Quality in Primary Care 2013;21:345–57 © 2013 Radcliffe Publishing

Research paper

Assessment of physician performance for diabetes: a bias-corrected data envelopment analysis model

Angela Testi BScEcon
Professor, Department of Economics, University of Genoa, Italy

Naleef Fareed PhD MBA
Assistant Professor, Department of Health Policy and Administration, The Pennsylvania State University, University Park, PA, USA

Yasar A Ozcan PhD MBA
Professor, Department of Health Administration, Virginia Commonwealth University, Richmond, VA, USA

Elena Tanfani PhD
Research Assistant, Department of Economics, University of Genoa, Italy

ABSTRACT

Background In most national health systems, especially when universal coverage is provided, family physicians act as gatekeepers, because most healthcare services are only delivered if there is a formal prescription provided by a primary care physician. Although the consumption of healthcare resources is initiated by prescriptions coming from family physicians, studies that evaluate their performance, especially those using a consolidated methodology (e.g. quality and efficiency) are limited in the literature. The specific aim of this paper is to propose a method for assessing primary care performance.

Methods The novelty of the proposed model is twofold. First, physician performance is assessed following a clinical pathway that focuses on homogeneous groups of patients, in this case, diabetes patients. Second, we argue that performance should not be limited to efficiency, but should encompass clinical effectiveness. Performance assessment is not based on the physician practice as a whole, but on a single disease, in this paper, diabetes. Data were collected from a sample of family physician practices in Italy, and Data Envelopment Analysis (DEA) is used to evaluate their efficiency performance.

Results We found that 35 of 96 practices were efficient based on the standard DEA model. The number of efficient practices decreased based on three restricted models that explored various behavioural preferences of physicians in relation to patient visits, medication administration and referrals to hospitals.

Conclusion The efficiency assessment is completed by a post-hoc evaluation of effectiveness, which in this study is defined as patient care adherence to the prescribed guideline. This study identified best practices both in terms of efficiency and effectiveness. The methods used in this paper are generalisable and could be applied to many other chronic conditions, which may constitute the prevalent activities within the primary care.

Keywords: data envelopment analysis (DEA), performance assessment, primary care

How does this fit with quality in primary care

What do we know?
Diabetes mellitus is a chronic disease that is mainly managed in a patient's home with appropriate, periodic controls on their lifestyle, use of medication when necessary, and, in some cases, admission to hospital. The management of patients with diabetes is a role of the primary care physician. Correctly managed, patients may not only avoid costly hospital admissions, but, much more importantly, also prevent complications that may lead to severe disability (such as retinopathy and blindness, nephropathy and diabetic foot).

What does this paper add?
Appropriateness and efficiency should be both evaluated and assessed, since healthcare delivery performance cannot be limited to the efficiency dimension. This requires, of course, that assessing performance should be done in strict collaboration between economists and clinicians, to know both the appropriate input mix as well the appropriate clinical pathway to compare with current clinical behaviours in family practices.

Introduction

Family physicians play an influential role in determining total healthcare expenditure, because in most health systems, every service referred to the different levels of secondary and tertiary care, such as diagnostics, drugs, hospital admission, must be prescribed by them. Assessing the primary physician's performance in contributing to resource allocation is, therefore, particularly important. In spite of this, the performance of primary care practices is not generally assessed or limited only to the total budget requirement and not to how patient needs are met.[1–4]

The starting point to assess performance is to identify the relevant production process: this involves the processes that transform inputs into outputs and the identification of the relevant inputs and outputs to be considered. Family physicians' activities include many different services such as disease prevention, visits, drug prescription and chronic disease management. In this paper, we assess primary care practices in managing patients with a chronic disease: diabetes mellitus. The proposed method, therefore, is not based on the whole activity of the family practice as it typically occurs in hospitals, but on a single disease. In particular, the study evaluates the manner in which different primary care practices deliver care to patients affected by the same chronic disease. This study uses the case of diabetes mellitus as a chronic illness that rapidly increases with the aging of a population.

