Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Suzanne van Veen

Comparative Efficiency Analysis from the Perspective of the Dutch Health Care Insurer

Determining the Usefulness of Efficiency Measures for Contracting Primary Care Organizations

MSc Thesis 2010-033

Comparative Efficiency Analysis from the perspective of the Dutch Health Care Insurer:

Determining the Usefulness of Efficiency Measures for Contracting Primary Care Organizations

S.H.C.M. van Veen

September 2010

Master Thesis Health Economics at the Erasmus University Rotterdam

Supervisors:

Dr. T. Van Ourti
Erasmus School of Economics

Prof. Dr. B.M. Balk
Rotterdam School of Management

Dr. P. Stam
SiRM - Strategies in Regulated Markets

Evaluators:

Prof. Dr. F.T. Schut
Institute for Health Policy and Management

Dr. T. Bago d' Uva
Erasmus School of Economics

Content

Summary

This study performed a comparative efficiency analysis from the perspective of the Dutch health care insurer in order to determine the usefulness of efficiency measures for contracting primary care organizations. The introduction of regulated competition on the Dutch health care market has forced the health care insurers to become more concerned with the performance of health care organizations. Health care insurers are interested into both the costs and the quality of the health care services when meeting the needs of their insured. Efficiency analysis can be a useful instrument that health care insurers can use to get insight into the (relative) efficiency level of health care organizations.

Several different efficiency assessment techniques are available to assess the efficiency of primary care organizations from the perspective of the health care insurer. Beforehand there is no 'best method' denotable. A comparative efficiency, where several techniques are applied and the results are compared, can narrow the uncertainty about the assumptions underlying the techniques that may bias the derived efficiency measures. This study has examined four efficiency assessment techniques, namely a Cost Effectiveness Ratio, Data Envelopment Analysis and two parametric techniques, one comparable to Corrected Ordinary Least Squares and one comparable to Stochastic Frontier Analysis.

This study used data from two health care insurers and nine primary health care organizations. The data from the health care insurers consists of administrative data, 1372 observations, and the data from the primary health care organizations consists of administrative and medical data, 265 observations. This study assessed the efficiency of primary care organizations by examining the relationship between the quality of care and the activities by health care providers. The output measures are specified as the score on quality indicators. This study is restricted to assessing the efficiency of primary health care organizations to deliver health care services to chronic patients, since quality indicators are developed for this patient group and it is feasible to define those quality indicators with the available data. Besides, the chronic patient population systematically causes high health care expenditures and therefore it is interesting to examine the efficiency level of the primary care organizations in providing care to this patient group.

Despite the fact that this study does not examine every model, 'only' four models are examined, there are some clear inconsistencies in the results among the different efficiency assessment techniques. It does matter which technique is used and which choices are made to derive the efficiency of the primary care organizations. Health care insurers prefer simple techniques for an efficiency analysis. However, techniques that are less easy might lead to

more appropriate efficiency measures. Thus, for an efficiency analysis from the perspective of the health care insurer, it is about finding a balance between the simplicity of the technique and the appropriateness of the derived efficiency measures. The stochastic parametric technique (comparable to the Stochastic Frontier Analysis) is a less easy technique than the other techniques examined in this study, but this technique might yield more appropriate efficiency measures since it can account for measurement error and noise in the data. Nevertheless, further research is needed to examine the validity of the obtained efficiency measures, because no secure conclusions may be drawn on the results in this study. The sensitivity analysis in this study has not been extensive enough and a new approach of the Stochastic Frontier Analysis is applied, namely a Stochastic Frontier Analysis with a binary output measure on individual level data and a rate on organizational level data. This model needs further investigation before the conclusion can be drawn that this technique is indeed more suitable for efficiency assessment of health care organizations from the perspective of the health care insurer than the other techniques examined in this study.

With the status quo, health care insurers have to rely on efficiency measures that are sensitive to the choice of technique and to the specification of the model. Therefore, it is not satisfactory to use those efficiency measures for contracting health care organizations. If the health care insurer wants to use efficiency assessment, then efficiency assessment could be used as an instrument to inform the health care insurer about the relative efficiency of the organization, while keeping in mind the sensitivity of the efficiency measures to the choice of technique and the specification of the model.

Preface

This study is the master thesis to graduate for the Master Health Economics at the Erasmus University Rotterdam (Netherlands). It was a challenge to write my master thesis about efficiency analysis in the health care sector, since it is an area with many unresolved issues. In the last five months, I worked with pleasure on this study and it was a challenge seeking solutions to deal with some unresolved issues. This study uses some new approaches to analyze the efficiency of primary health care organizations from the perspective of health care insurer. It goes beyond this master thesis to deal with all the unresolved issues in efficiency analysis in the health care sector and to review all the models that are available to assess the efficiency of organizations. This study hopes to contribute to further research on efficiency analysis in the health care sector.

In particular, I am grateful to some persons who have contributed to the study into its present form. I am grateful to Dr. P. Stam, Dr. X. Koolman and all the other members of SiRM - Strategies in Regulated Markets for the comments and the pleasant time working with you. This study has benefited from the discussion with you all. Thanks also to Prof. Dr. B.M. Balk and Dr. T. Van Ourti for the comments and efforts. The conversations and discussions have contributed to the study into the present form and it was pleasant to work with you.

Suzanne van Veen

September 2010

Introduction

This study considers assessing the efficiency of primary health care organizations from the perspective of the Dutch health care insurer. Efficiency is a concept that describes the relationship between the inputs and outputs of a production process as follows. A technically efficient organization maximizes the output of the production process given a fixed level of inputs (Farrell 1957). The interest of the Dutch health care insurer to measure the efficiency of the health care organizations is part of a larger demand for efficiency assessment in the (Dutch) health care sector. With the Dutch Health Care Reform in 2006, the Dutch government purposes to realize an efficient health care system. Within the Dutch context, the regulators monitor the competition on the market and the health care insurers purchase care on behalf of their insured. Since the introduction of regulated competition on the Dutch health care market, the health care insurers become more concerned with the performance of health care organizations. The regulators and the health care insurers are interested into the efficiency of health care organizations in order to judge whether the health care expenditures are in line with the quality of care and the consumers' preferences.

Efficiency assessment from the perspective of the health care purchaser is not a commonly used perspective for efficiency analysis. A more common used perspective, besides the traditional managerial perspective, is the perspective of the policy-maker. The efficiency studies in the Netherlands have analyzed the efficiency of health care organizations from the managerial perspective or the policy-maker perspective (Blank & Eggink 1998; Eggink & Blank 2001; Van der Veen et al 2001; Blank & Eggink 2004; Bakker 2004). Within the Dutch context, this study uses a perspective not used earlier in efficiency studies. In efficiency literature, efficiency analysis from the perspective of the policy-maker or the health care insurer is called the *regulatory perspective*. An efficiency analysis from the regulatory perspective is concerned with measuring the relative efficiency of organizations; this is the efficiency of one organization compared to the other organizations in the sample. The policy-maker as well as the health care insurer could use the information obtained by an efficiency analysis for benchmarking organizations, which is stimulating the organization to improve their own performance by comparing their performance to the performance of other organizations. The policy-makers could use the information obtained by an efficiency analysis for effective regulation, whereas the health care insurer could use this information for (selective) contracting the health care organizations (Parking & Hollingsworth 1997). The use of efficiency measures for explanatory or predictive purposes, such as monitoring the market or contracting health care organizations, lays a strong focus on the validity of the derived efficiency measures. A regulator or health care insurer should use only efficiency measures for explanatory or predictive purposes if those efficiency measures accurately reflect the

actual relative performance of the health care organizations. This study examines the usefulness of efficiency measures derived by an efficiency analysis to (selective) contract primary care organizations. This study will focus on the validity of the obtained efficiency measures.

Traditional efficiency assessment techniques developed for managerial purposes are in principle useful tools to measure the efficiency of organizations from the regulatory perspective. Several efficiency measurement techniques are available to assess the efficiency of health care organizations. Every efficiency measurement technique has its advantages and disadvantages. The assumptions underlying each technique should be examined in order to determine the validity of the efficiency measures. The derived efficiency measures can be sensitive to the choices made in the analysis (Jacobs et al 2006). A comparative efficiency analysis is a useful method to determine the relative merits of the assumptions underlying the efficiency assessment techniques. A comparative efficiency analysis is an analysis whereby the researcher assesses the efficiency of organizations by more than one technique and whereby the researcher compare the findings from these different techniques. This study uses a comparative efficiency analysis to examine the sensitivity of the efficiency measures to the assumptions underlying the efficiency assessment techniques.

This study applies four efficiency measurement techniques to assess the efficiency of primary care organizations. These techniques are the Cost Effectiveness Ratio, Data Envelopment Analysis, Corrected Ordinary Least Squares and Stochastic Frontier Analysis. Section 2 provides a theoretical discussion of these techniques. The data and the model specify the boundaries of an efficiency analysis (Coelli et al 2005). Section 1 describes the data of this study and the conceptual model. Section 3 presents the method and the analysis on this study. First, is questioned, how can the efficiency of primary health care organizations be assessed from the perspective of the health care insurer by each of the four techniques? Two questions need some consideration by analyzing the results, namely what is the sensitivity of the efficiency measures to the assumptions underlying the techniques and what are the reasons for the consistency or inconsistency in the results? Section 4 tries to answer the question how the health care insurer can use the derived efficiency measures in this study for contracting health care organizations. The discussion in this study is a balanced critique on the used data and on the specification of the model for assessing the efficiency of primary health care organizations. The last section of this study concludes and points out areas for further research.

Section 1: Data and conceptual model

The first step and probably the most difficult step in an efficiency analysis is the specification of the conceptual model. The available data specifies the boundaries for specification of the conceptual model. The data used for efficiency assessment is just as important as the efficiency measurement technique. This is because an efficiency measurement technique cannot overcome problems residing in the data, no matter how powerful the technique is (Coelli et al 2005). Efficiency studies in the health care sector often have to deal with missing data, omitted variables or small sample sizes (Jacobs et al 2006), which makes it hard to specify the model. The model should reflect the production process of the organization, which is the ability of the organization to convert inputs into valuable outputs (Coelli et al 2005). This section describes the data and the conceptual model used in this study.

§1.1 Data of this study

To perform an efficiency analysis from the perspective of the health care insurer, it is necessary to have data about the inputs and outputs of the production process and about environmental factors. Data about environmental factors is needed to control for factors that may influence the production or the quality of products, but that are beyond the control of the organization. This study uses data of health care insurers and primary health care organizations (PCOs) to define the inputs, outputs and the environmental factors.

Health care insurers have individual administrative data. This dataset consists of individual data about the volume and the costs of health care services, the volume and costs of drugs usage and individual characteristics such as age and gender. An efficiency analysis solely based on the dataset of health care insurers can be restrictive since data is missing about to 'product' of the health care organizations that is the improved health status and the responsiveness[1] to the expectations of individuals (Murray & Frenk 2000). When data is missing about the actual 'product' of the health care organizations is the researcher forced to use intermediate outcomes as the output measure, such as the number of activities (Jacobs et al 2006). The use of intermediate outcomes in an efficiency analysis imposes limitations, because the actual 'product' of health care organizations is not defined (Parking & Hollingsworth 1997; Jacobs et al 2006). In order to measure the efficiency of health care organizations from the perspective of the health care insurer, it is valuable to append data of health care insurers and data of health care organizations. Health care organizations register

[1] The World Health Organization defines responsiveness as enhancing the expectations of the individual in terms of respect for the person (dignity, autonomy and confidentially) and individual orientation (prompt attention, social support and choice of provider) (Murray & Frenk 2000).

both administrative data and medical data. The dataset of health care organizations consists of individual data about the volume of the health care service, the outcome of the health care service (outcomes of some medical tests) and individual characteristics. The use of administrative data as well as medical data enables the researcher to define the health status of individuals that can serve as a proxy for the 'product' of health care organizations, rather than relying on an intermediate output measure.

A dataset consisting of data from health care insurers as well as data from health care organizations can only be derived when both parties are willing to share their data. The willingness to share data is not a usual incentive in the current Dutch health care sector since health care organizations operate on a (regulated) competitive market whereby the health care insurers (selective) contract these health care organizations. An efficiency analysis can give insight into the relative efficiency level of organizations that the health care insurer might want to use for (selective) contracting. This study uses data of health care insurers and PCOs. The focus of this study is, therefore, on assessing the efficiency of PCOs. In the Netherlands, to the best of my knowledge, it is the first study that uses data of health care insurers and health care organizations to perform an efficiency analysis.

This study uses two datasets to analyze the efficiency of PCOs, because it was not possible to append the data of the health care insurers to the data of the PCOs for each unique individual in the datasets. The data for this study is obtained by appending 134 datasets of two health care insurers and nine PCOs[2]. The data is appended by using pseudonyms for each unique individual. First, one dataset, X, (1,372 observations) is the result of appending data of two health care insurers. Dataset X consists of individual administrative data records over the period 2007 to 2009. Another dataset, Y, (265 observations) is the result of appending data of the PCOs to the dataset X of the health care insurers. Dataset Y consists of individual administrative and medical data over the period 2007 to 2009. Dataset Y is used only when medical data is required to define the variable; in all other cases dataset X is used. This is because using dataset Y for the efficiency analysis would mean a large loss of information.

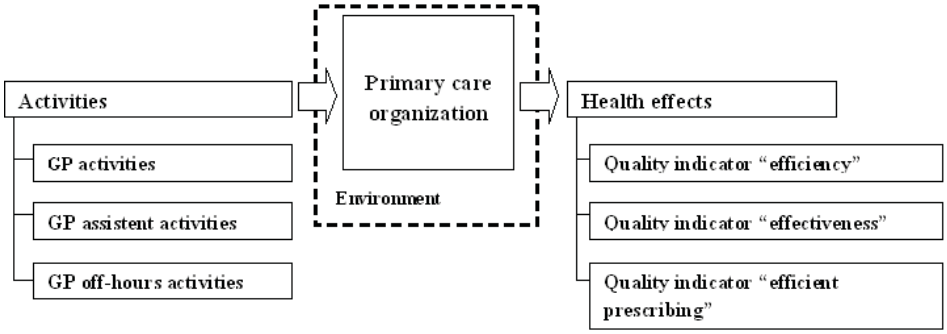
§1.2 Conceptual model

The conceptual model defines how the efficiency of the organizations in the sample is assessed. The model specification depends on the availability of data and the context of the efficiency analysis (Jacobs et al 2006). Figure 1 presents the conceptual model of this study. Figure 1 can be applied to cross-sections, but in this study, it is applied to panel data over the

[2] The name of the health care insurers and primary health care organizations are confidential

period 2007 to 2009.

Figure 1: Conceptual model



In particular, three separate inputs are specified to analyze the efficiency of PCOs. The inputs include the number of activities by a GP (AC_GP), a GP assistant (AC_GPA) and a GP in the off-hours (AC_GPO). These three separate inputs are defined, because these inputs are attributable to an individual. This makes it feasible to appropriately examine the relationship between the indicator-score per individual (output) and the corresponding consumed inputs. Data about aggregate inputs used in general in efficiency studies, such as labor, capital, energy, material and purchased services, is not available in the data of this study. The three separate inputs in Figure 1 are aggregate input measures of several activities performed by each health care provider. Those activities are presented in Table 1. The activities are aggregated, because the interpretation of the results may become complex when many inputs and outputs are specified. Moreover, many activities have a zero-value, which can cause problems when using one of the techniques applied in this study (Section 3). The different activities are weighted by their administrative price in order to specify a single input for each health care provider. The administrative prices of the several activities presented in Table 1 are constant over the period 2007 to 2009 (NZA 2007; NZ 2008; NZA 2009).

