression the pseudo R^2 rises from 0.0985 to 0.2322. Including these dummies into the instruments would make them much stronger. Unfortunately, when we test the validity of those hospital dummies by the Hansen J statistic (we check whether the instruments are uncorrelated with the error term of the cost equation), the null hypothesis of valid instruments is rejected. Finding valid instruments is often problematic and is the most important disadvantage of this methodology. Another important disadvantage of the treatment effect model is that it relies on relatively strong distributional assumptions (the residuals of both models should be bivariate normal with mean zero and a specific covariance matrix), which are not straightforward to test and correct. In this article we have assumed normality. Parameter estimates could be sensitive to these assumptions [33–35].

Neyt et al. [7] find that in patients receiving a BMS in Belgium, on average there is a cumulative probability of about 15% to have a re-PCI in the first year, but that less than half of these reinterventions are because of restenosis. If about two-thirds of these restenosis-related re-PCIs could be prevented by changing from DES to BMS, this would on average prevent less than an absolute 5% decrease of re-PCIs. The cost of a reintervention should be 20 times higher than the initial extra cost due to DES implantation before DES can result in cost savings.

Several authors have investigated different research designs for the evaluation of costs and effects in clinical research. The main belief is that there is a hierarchy of research design. Often a single RCT is considered to provide true results while results from any observational study are viewed with suspicion [36]. Research, however, indicates that considering research design as a rigid hierarchy is overly simplistic. Different studies that compare results of RCTs and observational studies conclude that average results are remarkably similar [36–38]. Recent publications also show that RCTs continue to generate conflicting results [36]. Concerning the use of stents in Belgium, the results based on RCT data and the results based on an observational study are contradictory. Both methods have their advantages and difficulties. This means that we cannot be sure about the true results and that more research on this topic is indispensable. Assuming that the RCT results are preferable would not be correct. "The results of a single RCT (or only one observational study) cannot be expected to provide a gold standard result for all clinical situations and should be interpreted cautiously" (p. 344) [36].

With the increasing need for valid data on the effectiveness, cost effectiveness, and budget impact of health technologies, there is an increasing need for a broader range of experimental research in the area of technology assessment [11]. Drummond et al. [39] state that when RCT data are not available more use should be made of techniques such as propensity scores, difference-in-difference techniques, time series analyses of natural experiments, and, where appropriate, more sophisticated econometric modeling and structural simulation modeling.

It is therefore interesting and necessary to gain more insight in the relative merits and disadvantages of using RCTs and naturalistic data. It is worthwhile to perform more analyses similar to this to better understand the differences in the results (e.g., repeat it for more years, more regions). Another way forward would be to set up "naturalistic" trials and compare its results with both the RCT and the treatment models.

Conclusion

In this article we investigated the 1-year direct medical cost (reimbursements + copayments) of patients receiving a PCI with DES or BMS in Belgium in a naturalistic setting, using established econometric techniques to account for overt and hidden selection bias. We found no significant cost differences between the two types of stents for diabetic patients and significant cost savings for DES in comparison to BMS for the non-diabetic population. These results are in contrast with the results of an HTA for Belgium using RCT-data in a model (which is standard practice in HTA) [7]. This may be due to the fact that the cost of antiplatelet drugs for patients who are not reimbursed is unavailable and thus not be accounted for in the new observational study. Also, it cannot be ruled out that the treatment effect model does not correct for all observable or unobservable characteristics and that the estimated treatment effect is therefore biased, possibly due to weak instruments and incorrect distributional assumptions.

In conclusion, it is interesting and necessary to explore the use of econometric tools in cost and cost effectiveness analysis to investigate the effect of a technology in everyday practice and to take into account patient and disease characteristics and model uncertainty. Further research is necessary to investigate how we can fully correct for selection bias when using observational data.

REFERENCES

[1] Mittmann N, Brown A, Seung SJ, Coyle D, et al. Economic evaluation of drug eluting stents (technology report no 53), Ottawa: Canadian Coordinating Office for Health Technology Assessment, 2005.

[2] Hill RA, Dündar Y, Bakhai A, et al. Drug eluting stents: an early systematic review to inform policy. Eur Heart J 2004;25: 902–19.

[3] Roiron C, Sanchez P, Bouzamondo A, et al. Drug eluting stents: an updated meta-analysis of randomised controlled trials. Heart 2006;92:641–9.

[4] Spaulding C, Daemen J, Boersma E, et al. A pooled analysis of data comparing sirolimus-eluting stents with bare-metal stents. N Engl J Med 2007;356:989–97.

[5] Abizaid A, Chan C, Lim Y-T, et al. Twelve-month outcomes with a paclitaxel-eluting stent transitioning from controlled trials to clinical practice (the WISDOM Registry). Am J Cardiol 2006;98:1028–32.

[6] Daemen J, Tanimoto S, Garcia-Garcia HM, et al. Comparison of three-year clinical outcome of sirolimus- and paclitaxel-eluting stents versus bare metal stents in patients with ST-segment elevation myocardial infarction (from the RESEARCH and T-SEARCH Registries). Am J Cardiol 2007;99: 1027–32.