Diabetes mellitus affected approximately 246 million adults worldwide in 2008, according to the International Diabetes Federation and this number is expected to soar to 380 million adults worldwide by 2025.[5] With the increasing number of patients with this disease, concerns are raised about how physicians will effectively manage their care. Diabetes and the complications associated with the disease account for a large number of hospital admissions each year, and hospitalisations due to diabetes are responsible for a significant proportion of hospital expenses.[6] In Italy, about 3 million people (4.9% of the population) have been diagnosed with diabetes, and have then been offered healthcare services.[7] The number is expected to increase to 5 milliion by 2030, because there are also 1 million pepole (1.6%) thought to have diabetes who have not yet been diagnosed. Moreover, another 2.6 million people (4.3%) are finding it difficult to maintain good control of diabetes, which puts them at a greater risk of developing associated complications in a short time.[7]

Diabetes mellitus is a chronic disease that should be managed mainly in a patient's home with appropriate, periodic controls on their lifestyle, use of medication when necessary, and only in some cases, admission to hospital. The management of patients with diabetes is a role of the primary care physician. Correctly managed patients may not only avoid costly hospital admissions, but, much more importantly, also prevent complications that may lead to severe disability (such as retinopathy and blindness, nephropathy and diabetic foot). In addition, diabetes increases the risk of developing stroke or heart attack.

Conceptual framework

The specific aim of this paper is to evaluate the performance of primary care practices in the treatment of their diabetic patients. In the prior literature, the methodology most used to assess healthcare delivery performance is data envelopment analysis (DEA).[8] The same methodology can also be used to evaluate the efficiency of practices. In Giuffrida and Gravelle's work,[9] DEA is cited as being a useful tool in determining efficient physician practices, along with the associated savings if 'best' practices were to be adopted

by inefficient physicians.[2] There are very few disease-based studies of DEA: examples include those conducted by Chilingerian *et al*[1] and Ozcan *et al* who investigated otitis media, sinisitus and ashtma.[2–4,11] DEA methodology utilises a linear combination of outputs over a linear combination of inputs to determine unit (in this case primary care physician) performance. The details of this technique are explained in the Methods.

In this paper, we follow Chilingerian and Sherman's and Ozcan's methods.[1,2,10,11] In particular, Ozcan noted that physician practices can be evaluated using DEA based on specific disease conditions such as diabetes;[11] the inputs and outputs of a physician practice follow the same logic used in hospital production, where patient treatments are outputs, and the resources used to produce these treatments are presented as inputs. For family physicians, however, the DEA model should take another point of view, which is a disease- or a patient-centred perspective. This means that the delivered health services (visits, laboratory tests, hospital admissions and so on) are not considered outputs as in the case of a hospital facility,[2] but the inputs of a production process where 'properly followed' diabetic patients are the outputs. In addition to the use of a standard DEA model for assessing the efficiency of physicians' practices in treating diabetes, concurrent examination of the values of the input weights (also called multipliers) may allow researchers to delineate how certain physicians' preferences (e.g. more drugs than visits or hospitalisation) influence healthcare delivery.[2,11]

Assessing efficiency alone is not sufficient, because the final effect on health outcome (that is effectiveness) is also important in healthcare delivery. Both dimensions are key aspects of performance assessment: effectiveness is the ability to achieve the delivered service goal and objectives (to do the right job), whereas efficiency means producing the services with a minimum resource level required (to do the job right).[11]

Efficiency is easier to determine because most earlier studies on production functions show consistent methods to compare outputs with inputs. A question remains about how effectiveness should be measured. Doing the right job is a clinical issue, which may be difficult to quantify. This study maintains that evaluating family physician effectiveness should be kept separate from outcome evaluation.[12] For instance, glycaemic level may measure health status, but not the effectiveness of physicians' actions, because it is generated not only by decisions taken by the physician, but also by patient behaviour and the different severity of their pathology and socioeconomic conditions.