Table 1: health care provider activities and their administrative prices

Activities	**Price (Euro's)**
GP consult†	9.00
GP consult > 20 minutes	18.00
GP visit	13.50
GP visit > 20 minutes	22.50
GP telephone consult	4.50
GP repeated prescription	4.50
Assistant off-hours consult	37.70
Assistant off-hours visit	49.10
Assistant off-hours telephone consult	16.40

† The same activities are taken into account for the GP assistant. The prices for a GP assistant are the same as for a GP. Source: NZA 2009

Notice that weighting the inputs by their administrative price does not mean estimation of a cost-function. The administrative prices do not reflect the costs of PCOs, since that are the costs on labor, capital, material and other resources needed to deliver the health care services. The administrative prices are used to attach a relative value to the different activities in order to aggregate them into a single input for each health care provider. This study examines the efficiency of PCOs through analyzing production relationships.

Unlike most previous efficiency studies in the health care sector where the output measure is defined as measures of activities, such as the number of patients treated, the number of operations undertaken or hospitals days (Hollingsworth 2003), the output measure in this study is a measure for improved health status. This study does not incorporate responsiveness as one of the 'products' of health care organizations, since the data to define this 'product' is not available[3]. This study uses quality indicators to define the output of PCOs whereby the score on those quality indicators is the output measure. Quality indicators are developed to measure the quality of care of organizations and health care insurers can use those quality indicators for contracting health care organizations. This study uses those quality indicators to assess the relationship with the activities of PCOs that is the efficiency of PCOs. For the health care insurer it can be interesting to examine the efficiency of PCOs to deliver care to chronic patients, because chronic patients systematically cause high health care expenditures. This study is restricted to analyze the efficiency of PCOs to provide health care services to the chronic patient population (Diabetes Mellitus, Chronic Obstructive Pulmonary Disease and Cardiovascular Risk Management). This is because quality indicators are developed for the health care services to chronic patient groups (ZN 2008). It is not feasible to define a variable for the improved health status of individuals that have not one of the three chronic disorders.

When an output measure 'health effects' is used, it is required to measure the health status of an individual before and after the health care service (Evans et al 2000). It is not feasible to make a comparison of the health status of an individual with and without an intervention, since it is in the health care sector not legitimate to let an individual untreated (Jacobs et al 2006). Health status measurement is therefore restricted to a before/after comparison. This study uses such a before/after comparison to define the health effects produced by PCOs. The assumption made here is that the health status of the individuals is the same among all chronic patients. Thus, being a chronic patient serves as the baseline ('before intervention'). The indicator score then presents the health effects produced by the PCO ('after intervention').

[3] Consumer Quality Index-data (CQ-data) could be used to define the 'product' responsiveness. CQ-data is data about the individual experience of the quality of care.

The output measure used here is a binary variable on individual level and a rate on organizational level. An example of a quality indicator is the number of patients with a chronic disease-related outpatient visit per patient per year (nominator) to the total chronic patient population (denominator). On individual level, the indicator score is a 'yes' (1) if the individual has visit outpatient care and a 'no' (0) if the individual has not visit outpatient care. On organizational level, the indicator-score is a rate from 0 to 100 whereby a 0 denotes a bad performer and a 100 stands for an excellent performer. One can imagine that an organization with a high indicator score, in other words a low prevalence on the nominator of the indicator, indicates being a good performer, because the PCO does succeed in preventing disease-related outpatient care. The output in this study differs from traditional output measures in efficiency assessments. Output measures at the individual and organizational level in traditional efficiency assessments are continuous variables that can take any positive value. The traditional efficiency assessment techniques, therefore, are developed for the use of continuous variables. This study uses a new approach; applying traditional efficiency assessment techniques with a binary variable on individual level and a rate at the organizational level. The use of an output measure as a binary variable or as a rate depends on the efficiency measurement technique used to analyze the efficiency of PCOs (Section 3).

This study uses three separate outputs. The output 'efficiency' (O_EFF) consists of indicators that measure the ability of the PCO to prevent outpatient care and hospitalization. The output 'effectiveness' (O_EFFEC) measures the medical outcomes of the health care service or the process of the health care service. The output 'efficient prescribing' (O_EP) measures the behavior of health care organizations to prescribe (expensive) drugs or combinations of drugs. The indicators are defined according to the definitions of the Dutch Health Insurance Association (ZN 2008). This study uses sixteen indicators to analyze the efficiency of PCOs, seven indicators for the category 'efficiency', eight indicators for the category 'effectiveness' and one indicator for category 'efficient prescribing'. Not all three chronic disorders are proportionally incorporated into the analysis of this study. Most of the indicators of the Diabetes Mellitus patient group are definable with the available datasets (Appendix 1). The indicator-score on each category is the average score of the quality indicators within the category. Assumed is that each quality indicator is of similar importance. Five quality indicators, three quality indicators of the category 'effectiveness' and two indicators of the category 'efficient prescribing', could not be defined with the available datasets (Appendix 2). The performance of the PCOs on those five quality indicators is unknown. The indicator-scores used in this study reflect a part of the overall quality of care of PCOs.

A comparative efficiency analysis from the regulatory perspective requires taking into account external influences (Jacobs et al 2006). In order to make a fair comparison among the organizations in the sample, it is necessary to account for factors that are beyond the control of the organization and that may influence the output of the organization. A method to account for differences in environmental factors is by adjusting the output measure for these differences (Jacobs et al 2006). This method is applied in this study.

This study accounts for case-mix differences. Other external influences, such as variations in policy regulation, variations in the organization of hospital care and differences in the preferences of the local community, are not taken into account. The reason for this is that the data to define those factors is not available. This study assumes that there is no significant variation among the PCOs in the mentioned external factors. This may be an appropriate assumption, because the PCOs in the sample are located in the same local geographical area. The case-mix of an organization is a measure for the characteristics of the patient population of the PCO, such as the gender, age, social-economic status and health status of the individuals. This study adjusts the output (O_EFF) for case-mix differences since for a PCO with a 'bad' case-mix, for example a relative old individual with co-morbidity, it may be harder to attain a health effect than for a PCO with a 'good' case-mix, that is a relative young individual with one chronic disease. The case-mix of the organization has no influence on the score of the outputs (O_EFFEC) and (O_EP) (ZN 2008) and therefore the score on those two outputs are not adjusted for case-mix differences. To adjust the output (O_EFF) for case-mix differences are four variables defined, a variable for age (AGE), a variable for gender (FEMALE) and two variables to proxy the health status of the individual, a variable for co morbidity (COMOB) and one for the usage of more than five drugs (POLYF). Assumed is that an individual with more than one chronic disease or using more than five drugs in a year has a relative less good health status than individuals with one chronic disease or using less than five drugs. Social economic status is not taken into account, because it is not feasible to define a variable for this with the available data. The method to adjust the output (O_EFF) for case-mix differences depends on the efficiency measurement technique used for the efficiency analysis (Section 3).

Table 2 reports the descriptive statistics of the variables in the conceptual model on organizational level. The rows are the PCOs per specific year and the columns are the variables used in the analysis. The output (O_EFFEC) in Table 2 is based on dataset Y, because medical data is needed to define this variable. All the other variables in Table 2 are based on dataset X. The three input measures in Table 2 are the weighted number of activities per patient per year. The output measures are the average score on the indicators within that

Table 2: Descriptive statistics

		Inputs			Outputs			Environmental factors			
	PCO	*AC_GP*	*AC_GPA*	*AC_GPO*	*O_EFF*	*O_EFFEC*	*O_EP*	*AGE*	*FEMALE (0 – 1)*	*COMOB (0 – 1)*	*POLYF (0-1)*
	1	68.75	1.25	2.728	93.77	70.694	83.333	54.389 (21.511)†	7 – 11	14 – 4	17 – 1
	2	69.9	0	0	84.779	70.804	66.667	43.267 (22.867)	23 – 22	41 – 4	45 – 0
	3	94.5	0	8.175	95.238	77.5	66.667	62.500 (27.550)	2 – 2	4 – 0	4 – 0
year	**4**	64.957	6.457	0	96.992	83.421	100	69.435 (10.286)	13 – 16	21 – 2	23 – 0
2007	**5**	59.46	10.08	1.745	87.27	30.556	57.143	58.720 (19.182)	36 – 39	65 – 10	75 – 0
	6	69.173	2.481	0.629	93.127	73.083	46.154	53.603 (24.843)	45 – 33	68 – 10	72 – 2
	7	106.031	3.234	2.556	86.259	78.69	66.667	57.906 (16.633)	16 – 16	29 – 3	30 – 6
	8	72.096	6.606	0	84.409	72.612	60	54.915 (20.184)	24 – 23	42 – 5	40 – 7
	9	80.209	6.231	0.359	85.602	74.841	56	59.703 (19.501)	42 – 49	83 – 8	84 – 7
	mean	76.12	4.037	1.8	89.716	70.244	66.959	57.160	23 – 23	41 – 5	43 – 3
	std.dev	15.037	3.486	2.628	4.986	15.444	16.038	7.133	208 – 205 (total)	367 – 46 (total)	390 – 23 (total)
	1	118.895	8.053	4.305	62.765	75.412	63.636	57.773 (17.636)	9 – 13	19 – 3	18 – 4
	2	67.827	0.409	0.595	60.594	79.042	76.923	44.872 (22.415)	24 – 23	44 – 3	43 – 4
	3	93.8	6.7	0	68.748	78.373	66.667	64.000 (22.415)	25 – 25	47 – 3	49 – 1
year	**4**	107.217	2.543	0	83	82	62.5	69.467 (12.300)	13 – 17	26 – 4	29 – 1
2008	**5**	66.32	9.262	1.43	74.72	79.734	53.846	57.547 (20.371)	47 – 48	77 – 18	92 – 3
	6	65.191	3.434	2.368	71.741	69.42	75	56.029 (25.190)	36 – 34	64 – 6	63 – 7
	7	83.25	2.966	1.116	75.13	81.917	69.231	54.368 (19.344)	18 – 20	34 – 4	36 – 2
	8	47.754	3.493	0	76.7	81.312	56.25	57.313 (19.403)	23 – 25	43 – 5	42 – 6
	9	47.571	0.468	0	73.649	79.602	60	60.640 (18.529)	21 – 29	42 – 8	49 – 1
	mean	77.536	4.147	1.090	71.894	78.535	64.895	58.001	24 – 26	44 – 6	47 – 3
	std.dev	25.115	3.169	1.461	6.972	3.982	7.880	6.751	216 – 234 (total)	396 – 54 (total)	421 – 29 (total)
	1	85.974	0.947	2.584	69.02	81.964	66.667	57.368 (17.902)	6 – 13	14 – 5	17 – 2
	2	74.945	0.491	2.975	78.878	73.077	93.333	48.491 (21.446)	25 – 30	48 – 7	53 – 2
	3	74	6.3	2.909	85.449	77.937	64.286	63.778 (15.072)	25 – 20	43 – 2	44 – 1
year	**4**	60.261	6.261	0	80.925	76.475	50	74.478 (9.105)	9 – 14	19 – 4	21 – 2
2009	**5**	69.684	14.112	1.909	75.326	75.643	81.818	58.854 (19.618)	53 – 50	86 – 17	100 – 3
	6	76.263	5.921	1.937	77.755	69.662	87.5	61.434 (18.653)	36 – 40	68 – 8	73 – 3
	7	96.034	4.602	2.23	73.829	82.451	83.333	56.955 (16.491)	18 – 26	40 – 4	41 – 3
	8	60.045	3.627	4.639	78.183	69.961	88.889	56.940 (18.422)	35 – 32	59 – 8	58 – 9
	9	69.896	1.636	3.188	77.358	77.592	76.19	59.416 (21.255)	33 – 44	65 – 12	72 – 5
	mean	74.122	4.877	2.486	77.414	76.084	76.891	59.746	27 – 30	49 – 7	53 – 3
	std.dev	11.477	4.14	1.247	4.574	4.596	14.052	6.938	240 – 269 (total)	442 – 67 (total)	479 – 30 (total)

† standard deviation in parentheses

Note: The descriptive statistics in Table 2 are on organizational level

output. The indicator-scores on (O_EFF) in Table 2 are not adjusted for case-mix differences in order to present the 'absolute' indicator-score instead of the relative indicator-score that is used in the analysis of this study. The variable for age (AGE) denotes the average age of the patient population of each PCO. The other three environmental factors are binary variables reported into frequencies. In this study, technical efficiency is defined as the ability of the PCO to maximize the indicator score given a level of consumed activities. In Table 2, PCO 4 seems to be an efficient PCO in 2007 since it uses a relative low number of activities per patient to produce the highest indicators-scores in that year. PCO 1 in 2008 seems to be a relative inefficient PCO, because it uses far more activities per patient and produces one of the lowest indicator-scores. Based on Table 2, it is difficult to determine the relative efficiency of each PCO in the sample, because PCOs uses more or less activities at one or two of the inputs and produce a higher or lower score on one or two outputs. Table 2 shows large variation in the environmental factors over time, such as PCO 3, while it is expected to have quite stable environmental factors or a (slightly) increase over the years due to an increase in the patient population of PCOs. This leads to the conclusion that data is missing, which sets doubts about the reliability of the data. The differences in the organizational efficiency level of the PCO in this study will not solely be attributable to the assumptions underlying the efficiency assessment technique, but the differences in the derived efficiency measures may also be (partly) attributable to the data used in this study. The quality of the data needs consideration when drawing conclusions on the derived efficiency measures in this study.

Dataset Y is a selective dataset of dataset X that means that there is selection bias in the score on the output (O_EFFEC) presented in Table 2 (Appendix 3). Most of the PCOs delivered an older and/or sicker population than the population in dataset X. The level of the quality of care on the output (O_EFFEC) is therefore an underestimation of the actual quality of care of PCOs on those indicators, when assuming that it is harder to attain a health effect by an older and/or sicker individual. The selection bias is largest for PCO 2 and PCO 4, since those two PCOs delivered the most selective patient group. The selection bias into the results of this study requires attention when determining the usefulness of the obtained efficiency measures for contracting PCOs. Underestimating the efficiency level of PCOs can cause problems when the efficiency measures in this study will be used to set a baseline for efficiency assessments in a next year(s). For a PCO it may be easier to improve their organizational efficiency in the next year(s), because the baseline is an underestimation of the actual efficiency level.

Since the efficiency analysis is restricted to examination of the efficiency of PCO in delivering health care services to the chronic patient population, it is necessary to select this population in the data. No data is available about the diagnosis of individuals in the data of

this study. The population for the efficiency analysis is identified by using registered drugs usage. Table 3 reports the variables used for the identification of the chronic patient population. There is no certainty that indirect identification of the population by the drugs usage of individuals has lead to identification of the actual patient population. It could be that individuals are missed with the selection procedure, because individuals can use other drugs than those used here. In the Dutch context, an individual using disease-specific drug over a longer period, six months, will be assigned to a Pharmaceutical Cost Group (PCG). PCGs are not used to identify the patient population in this study, because a PCG is more restrictive than the usage of some disease-specific drugs that would lead to a considerably loss of information (Appendix 3). Besides, it is unlikely that an individual that is not chronically ill uses one of the disease-specific drugs in Table 3.