[7] Neyt M, Van Brabandt H, Devriese S, et al. Drug eluting stents in Belgium: health technology assessment. Bruxelles: Belgian Health Care Knowledge Centre (KCE), 2007. Report No.: KCE reports 66C.

[8] Neyt M, De Laet C, De Ridder A, et al. Cost effectiveness of drug eluting stents in Belgian practice. Pharmacoeconomics 2009;27:313–27.

[9] Fransen GA, Van Marrewijk CJ, Mujakovic S, et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. British Medical Journal Medical Research Methodology 2007;7:16.

[10] Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. British Medical Journal Medical Research Methodology 2003;3:28.

[11] Freemantle N, Blonde L, Bolinder B, et al. Real-world trials to answer real-world questions. Pharmacoeconomics 2005; 23: 747–54.

[12] March JS, Silva SG, Compton S, et al. The case for practical clinical trials in psychiatry. Acad Psychiatry 2005;162:836–46.

[13] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. J Am Med Assoc 2003;290:1624–32.

[14] Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24:1–19.

[15] McKee M, Britton A, Black N, et al. Methods in health services research: interpreting the evidence: choosing between randomized and non-randomised studies. BMJ 1999;319:312–15.

[16] Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs – same or different? Value in Health 2008;12: 402–4.

[17] Taylor RS, Iglesias CP. Assessing the clinical and cost-effectiveness of medical devices and drugs: are they that different? Value Health 2009;12:404-6.

[18] Polsky D, Basu A. Selection bias in observational data. In: Jones AM, ed., The Elgar Companion to Health Economics. Cheltenham, UK: Edward Elgar Publishing, 2006.

[19] Basu A, Heckman JJ, Navarro-Lozano S, et al. Use of instrumental variables in the presence of heterogeneity and self-selection: an application in breast cancer patients. HEDG Working paper 07/07, ISSN 1751-1976.

[20] Neslusan CA, Hylan TR, Dunn RL, et al. Controlling for systematic selection in retrospective analyses: an application to Fluoxetide and Sertraline prescribing in the United Kingdom. Value in Health 1999;2:435–45.

[21] Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. J Am Stat Assoc 2008;95:573–85.

[22] Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. Value Health 2006;9:377–85.

[23] Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. Stata Journal 2002;2:358–77.

[24] Drake C. Effects of misspecification of the propensity score on estimators of treatment effect, Biometrics 1993; 49:1231–6.

[25] Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fiscal Studies 2000;21:427–46.

[26] RIZIV. Jaarverslag 2005. Available from: http://www.riziv.fgov.be/presentation/nl/publications/annual-report/2005/index.htm) [Accessed September 10, 2008].

[27] Stock JH, Yogo M. Testing for weak instruments in linear IV regression. NBER Technical Working Paper 284. Cambridge, MA: National Bureau of Economic Research. Available from: http://www.nber.org/papers/T0284 [Accessed November 1, 2010.

[28] Britton A, McKee M, Black N, et al. Choosing between randomized and non-randomized studies: a systematic review. Health Technology Assessment NHS R&D HTA Programme 1998;2:1–124.

[29] Basu A, Heckman JJ, Navarro-Lozano S, et al. Use of instrumental variables in the presence of heterogeneity and

self-selection: an application to treatments of breast cancer patients. Health Econ 2007;16:1133–57.
[30] Windmeijer F, Kontodimas S, Knapp M, et al. Methodological approach for assessing the cost-effectiveness of treatments using longitudinal observational data: the SOHO study. Int J Technol Assess Health Care 2006;22:460–8.
[31] Hill RA, Boland A, Dickson R, et al. Drug-eluting stents: a systematic review and economic evaluation. Health Technology Assessment NHS R&D HTA Programme 2007;11:1–221.
[32] Hill RA, Bagust A, Bakhai A, et al. Coronary artery stents: a rapid systematic and economic evaluation. Health Technology Assessment NHS R&D HTA Programme 2004;8:1–242.
[33] Greene WH. Econometric Analysis (6th ed). Upper Saddle River, NJ: Prentice Hall, 2008.
[34] Cameron AC, Trivedi PK. Micro-Econometrics: Methods and Applications. Cambridge, UK: Cambridge University Press, 2005.
[35] Maddala GS. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, UK: Cambridge University Press, 1983.
[36] Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? J Am Soc Exper NeuroTherap 2004;1:341–7.
[37] Concato J, Shah N, Horwitz RI. Randomized controlled trials, observational studies, and the hierarchy of research designs. New Engl J Med 2000;342:1887–92.
[38] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. New Engl J Med 2000;342: 1878–86.
[39] Drummond M, Wheatherly H, Claxton K, et al. Assessing the challenges of applying standard methods of economic evaluation to public health interventions. Public Health Research Consortium. University of York, York. Final report; p. 226.