This study proposes to measure the so-called 'appropriateness' of provided services, which depends only on the physician's decision, as a proxy of effectiveness. Appropriateness means that the physician ensures that the patient receives the right services at the right moment. In particular, physicians are attempting to monitor and manage their patients' glycaemic levels through activities such as more frequent consultations with patients, which may in turn prevent or delay complications from the disease.[5] The use of appropriate medications along with algorithms of diabetes disease management could contribute to better primary care for patients with diabetes.[13] Drugs such as metformin and insulin may be used to control the disease.[13] The focus on more interaction with patients and/or the use of the drugs may also help practice physicians in preventing their patients from having costly and life-threatening admissions to acute care facilities such as hospitals.

In summary, the novelty of the proposed approach is twofold. First, physician performance is assessed from the point of view of the patient following a clinical pathway. Second, we assume not only an economic point of view, which is based on the number of services delivered, but also a clinical one, which is whether the pathway is appropriate and evidence based. This last consideration is particularly important, because concerns have been raised about the excessive costs due to variations in medical practice that result in inappropriate care being provided to patients by physicians.[11]

Methods

Data collection

Following a clinical pathway perspective requires collection of data at the individual level, because a clinical pathway can be conceived as an algorithm detailing all treatments to be performed for a patient with a given pathology, with logic based on sequential stages.[14] In this study, data were available thanks to the collaboration of GP-LIGUR.net, the Primary Care Observatory of Regione Liguria in Italy. Data were composed of clinical and prescription information for about 200 000 citizens and more than 140 physicians. The physician data were collected using the same software (www.millewin.it) between January 2010 and December 2011. Because some patients entered the physician panel towards the latter part of the 2010, they were followed up for one year, which is why we collected data until the end of 2011 (two years).

Family physicians collected all clinical information regarding their patients. This was done on a voluntary basis and did not represent a new reporting obligation on family physicians. In fact, they were not obliged to record data in order to be paid, because in Italy doctors are paid based on capitation and not activity.

This might be a drawback, because they may not have recorded all the information, but also an advantage because they were not incentivised to register prescriptions for opportunistic reasons. From the physician's point of view, the only aim of registration was to follow his/her patient better by collecting all the relevant information. The data were therefore suitable for describing the appropriateness of the clinical pathways for patients with chronic illness requiring continuous monitoring by their physicians.

Data reliability was sufficiently guaranteed because the data were filtered and validated with the collaboration of the physicians who agreed a set of criteria identifying a 'good compiler' (e.g. the propensity to insert numerical values for the results of the prescribed examinations, the propensity to link the problem with the particular prescription, the propensity to record new problems using the internal ICD9 code and not a description only). Only the data from good compilers (96 of 140 physicians) was utilised for this study.

The final database included all the information needed to measure performance. Efficiency can be assessed using a large number of inputs and outputs. Among the different ways of evaluating primary care efficiency, we prefer defining inputs and outputs following Chillingerian *et al*, Ozcan and Amado *et al*,[1,2,11,15] which involves considering visits, hospital admissions and medications as inputs, and patients classified into severity levels as the particular output of the production process. With regard to effectiveness, for a chronic condition such as diabetes, once a person is diagnosed with the condition, the appropriate course is to control their disease, and prevent serious complications such as cardiovascular disease, kidney damage, blindness and lower limb amputation. From the available database it was possible to assess the extent to which the family physicians controlled each patient.

Variables used for performance evaluation

Efficiency assessment

Following Chilingerian[1] and Ozcan,[2–4] the inputs used for the study included the annual number of patient consultations with their physicians, the total annual hospital admissions for patients with diabetes (although data were collected for two years, data metrics were developed for each patient on a yearly basis for a given physician), and whether the physicians administered diabetes-related medication (metformin or insulin and other drugs) for a given patient during the year.