Table 3: variables for identification of chronic patient population

Chronic patient groups	**Drug usage**
Diabetes Mellitus type 2 (DM)	Oral diabetes drugs
Chronic Obstructive Pulmonary Disease (COPD)	Inhalation steroid drugs
Cardiovascular Risk Management (CVRM)	Beta-blocker drugs

In order to measure the efficiency of PCOs, it is necessary to assign each individual in dataset X to a PCO, because health care insurers do not register where the individual has received the health care service. Insurers only identify the individual providers working at the PCOs in order to pay the provider claims, not the PCOs where they work. In dataset Y, this information is available. This study has used the GP-code registered in the database by Vektis[4] to assign individuals to PCOs. From the Vektis database, it can be concluded that no GPs working at more than one of the PCOs in the sample of this study. This makes it appropriate to assign each individual uniquely to a PCO. Individuals with a missing value for the GP-code are excluded from the analysis, because it is not feasible to assign these individuals to a PCO (Appendix 3). The assumption made here is that each individual has received the health care service at the PCO where the GP works according to the Vektis database. Notice that the individual still may have received the health care service elsewhere in practice. It is unavoidable to make assumptions about the identification of individuals and the assignment of these individuals to PCOs when data of health care insurers is used, as long as the health care insurers do not register this information.

The restriction of an efficiency analysis to examine a part of the entire production process of

[4] Vektis is a database with general information, such as contact information and information about the working place, of all the health care providers in the Netherlands (www.vektis.nl).

an organization can impose limitations (Jacobs et al 2006). The assumption made in this study is that the health care services to chronic patients are independent of the health care services to other patients. This means that the potential influence of other actions of health care providers or other providers than those incorporated into the analysis that contribute to the health care service to chronic patients is ignored. The results should be judged as the efficiency of PCOs in delivering health care services to chronic patients insured by the two health care insurers. For the health care insurer the restriction of the analysis imposes no serious limitations, because the health care insurer is interested in the efficiency of PCOs in delivering health care services to their insured, as long as the efficiency measures reflect the actual relative efficiency level of the PCOs. The PCOs should be cautious with an interpretation of the results in this study. In this study, the efficiency of PCOs is analyzed for a part of their total patient population and not all quality indicators are incorporated into the efficiency analysis.

With the remarks of the data and model in mind, we now turn to the assessment of the efficiency of PCOs by several efficiency methodologies. The next sections focus on efficiency assessment methodologies and the comparison of the efficiency measures obtained by those efficiency assessment methodologies given the data and the specification of the conceptual model described in this section.

Section 2: Efficiency assessment methodologies

The focus of an efficiency analysis from a regulatory perspective lies on the relative efficiency of organizations. Frontier techniques measure the efficiency of an organization by comparing the efficiency level of that organization to the best practice in the sample. Frontier techniques are often used for efficiency assessment (Hollingsworth 2003). A common classification for these frontier techniques is to distinguish parametric from nonparametric techniques and stochastic from deterministic techniques. Parametric techniques estimate econometric functions in order to derive the efficiency of organizations, whereas nonparametric techniques use linear programming. Stochastic techniques accounts for noise in the data, while deterministic techniques ignores measurement error and noise in the data.

This study uses four efficiency measurement techniques to assess the efficiency of PCOs, namely the Cost Effectiveness Ratio (CER), Data Envelopment Analysis (DEA), Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA). DEA, a nonparametric deterministic technique, and SFA, a parametric stochastic technique, are the most frequently used techniques to measure the efficiency of health care organizations (Hollingsworth 2003). These techniques have some important properties that make them useful for efficiency assessment in the health care sector. Despite the fact that DEA and SFA are used frequently, they have been also highly criticized for efficiency assessments of health care organizations. Just as with every methodology, DEA and SFA require making assumptions. These assumptions can make efficiency assessment by DEA and SFA complex. For the health care insurer it is interesting to seek for simple methodologies for efficiency assessment. Both for DEA and SFA exists a technique that is almost similar, but that is more easy to use. A simple variant for DEA is a CER. A CER is just as DEA a nonparametric deterministic technique, but it is the only technique in this study that is not a frontier technique. The simple variant of SFA is the COLS, which is a parametric deterministic technique.

Unlike the 'standard' efficiency assessment techniques measure the efficiency of organizations on organizational level, this study will assess the efficiency of organizations by parametric techniques on individual level. Nonparametric techniques do not allow for individual data and therefore the two nonparametric techniques in this study will be applied to organizational data, which is the individual data aggregated to the organizational level. Parametric techniques do allow for individual data. Individual data has never been used before in efficiency studies. It is interesting to apply a new approach to the parametric techniques and to find out how individual data in efficiency studies could be used. Although the parametric techniques in this study are applied on individual data, the assumptions

underlying those techniques will be similar to the 'standard' COLS and SFA. This section provides a theoretical overview of the four efficiency assessment methodologies in order to clarify the assumptions underlying those techniques. The 'standard' COLS and SFA will be discussed here. Section 3.3 will describe the specification of the parametric techniques applied in this study. The model of the parametric techniques applied in this study will deviate from those discussed in this section. The discussion of the four efficiency measurement techniques in this section begins with the simplest technique to go on with techniques that increasingly become more complex.

§ 2.1 Nonparametric efficiency assessment methodologies

Nonparametric techniques are techniques that rely on linear programming. Linear programming is solving a record of problems by maximization or minimization of a subject to some constraints (Sarafadis 2002). The use of nonparametric techniques for efficiency assessment requires making some assumptions. This section discusses the two nonparametric techniques, the CER and DEA, used in this study.

§ 2.1.1 Cost effectiveness ratio

The simplest method to measure the efficiency of organizations is the CER. A CER is an index number that measures the efficiency of organization as follows:

$$TE_{CER,i} = (\sum u_s y_{si}) / (\sum v_m x_{mi}) \qquad \begin{array}{l} i = 1, 2\ldots, N \\ s = 1, 2,\ldots, S \\ m = 1, 2,\ldots, M \\ \text{, with u and v given} \end{array} \qquad (1.0)$$

$TE_{CER,i}$ denotes the ratio of the weighted outputs to the weighted inputs. The y represents the output of the i^{th} organization of N organizations in the sample; u denotes the weights to the s^{th} output; and x represents the m^{th} input of the i^{th} organization and v denotes the weight to the m^{th} input. The researcher specifies the weights u and v a priori. If the cost effectiveness of all the organizations in the sample is calculated, then the organization with the highest ratio is the most efficient organization, since this organization produces the most output given a level of inputs. All the organizations in the sample with a lower ratio are inefficient, whereby the inefficiency is the extent to which the organization can increase the units of outputs while consuming the same units of inputs.

The advantage of the CER above the frontier techniques is that the efficiency measure is not sensitive to the presence of other PCOs in the sample (Coelli et al 2005). The CER reports the

organizational specific index of the outputs to the inputs whereas the frontier technique reports the relative efficiency level. A second advantage of the CER is that the method is straightforward and easy to use for health care insurers. Besides, the CER is just as any other nonparametric technique free from the assumptions underlying parametric techniques (Section 2.2).

A common critique on the CER for efficiency assessment is that it is too simple to analyze complex multiple input-multiple output production processes such as the production process of health care organizations (Sarafadis 2002). The use of simple methods often corresponds to making strong assumptions (Jacobs et al 2006). The same accounts for the CER. The limitation of the CER is that certain strong assumptions needs to be hold to derive appropriate efficiency measures. Two assumptions underlying the CER are in particular of interest in this study since one of the other efficiency assessment techniques applied in this study relaxes one of the two assumptions. First, an important assumption underlie the CER is that the weights of the inputs and outputs needs to be known a priori in order to construct a single index. In the health care sector, the weights of the inputs and output are unknown, since there are no market prices available that can serve as a weight. The use of quasi-market prices may yield misleading results, because those prices do not reflect the relative importance of the inputs and outputs. Moreover, it is difficult to specify a weight for each organization in the sample. Therefore, often the weights take the same value for all the organizations in the sample (Sarafadis 2002). Organizations can operate in different circumstances that lead to variation in the importance to the inputs and outputs and therefore it may not always be legitimate to attach the same weight to all the inputs and outputs (Sarafidis 2002). DEA is a technique that does not require a priori specification of the weights (Section 2.1.2). Another important assumption underlying the CER is the assumption about stochastic. The CER is just as DEA a deterministic technique implying to assume that all relevant inputs and outputs is observed and is measured without error (Jacobs et al 2006; Coelli et al 2005). SFA is a technique that relaxes this assumption about stochastic (Section 2.2.2).

§ 2.1.2 Data Envelopment Analysis

DEA is a technique developed to measure the efficiency of organizations with multiple inputs and multiple outputs when the weights of these inputs and outputs are unknown. DEA involves solving a linear program. The (primal) linear program of DEA is (Coelli et al 2005):

$$\max TE_{DEAi} \quad \text{subject to } TE_{DEAi} = (\sum u_{si}y_{si}) / (\sum v_{mi}x_{mi}) \qquad (2.0)$$

$$\sum u_{si}y_{sj} - \sum v_{mi}x_{mj} \leq 1 \qquad i = 1,2,\ldots,N \qquad j= 1,2,\ldots,N$$

$$u_{si}, v_{mi} \geq 0 \qquad s = 1, 2,\ldots,S \qquad m = 1, 2,\ldots,M$$

Equation (2.0) is similar to equation (1.0), but the only difference is that DEA imposes some constraints to seek automatically the weights u and v. The constraint is that these weights are non-negative. Equation (2.0) will be solved N times, since for each organization the efficiency score will be calculated. DEA seeks weights that maximize the efficiency score of each organization with the constraint that the efficiency score cannot exceed one. In other words, DEA sheds each organization in the best possible light and compares only those organizations that are comparable to their size and input-output mix (peers) (Charnes et al 1981). DEA constructs a linear piecewise frontier. A frontier is the line that represents the best practice. DEA shapes the frontier by the maximum ratio of outputs to inputs in the sample. The frontier envelops all the observations in the sample. The organization(s) with the highest efficiency ratio lies on the frontier and this organization(s) is judged (technically) efficient. The organizations that lie below the frontier are judged (technically) inefficient. The efficiency score of each organization is a ratio where the nominator is the distance from the origin to the estimated point of the organization and the denominator is the distance of the origin to the frontier (Grosskopf & Valdmanis 1987; Charnes et al 1994). The assumptions underlying DEA needs consideration in order to determine which choices needs to be made to assess the efficiency of PCOs by DEA.

Assumption of the weights

The first important assumption is about the determination of the weights and the weight restrictions. The flexible weighting of DEA is an advantage above the CER, because the researcher does not need to specify weights a priori. However, this flexible weighting may in some situations lead to odd outcomes (Wong & Beasley 1990; Doyle & Green 1994; Jacobs et al 2006). In an extreme situation, an organization can be assigned efficient when it performs extremely well on one particular aspect, one input and output combination, and on all the other input and output combinations the organization is a bad performer. This extreme situation can occur since flexible weighting could attach a zero weight to some inputs and outputs if this yields the highest ratio for the organization (Doyle & Green 1994). For this reason, some authors have suggested ways of imposing restrictions (Allen et al 1997, Wong & Beasley 1990, Roll & Golany 1993). These weight restrictions can avoid such odd outcomes as mentioned above. However, the drawback of imposing weight restriction is that these restrictions do not rely on any economic theory (Jacobs et al 2006). Besides, a weight restriction requires specifying a priori some limits or weights which implies that the researcher should have an idea about the relative value of the inputs and outputs. A critique on applying weights restrictions to the data is that the pure objective of DEA gets lost (Roll & Golany 1993). DEA is a technique developed to measure the efficiency of organizations without the need to specify a priori the weights of the inputs and outputs. In this study, no

weight restrictions will be imposed in order to see what happens when DEA is applied in its pure objective.

Assumption of the operating scale

A second assumption that the researcher has to make relates to the production technology of the organizations under consideration. The operating scale, which is the relationship between the inputs and outputs of the production process, needs to be specified in order to run a DEA model. Two options are available to specify the operating scale, namely constant returns to scale (CRTS) and variable returns to scale (VRTS). A CRTS DEA model assumes that each additional unit of input results into the same additional proportional increase in a unit of output (Coelli et al 2005). With a VRTS DEA model, one assumes that each additional unit of input results into a varying additional increase or decrease in a unit of output (increasing or decreasing returns to scale). The relationship between CRTS and VRTS can be written as: $TE_{i,\,CRST} = TE_{I,VRST}$ x SE_I (Coelli et al 2005), where SE_i denotes the scale efficiency of the organization. The CRTS DEA model yields biased efficiency measures if the organizations under consideration are scale inefficient (Parking & Hollingsworth 1997). The choice between a CRTS and a VRTS model depends on the context of the analysis and the perspective of the study. It is legitimate to assume a CRTS DEA model when one measures the efficiency of organizations from the regulatory perspective (Smith 1997). This is because from a regulatory perspective - a societal perspective - one is interested into the long-term efficiency of organizations and not into the scale efficiency of organizations (Jacobs et al 2006).

Assumption of the orientation

A third assumption is the orientation of the analysis. There are two options: an input orientation or an output orientation (Charnes et al 1994; Coelli et al 2005). An input orientation focuses on reducing the inputs of the production process while holding the same production level. With an output orientation one wants to examine the extent to which the output can be increased with consumption of the same units of input. The choice between these two orientations is only relevant when one assumes a VRTS DEA model since the two orientations yields different efficiency measures under VRTS, but both orientations will yield the same efficiency measures under CRTS (Jacobs et al 2006). The choice between the two orientations depends on the assumption about the operating scale and the context of the analysis.

Assumption of slacks

The construction of a piecewise linear frontier has one inherent problem, which is the existence of slacks. A slack is the ability of an organization to improve the efficiency level by reducing the input quantity or increasing the output quantity while this organization lies on the estimated frontier (Coelli et al 2005). In other words, DEA judges this organization efficient while this organization indeed is inefficient. The existence of slacks is inherent to solving equation (2.0), because it is possible that solving the linear problem leads to a frontier that is for a part parallel to one of the axes. There is a debate about dealing with slacks in DEA (Cooper et al 2007). Some authors address the bias into the efficiency measures when slacks are present and propose alternatives for dealing with these slacks, Bessent et al 1988; Tofallis 2001, Ali & Seiford 1993, while others authors argue that we should use DEA, despite the inherent problem of slacks (Jacobs et al 2006; Coelli et al 2005). Coelli et al 2005 stated that slacks are the result of using finite samples since slacks would disappear with an infinite sample size. Regarding Coelli et al 2005 should the importance of the presence of slacks not be overestimated (Coelli et al 2005). There is no guideline about how the researcher should deal with slacks in an efficiency analysis by DEA, since it is an unresolved issue.

Besides the above assumptions, some other important properties of DEA need attention. First, DEA is a deterministic technique that implies that noise in the data is ignored. The efficiency measures derived by deterministic techniques do not distinguish between the inefficiency of the organization and measurement error when measurement error is (expected to be) present (Jacobs et al 2006). Second, the efficiency measures are very sensitive to sample size and outliers (Smith 1997). The inclusion or exclusion of organizations in the sample, the inclusion or exclusion of relevant variables or atypical input and output combinations can influence the efficiency measures, since DEA constructs the frontier by the observed data (Smith 1997). The model specification is highly important for deriving appropriate efficiency measures. A drawback of DEA is that no diagnostic test is available to determine whether the model is specified correctly, which is inherent to deterministic techniques (Smith 1997; Parkin & Hollingsworth 1997). Specification error is a serious problem when small sample sizes and (too) simple models are used for efficiency assessment of complex multiple input-multiple output production processes (Smith 1997).