With regards to the outputs, our database mainly contains elderly patients aged over 65 years who were classified into three different levels of patient severity (i.e. low, medium and high), based on age and the presence of associated comorbidities (hypertension, heart failure) among patients with diabetes seen by the practices. In particular, the three severity classes were considered in the model for three different outputs defined as:

- low severity, < 65 years old with no comorbidities
- medium severity, 65 years or older with no comorbidities
- high severity, 65 years or older and comorbidities.

Using this method we were able to construct what resources (hospitalisation, drugs) were used for each patient in a year, and classify the patient to a severity category.

Effectiveness assessment

For primary care facilities, we defined effectiveness as the ability of the physician to follow an evidence-based programme along the clinical pathway for the care of patients with diabetes, rather than as a judgement about the final health status of the patients.

In the healthcare literature, this particular concept is usually called 'appropriateness'. The underlying assumption is that the physician has done everything the clinical pathway prescribes for that particular patient. For instance, it is important that the physician checks that the patient does not smoke. Whether the patient eventually gives up smoking cannot be solely attributed to a physician's performance.

The appropriateness of the patient clinical pathway is assessed through a set of indicators chosen by the general practitioners taking part in the study, which is in agreement with the Italian Standards for Diabetes Mellitus (www.aemmedi.it/files/Linee-guida_Raccomandazioni/2007/2007-cura-diabete-mellito.pdf). They defined care pathway for diabetes is appropriate if, at least, the following parameters are checked:

- haemoglobin alpha 1 (HbA1) once every 12 months
- creatinine, once every 15 months
- microalbuminuria, once every 15 months
- low-density lipoprotein (LDL), once every 15 months
- smoking status (as a proxy for lifestyle).

More specifically, the paper uses a weighted 'appropriateness index score', which rewards physicians that provide a higher proportion of their diabetic patients with HbA1 (score of 1); HbA1, creatinine, microalbuminuria and LDL tests (score of 2.5); and HbA1, creatinine, microalbuminuria, LDL test and smoking assessment (score of 3). The values of these scores were agreed with the physicians taking part in the study and may be sensitive to different opinions. The sum of these scores was then standardised to a score of 0–1, with the practice with the highest weighted appropriateness index score being the reference for all the other practices.

Model specification

Assessment of the efficiency of the production process, in our case of a primary care practice using healthcare services to monitor patients with diabetes, consists of computing the ratio between outputs and inputs and comparing it with a benchmark or ideal value. Models differ as to how one might determine this comparison value. In the literature, the peculiarity of health services production and difficulties encountered in this estimation have led to a search for so-called 'non-parametric' methods of efficiency evaluation, different from the traditional econometric analysis or stochastic frontiers (parametric methods).[9] The usual choice is DEA, based on the economic concept of Pareto optimality,[16] and developed by Charnes and colleagues[17] as a linear programming problem.

Compared with parametric methods, the use of DEA has many advantages because it does not require the specification of a functional form and a distribution of the error, it may be applied to technologies that use multiple inputs for more than one output, and the benchmark can be identified among the other family practices in the sample. Consequently, applications of the DEA method in health have increased over the years, especially when focused on hospitals.[8] In DEA models, the efficiency of a decision-making unit (DMU), which in our application is a family practice, is evaluated as the weighted ratio between outputs and inputs for a given physician relative to their peers in the evaluation. The physician is efficient, and is assigned a score of 1, if this ratio is greater than or equal to the corresponding calculated ratio, using the same system of weights, for the other physicians in the evaluation set. Each physician is allowed to choose the weights so as to maximise efficiency score. The only constraints are that the weights must be non-negative and such as not to render > 1 the ratio for all the other physicians.

A detailed description of DEA methods and the calculations behind DEA have been described at length in the literature.[11,18] We provide a brief explanation for the calculation of DEA efficiency scores here using mathematical notations adapted from Ozcan (pp. 24–56).[11] The efficiency scores (θ_o) for a group of peer clinics ($j = 1 \ldots n$) are computed for the selected outputs (y_{rj}, $r = 1, \ldots s$) and inputs (x_{ij}, $i = 1, \ldots m$) using the following fractional programming formula:

$$\text{Max} \quad \theta_o = \frac{\sum_r u_r y_{ro}}{\sum_i v_i x_{io}}$$

$$\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \le 1 \qquad \forall j$$

$$u_r, v_i \ge 0$$

In this formulation, the weights for the outputs and inputs, respectively, are u_r and v_i, and 'o' signifies a focal clinic (to obtain efficiency scores, each clinic in turn becomes a focal clinic when its efficiency score is being computed relative to others). Note that the input and output values, as well as all weights are assumed by the formulation to be > 0. The weights u_r and v_i for each physician are determined entirely from the output and input data for all physicians in the peer group. Therefore, the weights used for each physician are those that maximise the focal physician's efficiency score. In order to solve the fractional programme described above, it needs to be converted to a linear programming formulation. Because the focus of this paper is not on the mathematical aspects of DEA, an interested reader is referred to Ozcan[11] for more detail on how the above equations are algebraically converted to a linear programming formulation. Other technical DEA books and papers may also be consulted for an in-depth exposure.[18,19] In summary, the DEA identifies a group of optimally performing physicians that are defined as efficient, and assigns them a score of 1. These efficient physicians are then used to create an 'efficiency frontier' or 'data envelope' against which all other physicians are compared.

Note that the mathematical formulation assumes a so-called input-oriented DEA model, which focuses on the extent to which input quantities can be reduced without changing output quantities, whereas output models focus on an organisation's attempt to maximise outputs without altering input quantities. In assessing physicians' performance, we preferred the input-oriented model, because the assumption is that physicians have control over their inputs and not their outputs (patient severity based on co-morbidities).

DEA estimates can be obtained by utilising a constant returns to scale (CRS) or variable returns to scale (VRS). The CRS model assumes that there is a linear, proportional change in outputs for changes in inputs, whereas the VRS assumes that returns are dependent on changes in volume.[19] Our model specification employs a VRS because the practices are different in size and hence cannot be assumed to have similar economies of scale.[11] More specifically, the VRS model accounts for the possibility that different practices may have different proportions of change in output for a given amount of input based on characteristics such as the number of physicians employed by the practice.

$$\text{Max} \quad \sum_r u_r y_{ro} - u_o$$

$$\sum_r u_r y_{rj} - \sum_i v_i x_{ij} - u_o \le 0 \qquad \forall j$$

$$\sum_i v_i x_{io} = 1$$

- $u_r, v_i \ge 0$
- u_o unconstrained

The negative sign of the additional variable u_o indicates increasing or diminishing returns if it is positive or negative, respectively.

In this study, two further specifications were used. First, because in the DEA approach the absence of an error term can create problems in measurements of errors or deviations,[20] a bootstrapping method was used to analyse the sensitivity of the measured efficiency scores to sampling variation. For this reason, in accordance with the recommendations of Simar and Wilson,[20] bias-corrected efficiency scores are also computed to overcome the inherent bias present in the construction of the DEA score of the practices. Using the FEAR command in the R statistical software package, the study used the homogenous bootstrap algorithm provided by Simar and Wilson with 1000 repetitions. The FEAR program was also used to generate the DEA scores for the standard model.

Second, DEA models do not prioritise certain inputs over others for a physician; for instance, how much they value the contribution of visits, drugs or hospital admissions. In the case of healthcare delivery, however, substitution among inputs can have different meanings, reflecting both different severity and also sometimes inappropriate actions. Clinical pathways are designed to constrain physicians to choose the right mix of inputs.