§ 2.2 Parametric efficiency assessment methodologies

The use of a parametric technique for efficiency assessment requires parameterization of the production process. These techniques rely on econometric estimation of a model. The researcher has to make some assumptions in order to estimate a parametric model for

efficiency assessment. There are two main assumptions underlie efficiency assessment by parametric techniques, these are an assumption about the function and one about the functional form.

Assumption of the function and functional form

The first assumption that the researcher has to make is the choice of the function. The researcher can choice either estimation of a production function or estimation of a cost function. Related to the assumption about the function is the assumption about the functional form. The functional form describes the relationship between the inputs and outputs. This study analyzes the efficiency of PCO through production functions. Production functions are used when no cost information is available or when it is not appropriate to assume cost-minimizing behavior (Jacobs et al 2006). In the datasets of this study, no cost information is available and therefore this study estimates production functions. The cost function will not be discussed in detail here. The drawback of the use of production functions is that multiple outputs needs to be aggregated into a single output, without the researcher can rely on a convenient weight such as market prices (Jacobs et al 2006; Coelli et al 2005).

A production function determines the maximum feasible operating scale that maximizes the output of the organization (Coelli et al 2005) and the econometric production model takes the form,

$$y_i = F(x_i; \alpha, ß) \qquad i = 1,2,\ldots N \qquad (3.0)$$

The y_i denotes the output in quantities of the i^{th} organization in the sample; x is the inputs; α and ß are the parameters to estimate and F(.) is the function that describes the relationship between the inputs and outputs. There are different functions available, but the Cobb-Douglas function is used most frequently in combination with a production function (Coelli et al 2005). A production function with logs of the variables, a Cobb-Douglas function, is usual linear in the parameters that leads to desirable statistical properties (Greene 2005). Other functions used frequently are the linear and translog functions (Coelli et al 2005). The choice of the functional form F(.) is important since it can yield different efficiency measures (Jacobs et al 2006). Coelli et al 2005 give some criteria for choosing the functional form. In general, one prefer those functional forms that are flexible, linear in parameters, regular (satisfying the statistical properties) and parsimonious (simplest model that is adequate) (Coelli et al 2005). If the variables in the model are expressed in their natural units, thus without transformation, then the researcher assumes a linear relationship between the inputs and outputs (Jacobs et al 2006). The appropriateness of the assumption of linearity depends on the context of the

analysis.

§ 2.2.1 Corrected Ordinary Least Squares

COLS is a technique that relies on econometric estimation of a function described in the section above. COLS estimates the model by Ordinary Least Squares (OLS) implying the COLS model having some similar statistical properties as usual econometric estimation by OLS. Those OLS properties are only of interest if one is concerned with the estimation of the parameters α and β. However, for efficiency assessment, you may also be interested in the estimation of the parameters α and β, lays a strong focus on the estimation of the residual ε. In this case, the OLS properties are not of main interest. The COLS production model takes the form:

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad i = 1,2,..., N \qquad (4.0)$$

, where y_i present the output of the production process for the i^{th} organization; α and β are the parameters that will be estimated; x present the input of the i^{th} organization; and ε is the residual. The residual is the unexplained variation of the performance of organizations that is used to derive the efficiency level of the organization. COLS derives the efficiency of the organization by first lifting the estimated regression line by OLS with the maximum residual in the sample. The lifted regression line denotes the frontier. Then, the inefficiency of the organization is the deviation of the observation to the theoretical frontier. The efficiency score of the organization is derived by,

$$TE_{i,\,COLS} = (\hat{y}_i + \varepsilon_i) / (\hat{y}_i + \max(\varepsilon)) \qquad i = 1,2,\ldots,N$$
$$0 < TE_{i,\,COLS} \leq 1$$

, where $TE_{i,\,COLS}$ is the technical efficiency score of the i^{th} organization in the sample; $\hat{y}_i$ is the estimated point of the i^{th} organization; ε_i is the estimated residual of the i^{th} organization and max(ε) is the maximum error in the sample.

The advantage of COLS is that it is the simplest parametric technique. However, the researcher should be aware of the underlying assumptions. The first assumption is that the entire residual of the organization is attributable to inefficiency of the organization and any other influence or other factor, such as random error or measurement error is ignored (Greene 2005). Just as other deterministic techniques, the limitation is likewise that the efficiency measures may be biased by measurement error.

§ 2.2.2 Stochastic Frontier Analysis with cross-sectional data

SFA is a stochastic frontier technique developed to account for random error and noise in estimation of the efficiency of organizations. SFA relaxes the assumption about stochastic underlying the deterministic techniques. The accompanying effect is that SFA becomes much more complex than the earlier mentioned efficiency assessment techniques. There are several SFA models available. Let us start with the simplest SFA model. The SFA production model for cross-sectional data is:

$$y_i = \alpha + \beta x_i + \nu_i - \upsilon_i \qquad i = 1,2,\ldots, N \qquad (6.0)$$

This is the same function as with COLS, but the only difference with COLS is that SFA splits up the residual ($\varepsilon_i = \nu_i - \upsilon_i$). The υ_i is a non-negative term interpreted as the inefficiency component. The ν_i is the random component.

The decomposition of the residual likewise corresponds with making an assumption. First, the researcher needs to specify a distribution for the inefficiency component and the random component in order to distinguish them econometrically (Jacobs et al 2006). To be consistent with classical OLS models is in the SFA cross-sectional model assumed that the random component is normally distributed. Several different distributions are available for the inefficiency component, such as the half-normal, the truncated normal, the exponential and the gamma distribution (Greene 2002), each yielding different efficiency measures (Jacobs et al 2006). There are no criteria available to guide the researcher which distribution should be assumed for the inefficiency component (Jacobs et al 2006). The drawback of the SFA model for cross-sectional data is that the assumption about the error distribution is unavoidable and may influence the obtained efficiency measures.

The inefficiency component should be estimated indirectly, because OLS will only estimate the residual ε. The partition of the residual will be done in a next step of the analysis on basis of the expected value of the inefficiency component conditioned on the residual (Jacobs et al 2006). The formula for this conditional expectation of the inefficiency component depends on the distribution chosen for the inefficiency component[5]. All these formulas are inconsistent estimators of the inefficiency component (Greene 2003). This implies that the efficiency measures obtained by a cross-sectional SFA model, no matter which distribution for the inefficiency component is chosen, is only an approximation of the actual efficiency of an organization (Jacobs et al 2006). The technical efficiency of the organization is derived by:

[5] Jacobs et al 2006 provides an overview of the formulas of the different models.

$TE_{i,SFA} = (\beta x_i - \upsilon_i) / (\beta x_i)$ $0 < TE_{i,\,SFA} \leq 1$

, where the ß is the coefficient of the estimated parameter. This formula only accounts when the variables are expressed into their natural units of measurement, thus without transformation.

§ 2.2.3 Stochastic Frontier Analysis with panel data

The use of panel data is in econometric analysis is in general preferred to the use of cross-sectional data. Panel data allows the researcher to correct for unobserved heterogeneity that is due to unobservable factors. These unobservable factors are 'picked up' into the estimation when cross-sectional data is used and it is unknown to the researcher whether these unobservable factors are captured into the residual and/or any other estimated parameter. The ability to correct for unobservable heterogeneity is an attractive property of the standard panel model (Greene 2003; Greene 2005).

The SFA production model for panel data takes the form:

$$y_{it} = \alpha + \beta x_{it} + \nu_{it} - \upsilon_i \qquad i = 1, 2, \ldots, N \qquad t = 1, 2, \ldots, T \qquad (6.0)$$

The interpretation of this function is similar to the SFA function for cross-sectional data, only this function specifies that the i^{th} organization of N organizations in the sample in time period t is followed over T periods. Two approaches can be adopted when estimating a SFA model with panel data, namely a time-invariant approach or a time-varying approach.

Time-invariant model

The time-invariant approach implies assuming that the efficiency level of the organization does not vary over time. The estimation of the function with this approach relies on fixed effects or random effects models (Greene 2005). The fixed effects (FE) model assumes that the efficiency level of the organization is captured into the firm specific constant, $\alpha_i = \alpha - \upsilon_i$. This model is estimated by within group least squares estimation. The FE approach is equivalent to a dummy variable regression, which means adding a dummy for all, except one, organization in the sample. The constant of the regression presents the constant of the remaining organization (the organization not added as a dummy variable). The constants of the other organizations are derived by the mean within group deviation of y_{it} and the estimated coefficients, the ß's, which is used to obtain the relative efficiency of each organization. The method to establish the efficiency estimates from a FE model is analogous to the COLS

method, only with a FE model the frontier is the regression line lifted by the maximum intercept α if a production function is estimated (Greene 2005). The advantage of the FE model is that it is free from the distributional assumption about the inefficiency component. The FE model also allows correlation between explanatory variables and the inefficiency component. However, the FE SFA model might lose the attractive property of correcting for unobservable heterogeneity. Unobservable firm specific characteristics that are not related to the efficiency level of the organization, still can be picked up in the estimation (Greene 2005; Polachek & Yoon 1996) and thus in the organizational specific constant that is used to derive the inefficiency of organizations. This may lead to a situation where the efficiency measures may be a mixture of unobservable organization-specific effects and the inefficiency of the organization (Polachek & Yoon 1996; Greene 2004). This requires making the same assumption as with the SFA model for cross-sectional data, namely all relevant factors have been observed and the model is correctly specified.

A random effect (RE) model assumes that the organizational specific effects are random draws from the population and therefore do not use the intercept α to extract the efficiency of the organizations. The random effects model estimates the inefficiency component just as the cross-sectional SFA model separately. This leads to that the RE model requires making an assumption about the distribution of the inefficiency component. In addition, the RE model requires a tighter parameterization than the FE model, namely the RE model assumes that the explanatory variables included in the model is uncorrelated with the inefficiency component (Greene 2005).

If long time series are used then a shortcoming of the FE as well as the RE model is the assumption of time invariance (Greene 2005). It is unlikely that the efficiency level of organization does not vary over long periods. In that case, it may be more appropriate to use a time varying approach.

Time-varying model

A time varying approach assumes that the efficiency level of the organizations changes over time. The time-varying model estimates an additional parameter, namely the residual will be decomposed into an inefficiency component, an industry-wide component and the random error component. There are several methods to estimate the industry-wide parameter (Lee & Smidt 1993, Kumbhakar 1990 and Battese & Coelli 1988). The time-varying model is the most complicated SFA model. This model will not be discussed in detail here, since this model will not be applied in this study. This study uses data with time-series of ‘only’ three years and it goes beyond the limits of this study to develop a time-varying SFA model with a

binary output measure on individual level.

The question that arises after discussing the four efficiency measurement methodologies in the above sections is which methodology the health care insurer should use for efficiency assessment. There is no consensus on the 'best method' for measuring the efficiency of (health care) organizations. Comparison of different techniques can provide useful insight into the relative merits of the assumptions underlying the different techniques. The next section describes the method for comparison of the four efficiency assessment techniques in this study and presents the comparative efficiency analysis of this study.

Section 3: Comparative efficiency analysis

A comparative efficiency analysis is a useful method to examine the relative merits of several techniques in order to narrow the uncertainty about the extent to which the assumptions underlying those techniques biases the results (Chirikos & Sear 2000). This study uses a comparative efficiency analysis to determine the role of the assumptions underlying each methodology. The comparative analysis in this study focuses on the main assumptions underlying the four techniques and compares the results of the different techniques as follows. Section 3.2 will make a comparison of the results from the nonparametric techniques, the CER and DEA. This section examines the role of the weights in deriving the efficiency measures. Both techniques share the assumption about stochastic. Section 3.4 compares the results from the parametric techniques, the COLS and SFA. This section examines the role of stochastic. Section 3.5 compares and discusses the results from DEA and SFA. This section focuses on a comparison of the two frequently used efficiency assessment techniques. The role of the weights, stochastic and the function and functional form will be examined here. Section 3.1 and 3.3 provides the specification of the model for each efficiency assessment technique. It goes beyond the limits of this master thesis to review every model that could be applied to examine the efficiency of PCOs. The sensitivity analysis in this study is therefore restricted to a comparison of the efficiency measurement techniques by four models, a model for each efficiency assessment technique. Those models are selected that are parsimonious and (as most as possible) similar to each other. The sensitivity of the results to the specification of the conceptual model will be examined in the discussion of this study. For the analysis of this study, the software packages STATA and Efficiency Measurement System (EMS) are used. EMS is a specific software package for DEA used to derive the weights chosen by DEA.

Comparison of several different techniques can be quite difficult when there is no consensus or guidance. It is unlikely that the different techniques applied in this study yields consensus, so that the efficiency measures are not sensitive to the choice of technique. Bauer et al 1998 argue that the following set of conditions should be met when efficiency measures will be used for explanatory or predictive purposes. “The efficiency estimates should be consistent in their efficiency levels (comparable means, standard deviations, and other distributions), consistent in their ranking, consistent in their identification of the best and worst performers and consistent over time” (Bauer et al 1998). The efficiency measures are useful for contracting purposes if these consistency conditions are satisfied. The use of efficiency measures for contracting health care organizations can have a large (financial) impact on the practice of health care organizations. Therefore, the health care insurer should make such contracting decisions only when the efficiency measures reflect the actual difference in

efficiency level across organizations rather than the analytical choices made in deriving those efficiency measures.

§ 3.1 Model specification nonparametric techniques

The first question that arises is how the efficiency of PCOs can be assessed by a CER and DEA. The CER and DEA are techniques that rely on an index approach, which is a simple ratio of output(s) to input(s). The CER and DEA examine the efficiency of organizations on organizational level. This has the following consequences for the specification of the variables in the conceptual model.

First, the output measures in this study are rates on organizational level. The score on the output (O_EFF) in Table 2 needs to be adjusted for case-mix differences across the PCOs in order to be able to compare the efficiency level of the PCOs. This leads to defining the output (O_EFF) as a case-mix indicator index, which is the ratio of the actual indicator score – as in Table 2 – to the expected indicator score on basis of the case-mix of the PCO. The expected indicator score is derived by first estimation of a regression for each quality indicator (dependent variable) separately with the case-mix variables (independent variables) on individual level. Then, the individual expected indicator scores are aggregated to organizational level that leads to the expected indicator score for each PCO. A score of '1' on the case-mix indicator index means that the PCO performs as expected on basis of their case-mix, a score of '<1' denotes that the PCO performs worse than expected and a score of '>1' stands for a better performance than expected. In order to express all the output measures in the same unit of measurement are the outputs (O_EFFEC) and (O_EP) also expressed as indices, whereby the nominator is the actual indicator score on organizational level and the denominator the average indicator score in the sample (Table 2). A score of '1' denotes an average performer, a score of '<1' stands for a more worse than average performer and a score of '>1' denotes a better than average performer. The input measures are defined as the 'average weighted number of activities per patient per year' as presented in Table 2.

Second, although the data of this study consists of panel data, cross-sectional data are used for specification of the nonparametric models instead. Year-specific models are estimated since the use of panel data in an index approach requires deflating the inputs and outputs over years to a base year (Coelli et al 2005). There is no appropriate deflator available, such as the national growth of the quality of care or other changed circumstances in the operational environment of PCOs. This leads to estimation of cross-sectional models.

Third, the CER requires a priori specification of the weights. The CER here attaches a weight of '1' to the all inputs and outputs. This is the simplest imaginable CER, since the researcher makes no judgment about the preferred input-output mix. The reason behind this is that comparison of the results from the CER and the results from DEA indicates the largest imaginable differences between no judgment about a preferred input-output mix and the preferred input-output mix that suits best regarding the observed data. The CER results will converge to the DEA results if the a priori chosen weights come closer to the weights applied by DEA.

Fourth, several DEA models could be applied to analyze organizational efficiency levels. This study applies only one of all the DEA models. This study examines a CRTS DEA model with an output-orientation. From the regulatory perspective, it is preferred to examine a CRTS model (Jacobs et al 2006). Moreover, a CRTS model is most similar to the parametric models applied in this study that are linear in the parameters. The output orientation is used frequently for estimation of production functions (Coelli et al 2005).

Fifth, some PCOs have a zero value for some inputs in the data of this study (Table 2). DEA does not accept zero input values in the data. The zero input values in the data therefore are changed into a value of 0.01 in order to be able to estimate a DEA model. Notice that the PCOs with the zero input values or the two input variables could be excluded. This option is not chosen, because that would mean a (large) loss of information due to exclusion of some PCOs in the sample and exclusion of two relevant inputs may yield misleading results, whereby the PCOs are benefited that consumes more units of those two inputs than another PCO, such as PCO 5. Appendix 4 reports the descriptive statistics of the data for the nonparametric techniques.

§ 3.2 Results nonparametric techniques: the CER and DEA

Table 4 reports the results of CER and DEA. The columns show the efficiency measure and their corresponding rank of the PCOs. The last two columns in Table 4 represent the weights that DEA has used to obtain the efficiency measures. Respectively those weights correspond to the inputs (AC_GP), (AC_GPA) and (AC_GPO) and to the outputs (O_EFF), (O_EFFEC) and (O_EP). An assumption made in the derivation of the cross-sectional efficiency scores is to not adjust the output measure for changes over years, such as changes in the circumstances across the PCOs. A panel of three years is likely to be short for major shifts in the operational environment of PCOs (Parking & Hollingsworth 1997), however still there may be temporal changes. This leads to careful drawing conclusions on cross-sectional results over time.

Table 4: results from the CER and DEA

	DMU	CER		DEA		Input weights			Output weights		
		TE_{CER}	rank	TE_{DEA}	rank	v_1	v_2	v_3	u_1	u_2	u_3
year 2007	**1**	0.045	2	1.000	1	0,01	0,08	0,00	0,00	0,00	0,81
	2	0.042	3	1.000	1	0,00	100,00$^{\Theta}$	69,09$^{\Theta}$	0,00	0,00	1,00
	3	0.031	8	1.000	1	0,00	100,00$^{\Theta}$	0,00	0,94	0,00	0,00
	4	0.053	1	1.000	1	0,02	0,00	0,10$^{\Theta}$	0,00	0,00	0,67
	5	0.032	7	0.983	6	0,02	0,00	0,00	1,03	0,00	0,00
	6	0.038	4	1.000	1	0,01	0,03	0,03	0,96	0,00	0,00
	7	0.028	9	0.677	9	0,01	0,02	0,00	0,00	0,89	0,00
	8	0.036	5	0.870	7	0,01	0,00	29,26$^{\Theta}$	0,00	0,97	0,00
	9	0.033	6	0.762	8	0,01	0,03	0,00	0,00	0,93	0,00
	mean	0.038	-	0.921	-	-	-	-	-	-	-
year 2008	**1**	0.021	9	0.424	9	0,01	0,00	0,00	0,00	0,00	1,02
	2	0.044	3	1.000	1	0,00	2,44	0,00	0,00	0,00	0,84
	3	0.030	7	1.000	1	0,01	0,07	45,52$^{\Theta}$	0,26	0,49	0,25
	4	0.029	8	1.000	1	0,00	0,11	72,36$^{\Theta}$	0,87	0,00	0,00
	5	0.037	5	0.706	7	0,02	0,00	0,00	0,00	0,98	0,00
	6	0.043	4	0.912	6	0,02	0,00	0,00	0,00	0,00	0,86
	7	0.036	6	0.659	8	0,01	0,00	0,00	0,00	0,00	0,93
	8	0.058	2	1.000	1	0,01	0,00	52,17$^{\Theta}$	0,93	0,00	0,00
	9	0.062	1	1.000	1	0,00	1,98	7,01$^{\Theta}$	0,00	0,99	0,00
	mean	0.040	-	0.856	-	-	-	-	-	-	-
year 2009	**1**	0.031	8	0.999	7	0,00	0,17	0,32	0,00	1,01	0,00
	2	0.042	2	1.000	1	0,00	2,04	0,00	0,00	0,00	0,83
	3	0.036	5	0.871	8	0,01	0,01	0,00	0,91	0,00	0,00
	4	0.042	2	1.000	1	0,00	0,00	100,00$^{\Theta}$	0,87	0,08	0,00
	5	0.036	5	1.000	1	0,01	0,00	0,09	0,00	0,17	0,78
	6	0.036	5	1.000	1	0,01	0,00	0,13	0,00	0,00	0,88
	7	0.030	9	0.808	9	0,01	0,02	0,14	0,00	0,00	0,93
	8	0.047	1	1.000	1	0,02	0,00	0,00	0,00	0,53	0,37
	9	0.041	4	1.000	1	0,01	0,05	0,03	0,00	0,95	0,00
	mean	0.038	-	0.964	-	-	-	-	-	-	-

† standard deviation in parentheses.

$^{\Theta}$ The corresponding observation has a zero value (Table 2).

Note: There are no slacks present for the PCOs judged efficient by DEA.

The CER and DEA results presented in Table 4 will be analyzed by considering the questions: what is the sensitivity of the results to the assumptions underlying those techniques and what is/are the reason(s) for this? The CER and DEA results will be compared by using their ranking. This is because the CERs are organizational specific indices whereas the efficiency measures by DEA are relative efficiency scores. If the results of the CER and DEA differ considerably, then it is expected that these differences primarily are affected due to the assumption about the weights.

Jacobs et al 2006 describes sorting the organizations into three groups when the researcher compares efficiency scores from different techniques (Jacobs et al 2006). The first group consists of organizations where the efficiency score is sensitive to the choice of technique. It is inadvisable to draw conclusions about the relative efficiency of those organizations (Jacobs et al 2006). Most of the PCOs in Table 4 would belong to this group, namely all PCOs over all the years except for PCO 4 and PCO 7 in 2007, PCO 1 and PCO 9 in 2008 and PCO 7 and PCO 8 in 2009. In the second group are organizations consisted that appearing to be efficient irrespective of the technique adopted or the model specified (Jacobs et al 2006). Those organizations seem to be efficient, however further analysis can be informative to identify which organization is indeed the 'best practice'. Identification of the organizations within this group requires attention, since the ranking can be very sensitive to specification of the model. According Table 4, PCO 4 in 2007, PCO 9 in 2008 and PCO 8 in 2009 would belong to this second group. The third group of organizations consists of organizations that appear to be inefficient whichever technique is applied (Jacobs et al 2006). Those organizations seem to be inefficient, and require further analysis to ascertain the reasons for their relative worse performance. In Table 4, only PCO 7 in 2007 and 2009 and PCO 1 in 2008 is the worst performer in both applications of CER and DEA. However, PCO 1 is efficient (or the second best performer by CER) in 2007. For PCO 4 in 2008, the extreme difference occurs in which a technique assigns a PCO efficient while the other technique judges the same PCO (one of the most) inefficient. To conclude, not all the consistency conditions by Bauer et al are satisfied. There are some similarities in the ranking of PCOs by CER and DEA, but there is no consistency in the ranking over time. The Spearman rank correlation coefficient show how close the rankings of PCOs are between two methods. The Spearman rank correlation coefficient for the results between the CER and DEA are $r_s = 0.6390$ for results in 2007, $r_s = 0.475$ for results in 2008 and $r_s = 0.749$ for results in 2009. Those correlation coefficients determine a weak correlation between the results from the CER and DEA in 2007 and 2008 whereas the two techniques yield in 2009 more consistent results. The data used in the analysis can be the reason for the large differences in the correlation coefficient across the years. Those correlation coefficients do not only show that the choice of technique is important, but that the data used for the analysis is just as, if not even more, important to derive the efficiency measures.

The following point can be made regarding the appropriateness of the weights used in the analysis to derive the results in Table 4. Variation in the inputs and outputs (Table 2) indicates that some of the PCOs are inefficient, but some of the variation might also be due to the operational environment of the PCOs. If one health care provider is more efficient in attaining a health effect than another health care provider, then is it not appropriate to attach the same

weight to all the inputs in this study. The same accounts for the outputs, if it is easier to attain a higher score on of the indicators, then the outputs should weighted differently. The weights DEA has used presented in Table 4 largely differ that suggest that it may not be appropriate to attach the same weight to all the inputs and outputs according to the observed data. Moreover, more PCOs are assigned efficient by DEA (Table 4) that also might indicate that some PCOs have another input-output mix, because DEA compares only those organizations that are comparable to their input-output mix. It is the question if the PCOs can adopt the input-output mix of any PCO in the sample. As long as this is unclear, one should be aware that the assumption underlying the CER might be too restrictive within the context of the analysis. Moreover, one can also have doubts whether the flexible weights by DEA represent the actual relative value of the multiple inputs and outputs (Table 4). DEA has attached many zero weights implying that DEA has judged a PCO efficient, because it performs well on one particular output and input mix, while the other inputs and outputs are ignored. Notice that observations with a zero input value has received a (extremely) high weight, such as a value of 100. The change of the zero inputs into a value of 0.01 has influenced the results, but it is a problem stemming from the data used when assessing the efficiency of PCOs by DEA.

Some other points that can be made regarding the sensitivity of the CER and DEA results in Table 4. The ranking of the CERs and the efficiency measures of DEA are sensitive to the presence of other PCOs in the sample. A difference of 0.003 in the CER determines if a PCO is ranked 2nd or 3rd (PCO 1 and PCO 2 in 2007). A difference of 0.001 in the efficiency measure of PCO 1 and PCO2 in 2009 by DEA determines if the PCO is ranked 7th or 1st. Exclusion of one of the PCOs yields different efficiency measures and ranks. In addition, there are relative many efficient PCOs assigned by DEA, respectively 5 PCOs (54%), 5 PCOs (54%) and 6 PCOs (67%) of the 9 PCOs in the sample are efficient. This is one of the drawbacks of using DEA for efficiency assessment when the sample size is small. The larger the fraction of the PCOs assigned efficient, the less discriminative power the analysis will have with the consequence that the efficiency assessment might be less informative. The usefulness of the results from DEA therefore may depend on the sample size and the comparability of the PCOs in the sample. Moreover, it is unclear whether the observed differences across the PCOs are significant. Confidence intervals can help to determine the significance of the observed differences in the efficiency measures, but confidence intervals are not applied in this study. This leads to careful drawing conclusions on the observed differences across the PCOs.

There are two other sources of bias into both the CER results and the DEA results in this study relating to other assumptions underlying those techniques than the assumption about the

weights. First, both techniques assume constant returns to scale. Misspecification of this assumption is most serious when the input variables have a low correlation (Smith 1997) and therefore may be expected that the returns to scale assumption largely have influenced the obtained results in Table 4 (Table 5). The derived results by CER and DEA can be biased by the scale efficiency of PCOs.

Table 5: correlation input measures

	AC_GP	**AC_GPA**	**AC_GPO**
AC_GP	1.000		
AC_GPA	0.000	1.000	
AC_GPO	0.330	-0.145	1.000

Another source of bias is into the results of the CER and DEA relates to the assumption about stochastic. In this study, and in general in efficiency assessment in the health care sector (Jacobs et al 2006; Coelli et al 2005), is it unlikely that all the relevant factors are observed and measured without error. Measurement error into one of the variables can have a large impact into the results, because the results are very sensitive to the efficiency score of the other PCOs in the sample. The role of stochastic into the results needs to be examined in order to establish whether the assumption underlying deterministic technique is appropriate within the context of the analysis.

Conclusion

The results in this section show that not all consistency conditions of Bauer et al are satisfied. This leads to concluding that the health care insurer should be cautious with using the obtained results from the CER and DEA for contracting health care organizations, because the results of the CER and DEA are sensitive to the assumptions underlying the technique and the data used to analyze the efficiency of PCOs. The ranking is used to judge whether a PCO is efficient. However, it is questionable if the observed differences in this section are significant. The results are very sensitive to the presence of other PCOs and the difference across some of the PCOs is (very) small. Measurement error into one of the variables in this study can therefore have a large impact on the relative efficiency of those PCOs. The next section will focus on the assumption about stochastic, in order to determine in what extent the results in this section are biased.

§ 3.3 Model specification parametric techniques

This section describes the specification of the parametric models used in this study. Here, the parametric models will be estimated on individual panel data. The advantage of parametric techniques above nonparametric techniques is that panel data can be used more easily,

because parametric techniques do not rely on an index approach where it is necessary to deflate the variables to a base year. The two parametric techniques in this study both use panel data. The panel in this study is unbalanced; not every individual has an observation for each year. According to the definition in literature, one should not label a COLS or SFA model as a COLS or SFA model when it is applied to individual data, despite the fact that the model may have the same properties. To be analogous to the definitions in literature, this study will refer to the COLS model by a deterministic parametric technique and the SFA model by a stochastic parametric technique.

In this study, estimation the efficiency of PCOs on individual level data leads to a binary output measure. To the best of my knowledge is such an output measure never been used before in an efficiency assessment. The deterministic parametric approach for efficiency assessment in this study uses a logistic regression. This leads to estimation of a logistic production function whereby the variables are defined into their natural units of measurement. The assumed functional form F(.) is the logistic cumulative distribution function. Hence, this leads to a deterministic parametric model formulated as:

$$y_{it}^{*} = \alpha + \beta_m x_{mit} + \beta S_{it} + \varepsilon_{it} \qquad i = 1, 2,\ldots,n \quad \text{(individual)} \qquad (7.0)$$

$$t = 1, 2,\ldots,T$$

$$m = 1, 2,\ldots,M$$

$y_i^{*} = 1$ if $y_i \geq 1$ (the nominator of the quality indicator is 'yes')

$y_i^{*} = 0$ if $y_i = 0$ (the nominator of the quality indicator is 'no')

The y_i* denotes the score on the quality indicator; x is the m^{th} input of M inputs defined as 'the weighted number of activities' per individual in the data; the case mix variables denoted by S are regressed against the output variable. The 'i' denotes the i^{th} unique individual in a specific year of n unique individuals in a specific year. The deterministic parametric model uses the residual ε_{it} to extract the efficiency score. The residual for each individual is aggregated to the organizational level whereby the average residual per PCO is used to derive the efficiency of that PCO. This method is comparable to the 'standard' COLS approach, since the residual in equation (7.0) is used to extract the organizational efficiency. The PCO with the highest average residual serves as frontier and all the other PCOs are compared to this frontier. The limitation of the parametric model used in this study requires specifying a single output, y_i, instead of multiple outputs as defined in the conceptual model of this study. Therefore, both parametric models estimate the efficiency of PCO for each quality indicator separately. Then, the average score on the separate indicators within each output is the efficiency score on this output. The efficiency of a PCO is the average score of the three

outputs.

The stochastic parametric model estimates the efficiency of PCOs by a dummy variable logistic regression. The stochastic parametric model in this study takes the form:

$$y_{it}^{*} = \alpha + \beta_m x_{mit} + \beta S_{it} + \beta_I D_I + \nu_{it} \qquad (8.0)$$

$m = 1,2,\ldots,M$

$i = 1,2,\ldots,n$ (individual)

$I = 1,2,\ldots,N$ (organization)

$t = 1,2,\ldots,T$

, where just as in equation (7.0) the y_{it}^{*} denotes the score on the quality indicator; the x the input variables and the S the case-mix variables. The D stands for the dummies for all, except one, PCO. The ν_{it} is the random error component. The intercept α is the constant of the remaining PCO (not included with dummies). The intercept, α, and the β-coefficients of the organization-dummies are used to derive the organizational specific constant: $\alpha_I = \alpha + \beta_I$ is the constant for the I^{th} PCO of N PCOs in the sample. The inefficiency of a PCO is obtained by, $\upsilon_I = \max(\alpha_I) - \alpha_I$ and then the efficiency score is:

$$TE_{I,SFA} = (\max(\alpha_I) - \upsilon_I) / \max(\alpha_I) \qquad \text{,with } 0 < TE_{I,\,SFA} \leq 1.$$

An approach, comparable to the FE SFA approach (Section 2.2.3), is used for the efficiency assessment in this study, since this approach does not require specification of distributional assumptions nor assuming that the inefficiency component is uncorrelated with the input variables. The drawback of the use of this approach is that it requires existence of sufficient variation in the explanatory variables in the data for some individuals or over time. The variables that do not vary across individuals and over time are automatically omitted from the estimation procedure. In this study, the observations with zero-input values are omitted from the estimation, because for all the individuals within this PCO the explanatory variable does not vary. In that case, the estimation is based solely on the remaining input variables. With a approach comparable to SFA FE, there may also not be a perfect linear relationships with the explanatory variables. In this study a perfect linear relationship occurs when a PCO produces an indicator score of '100'. This PCO in that case is omitted from the analysis. The assumption made here is that the PCO is efficient on the indicators where it produces a score of '100', thus the PCO has received an efficiency score of '1'. PCO 4 has benefited most of this assumption, since PCO 4 produces relative to the other PCOs more times an indicator-score of '100' (Appendix 5).

Table 6 reports the estimated parameters that are used to derive the efficiency of the PCOs. Most of the parameters are not significant and the coefficient of the constant is large compared to the other coefficients, which may indicate the presence of unobservable factors into the estimated parameter(s). The advantage of parametric techniques is that statistical tests can be used to improve the model specification, however, this study assesses the efficiency of PCOs by using the model as specified in Section 1. The coefficients presented in Table 6 are used to derive the efficiency score of the PCOs.

Table 6: Estimated parameters

	Indicators output (O_EFF)							Output
variables	DM_1 Coef.	DM_2 Coef.	COPD_1 Coef.	COPD_2 Coef.	CVR_1 Coef.	CVR_2 Coef.	COPD_3 Coef.	(O_EP) Coef.
AC_GP	- 0.001	0.004	- 0.008	0.034	0.007♠	0.003	0.000	0.011
AC_GPA	0.005	- 0.025	0.020†	(omitted)	- 0.007	- 0.009	-0.001	- 0.003
AC_GPO	0.013	0.023	(omitted)	(omitted)	0.045♠	- 0.015	0.056♠	0.012
FEMALE	0.466	0.792	1.080†	(omitted)	0.020	- 0.145	⊖	⊖
AGE	0.023†	0.053†	0.037†	0.077	0.031♠	0.034♠	⊖	⊖
COMOB	1.404♠	1.852♠	- 0.607	(omitted)	0.463	- 0.010	⊖	⊖
POLYF	0.431	- 0.873	0.571	(omitted)	0.704	- 1.319	⊖	⊖
PCO 1	0.999	0.207	1.430	(omitted)	0.544	- 0.453	0.784	0.606
PCO 2	1.062	0.611	0.533	(omitted)	1.771	- 0.102	0.423	0.175
PCO 3	- 0.597	(omitted)	- 0.171	(omitted)	- 0.002	- 0.697	0.112	- 0.014
PCO 4	(omitted)	- 0.924	(omitted)	(omitted)	1.440	- 0.058	-0.055	- 0.054
PCO 5	- 0.045	0.347	- 0.919	(omitted)	0.851	0.028	-0.105	0.132
PCO 6	0.818	- 0.552	- 1.717	(omitted)	0.963	- 0.225	0.484	- 0.049
PCO 7	- 0.766	- 0.909	0.508	(omitted)	1.097	- 0.330	0.528	0.357
PCO 8	1.064	- 0.174	- 0.650	(omitted)	1.413	- 0.031	0.280	0.226
cons	- 5.922	- 8.653	- 6.307	- 15.810	- 7.050	- 4.953	-3.149	- 2.486

	Indicators output (O_EFFEC)							
variables	DM_3 Coef.	DM_4 Coef.	DM_5 Coef.	DM_6 Coef.	DM_7 Coef.	DM_8 Coef.	CVR_3 Coef.	COPD_4 Coef.
AC_GP	0.001	- 0.005	- 0.002	-0.014	0.001	0.005	0.001	(omitted)
AC_GPA	0.002	- 0.006	- 0.007	(omitted)	- 0.011	- 0.005	0.001	(omitted)
AC_GPO	- 0.014	0.005	0.003	(omitted)	0.000	0.003	- 0.002	(omitted)
PCO 1	0.137	16.294	0.053	(omitted)	- 0.594	- 1.967	0.039	(omitted)
PCO 2	- 0.603	15.772	- 0.440	14.233	- 0.400	- 1.785	- 0.291	(omitted)
PCO 3	- 0.594	15.075	- 0.798	(omitted)	- 0.945	- 0.878	0.293	(omitted)
PCO 4	- 0.661	15.881	- 0.091	(omitted)	- 0.725	- 1.195	0.348	(omitted)
PCO 5	0.325	13.665	- 1.124	(omitted)	- 2.499	(omitted)	0.267	(omitted)
PCO 6	- 0.294	15.864	- 0.168	(omitted)	- 0.326	- 1.702	0.031	(omitted)
PCO 7	- 0.406	16.045	- 0.103	14.376	- 0.364	- 0.598	0.319	(omitted)
PCO 8	0.141	16.382	0.327	(omitted)	- 0.178	- 0.905	0.085	(omitted)
_cons	- 1.543	-17.453	- 1.687	-17.709	- 1.707	- 2.028	- 0.984	(omitted)

† significant at 0.05 significance level

♠ significant at 0.01 significance level

⊖ adjusting indicator-score for case-mix differences not required (ZN 2008).

§ 3.4 Results parametric techniques

This section focuses on sensitivity of the results to the assumptions underlying the parametric techniques, which are the assumption about stochastic and the assumption about the function and functional form. The two parametric models used here have not been applied before and therefore the model specification of the parametric techniques needs to be discussed in order to examine the validity of the obtained efficiency measures. Table 7 presents the efficiency measures by the deterministic parametric technique and the stochastic deterministic

technique. The column 'mean TE' represents the average efficiency score on the three outputs. The ranking is based on this 'mean TE'.

Table 7: efficiency scores derived from two parametric techniques

PCO	Deterministic parametric model		Stochastic parametric model	
	mean TE	rank	mean TE	rank
1	0.547	8	0.810	8
2	0.614	6	0.799	9
3	0.642	5	0.909	4
4	0.692	2	0.938	1
5	0.652	4	0.926	2
6	0.728	1	0.906	4
7	0.448	9	0.836	7
8	0.609	7	0.886	6
9	0.675	3	0.922	3
mean	0.623	-	0.881	-
std.dev	0.084	-	0.053	-

In Table 7, the efficiency measures and the rankings of the two parametric techniques differ considerably, although the differences are not such extreme as with the CER and DEA, whereby one technique assigns a PCO efficient and the other technique judge this PCO inefficient. PCO 4 seems to be again a good performer, it is assigned efficient by the stochastic parametric technique and second best by the deterministic parametric technique. The Spearman correlation coefficient for the results in Table 7 is $r_s = 0.7364$, which indicates a (quite) strong correlation. However, it still does matter which techniques is used to assess the efficiency of PCOs, since both techniques judge a different PCO as the best and worst performer.

One of the reasons for the differences in the results in Table 7 is the assumption about stochastic. Consistent with theory, the stochastic parametric technique yields higher efficiency scores than the deterministic parametric technique. The efficiency estimates in Table 7 suggest that the efficiency measures by deterministic technique are a mixture of inefficiency, measurement error and unobservable individual characteristics. This leads to the conclusion that stochastic techniques may yield more appropriate efficiency measures within the context of this study than the three deterministic techniques applied in this study. The question is whether the efficiency measures of the stochastic parametric technique are appropriate.

The following points can be made regarding the sensitivity of the results in Table 7. First, one may question whether the approach used here is an appropriate method within the context of this study. An approach comparable to the FE model can have two limitations, one related to

all panel data applications and the other particular to efficiency assessment application (Jacobs et al 2006). First, the appeal to use an approach comparable to the FE approach reduces if there is insufficient variation within the PCOs in the explanatory variables over time. In this study, there is lack of within variation in the explanatory variables for some of the individuals. Lack of variation in the data can make it difficult to determine the relative efficiency of PCOs and to obtain appropriate efficiency measures. Jacobs et al 2006 proposed that the lack of within variation might indicate that the organizations are at a similar level of efficiency (Jacobs et al 2006). The stochastic parametric model yield high efficiency measures, on average 0.881, and there is no large variation in the efficiency measures across the PCOs. Based on the results in Table 7, no secure conclusions can be drawn whether a PCO is indeed more efficient than another PCO.

An important limitation of the approach used here that needs attention is the fact that time-invariant (unobservable) factors are captured into the organization-specific constant, which means that the model fails to distinguish between organizational inefficiency and time-invariant unobservable heterogeneity (Jacobs et al 2006). The results in Table 7 underlie a strong assumption, namely that all the relevant data is observed and that the model is correctly specified so that the 'effects' arises solely from inefficiency. An unobserved organizational effect that might be captured into the results of this study can be the location of the PCO. It can be expected that the (average) health status of individuals vary across social-economic areas and thus across the PCOs. This external factor can influence the output of PCOs in this study. This study does not take into account this external factor and therefore one can have doubts about the appropriateness of the results to reflect solely the inefficiency of PCOs rather than unobserved factors.

Although the RE estimator can account for unobservable heterogeneity, it is not an alternative method for the method in this study, since it goes beyond the time limits of this master thesis to develop a stochastic parametric RE model with a binary output measure. The stochastic parametric RE model with a binary output measure is more complicated than the RE approach with continuous variables (already available in software package STATA) and the stochastic parametric model as used in this study. Development of a stochastic parametric RE model might provide useful insight for efficiency assessment from the perspective of the health care insurer when one is interested in the relationship between the quality of care and the corresponding inputs.

Conclusion

The results in this section show that the consistency conditions by Bauer et al are not satisfied. The efficiency measures are sensitive to the assumptions underlying the parametric techniques. The stochastic technique may yield more appropriate efficiency measures than deterministic techniques due to presence of measurement error in the data. However, this comes at a price. The stochastic parametric technique is not an easy efficiency assessment technique to use for the health care insurer. The striking question is in what extent the stochastic parametric results deviate from those of a relative simple technique, such as DEA. In other words, does the use of a less easy efficiency assessment technique indeed leads to considerably different efficiency measures and ranking of the PCOs than those obtained by a relative simple technique?

§ 3.5 Results of comparison DEA and the stochastic parametric model

This section compares the results of DEA and the stochastic parametric model. To ascertain the reason(s) for the differences in the results from DEA and the stochastic parametric model, we compare the results to the results from a second stochastic parametric model. The second stochastic parametric model is similar to the stochastic parametric model defined by equation (8.0), but applies the weights of DEA presented in Table 4 to the inputs and outputs. Table 8 reports the results from DEA and the two stochastic parametric models. The DEA scores in Table 8 are the average efficiency scores over the three cross-sectional efficiency scores in Table 4. The results from the second stochastic parametric model are presented in the last column in Table 8.

Table 8: efficiency scores derived from DEA and the two stochastic parametric models

PCO	DEA		Stochastic parametric (1)		Stochastic parametric (2)	
	average TE	rank	average TE	rank	average TE	rank
1	0.808	8	0.810	8	0.204	7
2	1.000	1	0.799	9	0.127	8
3	0.957	4	0.909	4	0.632	1
4	1.000	1	0.938	1	0.382	3
5	0.896	7	0.926	2	0.324	5
6	0.971	3	0.906	4	0.233	6
7	0.715	9	0.836	7	0.120	9
8	0.957	4	0.886	6	0.376	4
9	0.921	6	0.922	3	0.579	2
mean	0.914	-	0.883	-	0.33	-
std dev	0.095	-	0.053	-	0.183	-

This section will examine the sensitivity of the results to the choice of technique as follows. If the results from the second stochastic parametric model converge to the results from DEA then the role of stochastic and other properties of the parametric models are small, which

makes the assumption about stochastic, the assumption about the functional form and the function of less importance than the assumption about the weights. Whereas, if the results from the second parametric model converge to the result from stochastic parametric model then the role of the weights is small that makes the assumption about stochastic, the assumption about the functional form and function of greater importance than the assumption about the weights.

The results from the second stochastic parametric model considerably differ from those of DEA and the stochastic parametric model. The Spearman rank correlation coefficient for the results from the two stochastic parametric models is $r_s = 0.6778$ and for the results from DEA and the second stochastic parametric model is this correlation coefficient $r_s = 0.2689$. The results from DEA and the stochastic parametric model ($r_s = 0.1646$) are more weakly correlated than those from the two stochastic parametric models. Thus, there is more internal consistency within the stochastic parametric models than between DEA and the stochastic model. The second stochastic parametric model more converges to the stochastic parametric model than to DEA, although the correlation is weak for all results. This leads to concluding that the assumption about the weights is of minor importance than the assumptions underlying the parametric techniques.

The primary reason for the difference between the results of DEA and stochastic parametric model relies on the assumption of stochastic and the assumption about the function and functional form. A part of the difference between DEA and stochastic parametric model is due to the presence of measurement error in the efficiency measures from DEA. Another part of the difference between the results from DEA and stochastic parametric model may be due to the imposed function and functional form of the stochastic parametric model in this study. The stochastic parametric model assumes that the relationship between the quality of care and the activities follows a smooth logistic relationship. DEA displays a frontier that fits the observed data and since the results from DEA and the stochastic model considerably differ, it can be concluded that the data might not seem to follow a smooth logistic relationship. Nevertheless, the efficiency measures from DEA are sensitive to the presence of other PCOs in the sample and to outliers in the data. This may perhaps have resulted in a (part of the) frontier being inappropriately shaped en positioned. No secure conclusion can be drawn whether the ‘true’ frontier of PCOs is shaped. The imposed function and functional form by the stochastic parametric model should approximate the ‘true’ frontier, which may not be an easy task. The assumption about the function and functional form may be (too) restrictive given the data of this study. This may have influenced the obtained results by the stochastic parametric model.

Conclusion

Efficiency assessment from the perspective of the health care insurer is about finding a balance between using a simple methodology and deriving more appropriate efficiency measures. Stochastic parametric techniques seem to be a more appropriate efficiency assessment technique within the context of the analysis. However, the analysis in this study is too restrictive to conclude that stochastic techniques truly yield efficiency measures approximating the relative efficiency of PCOs. There are doubts about the appropriateness of the stochastic parametric model and the accuracy of the efficiency measures. Moreover, no hard conclusion may be drawn whether the observed differences in the efficiency level of PCOs is significant. To conclude, the health care insurer should be cautious with drawing conclusions based on the results in this study for contracting health care organizations. The question that remains is what the usefulness of the efficiency assessment is, if the obtained efficiency measures are sensitive to the choice of technique.

Section 4: Decision-making for efficiency assessment from the perspective of the health care insurer

From the regulatory perspective, measuring the efficiency of health care organizations lays a strong focus on the validity of the results, because the efficiency assessment will be used for explanatory or predictive purposes (Parking & Hollingsworth 1997; Newhouse 1994a; Newhouse 1994b). In the context of this study, it is not satisfactory to use results that are sensitive to the choice of technique for contracting health care organizations. The validity of the results is not the only criteria that should be satisfied before the health care insurer can use the efficiency measures for contracting organizations. The criteria that should be satisfied are practicality, parsimony, freedom from bias (validity), plausibility, acceptability and freedom from reverse incentives (Jacobs et al 2006). Thus, even more criteria than the validity criteria need to be satisfied. The extensiveness of those criteria shows the complexity of an efficiency assessment from the perspective of the health care insurer for contracting health care organizations. The question is whether the health care insurers should only use the efficiency measures for contracting health care organizations when all the criteria proposed by Jacobs et al 2006 are satisfied. What happens if the health care insurer uses the efficiency measures in this study for contracting health care organizations?

Newhouse 1994a raises an important issue when efficiency measures derived by the efficiency assessment techniques discussed in this study are used for regulatory of contracting purposes (Newhouse 1994a). Newhouse argues that the use of imprecise efficiency measures for reimbursement or other financial aspects may lead to welfare losses in terms of making wrong judgments about the relative efficiency level of PCOs in the sense that the best performer is not contracted and in terms of creating reverse incentives for the health care organizations. Therefore, one should pay attention to the degree of misestimating of the analysis. The next section will discuss the conceptual model of this study to point out the degree of misestimating in this study. For here, an efficiency analysis from the perspective of the health care insurer may always have some kind of misestimating, since the data used for the analysis is incomplete. The purpose of an efficiency analysis is to make a judgment about the relative efficiency level of an organization. Hereby the health care insurer uses their administrative data, which is data from outside the organization, and in the 'optimal' situation, the health care insurer may also use data from health care organizations. However, the health care organization will always have more relevant data about their production process than the health care insurer. This may lead to patient selection or the health care organization may only focus on those aspects incorporated into the analysis. This study does not incorporate all quality indicators and therefore there may be incentives for the PCO to focus solely on the indicators incorporated into the analysis (Appendix 1). The health care insurer can reduce this

incentive by improving the dataset of this study so that a coherent model can be specified. It is unclear in what extent the PCOs can select patients in practice. Therefore, the health care insurer should be aware of the reverse incentives that an efficiency analysis might create.

The extent to which a welfare loss will truly occur depends on the behavior of the health care insurer to differentiate across the PCOs. Misestimating of the analysis will only lead to welfare losses when efficiency measures do not reflect the actual relative efficiency level of health care organizations. In that case, the health care insurer would make the wrong decision about which PCO performs better than another PCO. As long as the misestimating of the analysis does not affect the relative efficiency level, then misestimating in the analysis will not lead to wrong decision-making from the perspective of the health care insurer. The results in this study indicate that there are good and worse performers among the PCOs. PCO 4 seems to be a good performer, since this PCOs is judged (one of) the best performer in most of the cases. However, based on the results in this study cannot be concluded whether this PCO is truly significant better than the other PCOs. This should need further investigation.

With the status quo, it is inadvisable to make a real differentiation across PCOs based on the efficiency measures in this study. This study could not identify the 'best performer' with certainty. Therefore, the health care insurer should use an efficiency analysis not to differentiate among PCOs, but if the health care insurer wants to perform an efficiency assessment, then the efficiency assessment should be used in conjunction with other instruments to contract health care organizations. To conclude, the health care insurer has to trade-off the possibility of occurrence of reverse incentives with the corresponding consequences and the ability to get more insight into the relative efficiency level of organizations. This may lead to contracting those health care organizations that are a relative better performer.

Discussion

In the current situation, the health care insurer should rely on incomplete information for contracting health care organizations, because the efficiency techniques cannot provide accurate estimates of the organizational efficiency. One of the reasons is the validity issues discussed earlier. Another reason is the difficulty of specifying a model that captures the multidimensionality and heterogeneity of the production process of PCOs. This section points out the strengths and weaknesses of this study related to the data and the specification of the conceptual model.

In this study, two broad uncertainties are present, namely data uncertainty and model uncertainty. Data uncertainty relates to the limits of the data used in this study. This study has appended data registered by health care insurers and health care organizations in order to perform an efficiency analysis. The dataset from the health care organizations is a selective dataset of the dataset from the health care insurers. This imposes limitations to the usefulness of the efficiency measures, because the efficiency measures on the output (O_EFFEC) are an underestimation of the actual efficiency level of PCOs. The health care insurer should be aware of the underestimation of the scores in this study if the efficiency measures will be used to set a baseline for contracting health care organizations. Second, there is uncertainty about the identification of the actual patient population, which may have introduced measurement error in the analysis. As long as health care insurers do not register the diagnosis and the health care organizations where the individual receives the health care services, then indirect identification of the chronic patient population is unavoidable. Another limitation related to the data is that not all quality indicators are incorporated into the analysis, because it is not feasible to specify them all with the available data. This restriction may lead to misestimating of the organizational efficiency of PCOs in delivering health care services to chronic patients. An improvement of the quality of the dataset can reduce the degree of misestimating of the organizational efficiency levels.

The second uncertainty introduced in this study relates to the specification of the model. The first problem that arises with efficiency assessment in the health care sector is the specification of the ‘product’ and the quality of the ‘product’. The output measures used here are measures of the quality of care to proxy the health effects produced by the PCOs. This is an improvement over other efficiency studies that rely on specification of intermediate outcomes as the output measure of health care organizations. However, also the output measure in this study is not complete. The actual ‘product’ of (primary) health care organizations is the health gains in terms of health gains and in terms of responsiveness. The difficulty of efficiency assessment is in first place the specification of the actual ‘product’,

followed by adjusting this 'product' for quality differences, which is the quality of care. The quality of care can be measured by quality indicators, which are used in this study as the output measure of PCOs. Adjusting the 'product' for quality differences requires determination of the level of quality one is willing to pay, in other words, what quality level would be an efficient organization? With the status quo, it is not feasible to define the actual 'product' of health care organizations and to adjust this 'product' for quality differences. One should be aware that the output measure in this study is still a proxy of the actual 'product' of PCOs and the influence of other external influences than case mix differences are ignored. Error can be introduced in the analysis of this study, because another output than the actual output of an organization is specified, which may lead to the presence of unobservable factors in the results of this study.

Besides the misspecification in the output measures, there are doubts about the accuracy of the inputs measures in this study. This study specifies the input of the production process as the number of activities by a GP, GP assistant or GP in the off-hours instead of relying on the input measures capital, energy and material. Not incorporating inputs that are used in practice to deliver health care services and assuming that those inputs are (partially) substitutes of the input measure used here, introduces error in the results since those omitted consumed resources may appear in the inefficiency component. For example, a PCO that may have invested in capital and material that contribute to the quality of care (the output measure in this study) are benefited in this analysis upon those who did not invest in such capital and materials.

An important limitation of this study is the restriction to examine a part of the entire production process of PCOs. Thereby is assumed that the production process of each PCO to deliver care to the chronic patient population is independent of the health care services to other patients. This is a strong assumption, since actions of other health care providers and other actions than those incorporated into the analysis are ignored. The PCOs in the sample of this study are part of a joint venture that might lead to (some) collaboration in the delivery of health care services in practice. For example, one can have doubts about the independence of the health care services by the GP in the off-hours. Ignorance of cooperation among the PCOs may have introduced an underestimation of the efficiency of PCOs or an overestimation of the efficiency of PCOs, depending on the extent of independence of the production process of each PCO that is unknown in this study.

Based on this can be concluded that it is highly complex to specify a model that accurately reflect the production process of health care organizations, certainly from the data available to

health care insurers. The health care insurer may not have all the relevant information to define a coherent model. Care should be placed on the interpretation and the use of efficiency measures obtained by an efficiency assessment that relies on incomplete data and a misspecification of the model. When a health care insurer wants to use efficiency assessment for contracting health care organizations, then the health care insurer should be aware of the error that stems in the analysis and the consequences that the use of those efficiency measures may have. Yet, there is no guidance available about the conditions that a model should satisfy in order to establish an appropriate model. This can make efficiency assessment difficult and there might always be doubts about the appropriateness of the specification of the model in the study.

Conclusion

The purpose of this study was to determine the usefulness of efficiency measures for contracting health care organizations. The analysis in this study showed that the health care insurer cannot be indifferent to the specification of the model and the choice of technique. The specification error in the model stems from the availability of data and the quality of the data. This data used in this study consists of valuable individual panel data, but there are doubts about the reliability of the data of this study. Although the quality of the data should be improved in order to be able to decrease the degree of misestimating, one should keep in mind that the health care insurer always has to rely on their administrative data or in the 'optimal' situation, their data appended to data from health care organizations as the dataset of this study. In such a dataset there is (always) data missing to define a coherent model of the organizational efficiency. It is even questionable whether from the perspective of the health care insurer a coherent model could be defined.

The analysis in this study showed that each efficiency assessment technique yields different efficiency measures and in most cases a different ranking of the organizations. The choice of technique is an important aspect for deriving efficiency measures for contracting health care organizations. This study concludes that from the perspective of the health care insurer, it is about finding a balance between the appropriateness of the efficiency measures that will be obtained and the easiness of the technique. The use of a simple efficiency assessment technique c.q. too simple technique for deriving the efficiency of organizations with a complex production process lays a strong focus on the specification of the model. This can become a highly difficult task and certainly with the data available to health care insurers. The use of a less easy efficiency assessment technique, such as a stochastic parametric model may yield more appropriate efficiency measures than those obtained by deterministic efficiency assessment techniques, which are the Cost-Effectiveness Ratio, Data Envelopment Analysis and the technique comparable to Corrected Ordinary Least Squares. However, the researcher should be aware that the other assumptions underlying the stochastic parametric technique are appropriate assumptions in order to obtain truly more appropriate efficiency measures. The results of the stochastic parametric technique in this study needs to be interpreted cautiously, because this study applied a new approach of Stochastic Frontier Analysis, namely a Stochastic Frontier Analysis with a binary output measure and the use of data on individual level. No hard conclusions can be drawn whether the assumptions underlying the stochastic parametric technique are appropriate assumptions.

To conclude, it is not satisfactory to use efficiency measures that are biased by the specification of the model and the assumptions underlying the techniques for contracting

health care organizations. The efficiency measures may reflect a mixture of the inefficiency level, measurement error and unobservable heterogeneity rather than solely the organizational efficiency. In this study, one primary health care organization, organization 4, seems to be a good performer, but the sensitivity analysis in this study has not been extensive enough to conclude whether this organizations is truly the 'best practice'. This study concludes that if the health care insurer wants to use an efficiency analysis for contracting health care organizations with the current state-of-art, then the efficiency assessment should be used in conjunction with other instruments, because the efficiency measures cannot provide a coherent overview of the performance of organizations.

This study is restricted to examine some of all the efficiency assessment techniques and models that in principle are suitable for efficiency assessment. Further research can make a valuable contribution, since many questions remain to answer. Topics for further research are:

1. Investigation on a conceptual model that specifies the actual 'product' of health care organizations in terms of health gains and responsiveness. Relying on incomplete models for efficiency analysis imposes limitations on the usefulness of efficiency measures for contracting health care organizations, since the efficiency measures consist of specification error.
2. Health care insurers have individual (panel) data of their insured and will not have a complete dataset to assess the efficiency of an organization. From this perspective, it is interesting to investigate, more in detail than is done in this study, the 'standard' Stochastic Frontier Analysis models with the use of individual (panel) data. In addition, incorporation the selection-effects in the models are also of interest in order to obtain reliable efficiency measures.
3. Development of a Random Effects model to account for unobservable heterogeneity that is not related to differences in the organizational efficiency level (with the approach applied in this study, unobservable factors might be captured into the organization-specific constant, those constants are indeed, compared to the other coefficients, large). Moreover, it is interesting to develop such a Random Effects model for multiple binary dependent variables, because health care organizations have multiple outputs. Development of such a Random Effects model goes beyond the limits of this study.
4. Development of confidence intervals for parametric efficiency assessment techniques. Confidence intervals are not incorporated into the 'standard' stochastic efficiency assessment techniques, but it can help to determine whether the observed differences in organizational efficiency levels are significant. Development of those confidence intervals goes beyond the limits of this study.

References

Allen, R., A. Athanassopoulos, R.G. Dyson, E. Thanassoulis. 1997. Weight restrictions and value judgements in data envelopment Analysis: evolution, development and future directions. *Journal of the Operational Research Society* 73:13 – 34.

Battese, G.E. & Coelli, T. 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics Vol.* 38, No. 3:387-399.

Bakker, P. 2004. Het kan écht: betere zorg voor minder geld. Sneller beter – De logistiek in de zorg. Eindrapportage TPG.

Bauer, P.W., A.N. Berger, G.D. Ferrier, D.B. Humphrey. 1998. Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency models. *Journal of Economics and Business* 50: 85-114.

Blank, J.L.T. & E. Eggink 2004. The Decomposition of Cost Efficiency: An Empirical Application of the Shadow Cost Function Model to Dutch General Hospitals. *Health Care Management Science*, Vol. 7, No. 2:79-88.

Blank, J.L.T. & E. Eggink. 1998. *Tussen bed en budget: een empirisch onderzoek naar de doelmatigheid van algemene en academische ziekenhuizen in Nederland.* Sociale en Culturele Studies-26. Den Haag: Sociaal en Cultureel Planbureau/Elsevier.

Cooper, W.W., L.M. Seiford, K. Tone, J. Zhu. 2007. Some models and measures for evaluating performances with DEA: past accomplishments and future prospects. *The Journal of Production Analysis.* DOI 10.1007/s11123-007-0056-4. Springer Science+Business Media, LLC.

Coelli, T.J., D.S. Prasada Rao, C.P. O'Donnell, G.E. Battese. 2005. *An introduction to efficiency and productivity analysis.* Second edition. Springer Science+Business Media, LLC (United States).

Charnes, A., W.W. Cooper, A.Y. Lewin, L.M. Seiford. 1994. *Data Envelopment Analysis: Theory, Methodology and Application*. Boston: Kluwer Academic Publishers.

Charnes, A., W.W. Cooper, E. Rhodes. 1981. Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. *Management Science* Vol. 27, No. 6:668-697.

Chirikos, T.N. & A.M. Sear. 2000. Measuring Hospital Effciency: A Comparison of Two Approaches. *Health Service Research 34:6.*

Doyle, J. & R. Greene. 1994. Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses. *Journal of Operational Research Society*, Vol. 45, No. 5:567-78.

Eggink E & J. Blank. 2001. *Verstandig verzorgd. Een empirisch onderzoek naar de efficiëntie van de intramurale zorg voor verstandelijk gehandicapten.* Den Haag: Sociaal en Cultureel Planbureau.

Evans, D.B., Tandon, A., Murray, C.J.L., Lauer J.A. 2000. *The comparative efficiency of national health systems in producing health: an analysis of 191 countries.* GPE discussion paper Series, no. 29, EIP/GPE/EQC, World Health Organization.

Farrell, M.J. 1957. *The Measurement of Productive Efficiency.* Journal of the Royal Statistical Society, Series A, Vol. 120, No. 3:253-290.

Greene, W. 2002. The Bahavior of the Fixed Effects Estimator in Nonlinear Models. New York University.

Greene, W. 2003. *Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Anaysis of the World Health Organization's Panel Data on National Health Care Systems. New York University.*

Greene, W. 2005. *Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of econometrics 126:269-303.*

Grosskopf, S. & Valdmanis, V. 1987. *Measuring Hospital Performance: a nonparametric approach.* Journal of Health Economics 6:89-107.
Hollingsworth, B. 2003. Nonparametric and Parametric Applications Measuring Efficiency in Health Care. Health Care Management Science 6:203-218.

Hollingsworth, B. 2003. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science* 6:203-218.

Jacobs, R., Smith, P.C., Street, A. 2006. *Measuring efficiency in health care*. Analytical techniques and health policy. Cambridge University Press, United Kingdom.

Kumbhakar, S.C. 1990. *Estimation of technical inefficiency in panel data models with firm and time-specific effects.* Economics Letters 36:43-48.

Lee, Y.H. & Smidt, P. 1993. A production function with flexible temporal variation in technical efficiency, in H.O. Fried, C.A.K. Lovell, S.S. Smidt. *The Measurement of Productive Efficiency: techniques and applications.* New York: Oxford University Press, pp. 237-255.

Murray, C.J.L. & J. Frenk. 2000. *A Framework for assessing the performance of health*

systems. Bulletin of the World Health Organization, 78(6).

Nederlands Zorgautoriteit. 2007. Tariefschikking. 5000-1900-07-4, volgnr. 17.

Nederlands Zorgautoriteit. 2008. Tariefschikking. 5000-1900-08-1, volgnr. 18.

Nederlands Zorgautoriteit. 2009. Tariefschikking. 5000-1900-09-3, volgnr. 24.

Newhouse, J.P. 1994a. *Reimbursement under uncertainty: what do if one cannot identify an efficient hospital.* RAND/UCLA/Harvard Center for Health Care Financing Policy Research. RA981.A2N484.

Newhouse, J.P. 1994b. *Frontier estimation: how useful a tool for health economics?* Journal of Health Economics 13:317-322.

Parkin, D. & B. Hollingsworth. 1997. Measuring production efficiency of acute hospitals in Scotland, 1991-94:validity issues in data envelopment analysis. *Applied economics,* 29, 11: 1425-1433.

Polachek, S.W. & B.J. Yoon. 1996. *Panel estimates of a two-tiered earnings frontier.* Journal of Applied Econometrics, Vol. 11:169-178.

Roll, Y. & B. Gollany. 1993. Alternate Methods for Treating Factor Weights in DEA. *OMEGA International Journal of Management Science.* Vol. 21, No. 1, pp. 99-109.

Sarafidis, V. 2002. *An Assessment of Comparative Efficiency Measurement Techniques*. Europe Economics, London.

Smith, P.C. 1997. Model Misspecification in Data Envelopment Analysis. *Annals of Operations Research* 73:233-52.

Veen A.A. van der, S.L.N. Zwakhals, B. Hazelzet-Crans, J.W. van Manen. 2001. *Niet zonder zorg. Een onderzoek naar de doelmatigheid en kwaliteit van de ambulancezorg.* RIVM-rapport nummer 270556002. Rijksinstituut voor Volksgezondheid. Houten: Bohn Stafleu Van Loghum.

Wong, Y.H.B. & J.E. Beasley. 1990. Restricting Weight Flexibility in Data Envelopment Analysis. *The Journal of the Operational Research Society*, Vol. 41, No. 9:829-35.

Wooldridge, J.M. 2000. *Introductory Econometrics: A Modern Approach.* South-Western College Publishing.

Zorgverzekeraars Nederland. 2008. Basisset prestatie-indicatoren geïntegreerde eerstelijnszorg en aandoeningspecifieke ketenzorg. Ten behoeve van zorginkoop 2009.

Appendix 1: Definition output measures

Variables	Description and definitions
O_EFF	*The indicators within the category "Efficiency" focuses on an effective organization of the primary care. This may lead to reduction of outpatient care and hospitalization.*
DM_1	Nominator: The number of patients with a DM-related outpatient visit per year. Denominator: The total DM population.
DM_2	Nominator: The number of patients with a DM-related hospitalization per year. Denominator: The total DM population.
COPD_1	Nominator: The number of patients with a COPD-related outpatient visit per year. Denominator: The total COPD population.
COPD_2	Nominator: The number of patients with a COPD-related hospitalization per year. Denominator: The total COPD population.
COPD_3	Nominator: The number of patients with a prescription 'Prednison'. Denominator: The total COPD population. (The drug 'Prednison' can give an indication for outpatient care).
CVR_1	Nominator: The number of patients with a CVR-related first outpatient visit per year. Denominator: The total CVR population.
CVR_2	Nominator: The number of patients with a CVR-related first outpatient visit per year. Denominator: The total CVR population.
O_EFFEC	*Disease-Management may have a positive effect on the health status of chronic patients. The registration of medical information is important for good disease-management. The indicators within this category focuses on the process of disease-management (registration of medical information) and medical outcomes.*
DM_3	Nominator: The number of diabetes patients that uses a statin. Denominator: The total DM population. (Statin lowers the LDL level, a lipid in the blood, and may contribute to a better health status).
DM_4	Nominator: The number of DM patients where the HbA1C level is known. Denominator: The total DM population. (HbA1C level is the glucose level in the blood)
DM_5	Nominator: The number of DM patients with a HbA1C level below 8. Denominator: The total DM population.
DM_6	Nominator: The number of diabetes patients with a HbA1C level above 9.5. Denominator: The total DM population.
DM_7	Nominator: The number of DM patients where their lipid profile is known. Denominator: The total DM population.
DM_8	Nominator: The number of DM patients with a LDL below 2.5. Denominator: The total DM population whereby the LDL profile is known.
COPD_4	Nominator: The number of COPD patients participating in a 'stop with smoking'- program. Denominator: The total COPD with a positive smoking status.
CVR_3	Nominator: The number of CVR patients that uses a statin Denominator: The total CVR population.
O_EP	*The indicators within the category "Efficient Prescribing" focuses on the behavior of health care providers to prescribe generic drugs.*
EP_1	Nominator: The number of generic prescriptions 'proton inhibitors' Denominator: The total number of prescriptions 'proton inhibitors'.

Appendix 2: Indicators not incorporated in the analysis

The following quality indicators are not definable with the available data.

Category 'effectiveness':

1. The number of DM patients where the following medical information is known: the blood pressure, the kidney functioning, the BMI, the outcome of a foot and eye test, of the total DM population.
2. The number of COPD patients where the following medical information is known: the GOLD classification, the daily activity pattern, the BMI, the lung functioning and the outcome of lung tests, of the total COPD population.
3. The number of CVR patients where the following medical information is known: the blood pressure, the daily activity pattern, the BMI, the smoking status, the family history, the lipid level and risk factors, of the total CVR population.

Category 'efficient prescribing':

4. The number of prescriptions for DM population of the total drugs costs.
5. The number of antibiotics prescriptions (exclusive chinolon, cefalosporin, amoxicillin) of the total antibiotics prescriptions.

In the NHG-standards, those are protocols for PCOs for the health care service to chronic patients, are descriptions about the prescription of drugs for chronic patients available that are more detailed. However, some of the information in those guidelines is not defined yet as quality indicators for contracting health care providers (NHG-standard Diabetes Mellitus type 2; NHG standard COPD; NHG standard Cardiovascular Risk Management, available at: www.nhg.artsennet.nl).

Appendix 3: Selection procedure data

The datasets used in this study are a selection of the original datasets of the health care insurers and health care organizations. Table 9 reports the selection procedure to derive the dataset X and dataset Y.

Table 9: selection procedure of data records

Dataset X					
	original	**duplicates**	**group selection by drugs usage**	**PCO selection 1**	**PCO selection 2**
number of observations†	51,900	51,885	4,908	1,665	*1,372*
	original	**duplicates**	**group selection by PCG's**	**PCO selection 1**	**PCO selection 2**
number of observations	51,900	51,885	2,026	590	581
Dataset Y					
	original	**duplicates**	**group selection by drugs usage**	**PCO selection 1**	**PCO selection 2**
number of observations	2209	362	284	284♦	*265*
	original	**duplicates**	**group selection by PCG's**	**PCO selection 1**	**PCO selection 2**
number of observations	2209	362	152	152♦	152

† An observation in the datasets stands for a unique individual in a specific year.

♦ The PCO is known for each individual.

The first step in the selection is the exclusion of duplicates. A duplicate is an individual that appears more than one time in a specific year. The duplicates are excluded in dataset X. The reason for the many duplicates in dataset Y is that the PCOs delivered a dataset whereby the data records denotes each activity that is done for a unique individual instead of the data record denotes an unique individual in a specific year. In the analysis, only those data records are selected that are required for defining the variables (temporary exclusion of those duplicates if it is necessary to define this variable). The second step in the data selection is the identification of the chronic patient population. Identification of individuals by Pharmaceutical Cost Groups (PCGs) is more restrictive than with identification by disease-specific drugs. Therefore, the drugs usage of some specific drugs is used to identify the chronic patient population. The third step is assigning individuals to a PCO. Individuals that have not received a health care service at one of the PCO in the sample are excluded. The last step in the selection is excluding individuals that could not be identified due to a missing value for the GP-code.

In the first column in Table 10 is the completeness of the data delivery of PCOs presented. Table 10 shows a (large) variation among the PCOs and dataset Y represents on average 19.315% of the observations in dataset X. The question is whether the PCOs delivered a selective dataset. In last three columns in Table 10 is the prevalence reported of some characteristics of the individuals, namely the age (AGE), the PCGs and Diagnostic Cost Groups (DCG). The prevalence is a ratio whereby the nominator is the prevalence of the variable in dataset Y and the denominator is the prevalence of that variable in dataset X. In Table 10, a score of '100' denotes that the PCO delivers the same population as in dataset X occurs. A score '<100' represents that in dataset Y relative less observations of this variable appears than in dataset X. A score of '>100' stands for a higher prevalence than in dataset X. Based on Table 10 can be concluded that all PCOs delivered a selective patient group whereby most PCOs delivered a older and sicker patient population than the population in dataset X. PCO 2 and PCO 4 delivered the most selective patient group.

Table 10: Selection bias in Dataset Y

PCO	Observations (%)	Characteristics of the population		
		Age (%)	PCG (%)	DCG (%)
1	35.593	123.653	124.868	70.238
2	21.769	131.677	259.009	275.625
3	7.071	118.659	145.588	0.000
4	19.737	105.673	154.203	316.667
5	10.256	118.512	192.188	162.500
6	18.304	123.458	153.242	28.755
7	26.316	95.918	152.000	126.667
8	25.926	115.415	160.220	148.352
9	22.477	108.991	132.927	121.336
mean	19.315	115.190	161.811	129.528

Appendix 4: Descriptive statistics nonparametric techniques

		Inputs			Outputs		
	DMU	*AC_GP*	*AC_GP*	*AC_GPO*	*O_EFF*	*O_EFFEC*	*O_EP*
year 2007	**1**	68.75	1.25	2.73	1.05	1.01	1.24
	2	69.90	0.01♦	0.01♦	0.94	1.01	1.00
	3	94.50	0.01♦	8.18	1.06	1.10	1.00
	4	64.96	6.46	0.01♦	1.08	1.19	1.49
	5	59.46	10.08	1.75	0.97	0.43	0.85
	6	69.17	2.48	0.63	1.04	1.04	0.69
	7	106.03	3.23	2.56	0.96	1.12	1.00
	8	72.10	6.61	0.01♦	0.94	1.03	0.90
	9	80.21	6.23	0.36	0.95	1.07	0.84
year 2008	**1**	118.89	8.05	4.31	0.87	0.96	0.98
	2	67.83	0.41	0.59	0.84	1.01	1.19
	3	93.80	6.70	0.01♦	0.96	1.00	1.03
	4	107.22	2.54	0.01♦	1.15	1.04	0.96
	5	66.32	9.26	1.43	1.04	1.02	0.83
	6	65.19	3.43	2.37	1.00	0.88	1.16
	7	83.25	2.97	1.12	1.05	1.04	1.07
	8	47.75	3.49	0.01♦	1.07	1.04	0.87
	9	47.57	0.47	0.01♦	1.02	1.01	0.92
year 2009	**1**	85.97	0.95	2.58	0.89	0.99	0.87
	2	74.95	0.49	2.97	1.02	1.04	1.21
	3	74.00	6.30	2.91	1.10	1.03	0.84
	4	60.26	6.26	0.01♦	1.05	1.08	0.65
	5	69.68	14.11	1.91	0.97	1.05	1.06
	6	76.26	5.92	1.94	1.00	0.91	1.14
	7	96.03	4.60	2.23	0.95	1.08	1.08
	8	60.04	3.63	4.64	1.01	1.07	1.16
	9	69.90	1.64	3.19	1.00	1.05	0.99

♦ Those input values have been changed, since DEA do not accept zero input values.

Appendix 5: Descriptive statistics outputs parametric techniques

O_EFF

PCO	DM_1		DM_2		COPD_1		COPD_2		COPD_3		CVR_1		CVR_2		**O_EP** EP_1	
	0	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*
1	55	4	57	2	55	4	59	0♦	52	7	56	3	57	2	41	18
2	141	6	144	3	144	3	147	0♦	136	11	138	9	141	6	117	30
3	98	1	99	0♦	97	2	99	0♦	93	6	97	2	95	4	80	19
4	76	0♦	75	1	76	0♦	76	0♦	73	3	71	5	70	6	64	12
5	267	6	266	7	270	3	273	0♦	260	13	264	9	257	16	224	49
6	214	10	221	3	223	1	224	0♦	206	18	214	10	213	11	186	38
7	113	1	113	1	111	3	114	0♦	104	10	108	6	109	5	82	32
8	153	9	160	2	160	2	162	0♦	151	11	153	9	154	8	131	31
9	213	5	213	5	213	5	217	1	207	11	214	4	204	14	179	39
total	1,330	42	1,348	24	1,349	23	1,371	1	1,282	90	1,315	57	1,300	72	1,104	268

O_EFFEC

PCO	DM_4		DM_5		DM_6		DM_7		DM_8		DM_3		COPD_4		CVR_3	
	0	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*
1	63	9	65	7	72	0♦	64	8	70	2	47	12	72	0♦	42	17
2	153	17	156	14	169	1	154	16	162	8	131	16	169	1	114	33
3	43	3	43	3	46	0♦	42	4	44	2	88	11	46	0♦	65	34
4	51	6	51	6	57	0♦	52	5	54	3	68	8	57	0♦	49	27
5	88	1	86	3	89	0♦	88	1	89	0♦	208	65	89	0♦	181	92
6	189	24	191	22	213	0♦	189	24	205	8	192	32	213	0♦	160	64
7	120	13	122	11	132	1	120	13	122	11	99	15	133	0♦	74	40
8	187	35	190	32	221	1	192	30	208	14	129	33	222	0♦	114	48
9	252	3	221	34	255	0♦	221	34	224	31	178	40	255	0♦	157	61
total	1,146	111	1,125	132	1,254	3	1,122	135	1,178	79	1,140	232	1,256	1	956	416

♦ The corresponding DMU is omitted from the regression, because there is no within variation. The assumption made here is that those DMUs are efficient (has received efficiency score on this indicator of 1).

- Note: the inputs are the total weighted number of activities as defined in Table 2.