In the case of diabetes, particularly in the early stages of the disease, glycaemic control and a healthy lifestyle are sufficient, whereas in the second stage one begins to administer specific drugs (e.g. metformin), and hospital admissions in all cases should be avoided and used only in extreme cases, not being a substitute for other services. For this reason, the impact of substitution among inputs, concurrent use of a restricted model (RM) in addition to the use of a standard DEA model may allow delineation of different practice styles, that is the relative contribution in the preferred pathways of visits, drugs and hospital admission.[2,11] In particular, to obtain the efficiency scores of the restricted DEA models, upper (80th percentile) and lower (69th percentile) bound weights were assigned to physicians that preferred to administer medication over hospital admissions (RM1), preferred to have patients visit them over hospital admissions (RM2) and preferred to administer medication and have visits over hospital admissions (RM3). The literature suggests that if there is no *a priori* information for weight restrictions to confine physician practices to a norm (or practice pattern), then using percentiles or quartiles of the weight distributions generated by non-weight restricted DEA models is appropriate.[2,11] The decision to use the 69th percentile for lower bound restriction and the 80th percentile for upper bound restriction was rendered after examining the weight structure of these variables and running the initial unrestricted model.

In summary, DEA models allowed the computation of a set of scores for each physician depending on the particular specification of the model (standard, bootstrapped and restricted). This assessment concerned only efficiency and not appropriateness; this comes down to the physicians' choices regarding following specific types of evidence-based programmes to manage the disease condition. Final performance was therefore assessed taking into account efficiency and appropriateness scores for each physician. The latter score was obtained with a post-hoc evaluation with regard to the effectiveness of physicians in delivering for their patients the appropriate clinical pathways following clinical guidelines.

Results

A descriptive summary of the utilised database is given in Table 1. On average, in Italy, a primary care practice provides care for more than 120 450 patients older than 14 years, and about 8% are affected by diabetes mellitus. Most patients with diabetes are elderly (73%) and diabetes is associated with at least one other chronic pathology in 77% of patients.

A summary of the efficiency scores, input, and output variables is presented in Table 2. Inputs are the average number of yearly contacts the physician had with their patients with diabetes, the number of hospital admissions in the previous five years and the percentage of prescriptions for metformin, insulin and other antidiabetic drugs. Outputs are the percentage of patients classified as being low, medium and high severity.

In the standard model (first row of Table 2), the mean input-oriented, VRS efficiency score of the sample practices was 0.86, with 35 practices (fewer than half) being efficient. In the bias-corrected model, mean efficiency fell to 0.78 and none of the practices was efficient (see Appendix for detailed information about individual practice efficiency scores and confidence intervals). A *t*-test comparing the standard efficiency and bias-corrected scores indicated that their means were significantly different ($P < 0.001$), although we focus on the standard efficiency score for convenience.

The restricted models had a lower mean efficiency score in relation to the standard efficiency scores, and RM2 had more efficient practices than RM1 and RM3, but still fell short in relation to the standard efficiency model.

In the post-hoc evaluation, practice (DMU) efficiency and standardised appropriateness scores were compared. Detailed scores for the first practices are presented in Table 3 to show that the efficiency score

Table 1 Descriptive statistics from the utilised database (January 2000 to December 2011, GP-LIGUR.net)

Variable	
No. of family doctors (DMUs)	96
Total no. of patients (> 14)	120 450
No. of diabetic patients	9532
% diabetic patients	7.90
% diabetic patients aged < 65	27
% diabetic patients aged 65–75	30
% diabetic patients aged > 75	43
% diabetic patients with no co-pathology	23
% diabetic patients with 1 co-pathology	50
% diabetic patients with > 1 co-pathology	27

can be misleading without the appropriateness assessment.

With reference to the standard efficiency model, for example, Practice 3 is both efficient and appropriate, Practice 9 is efficient (score of 1), but poor at following medical guidelines with a score of only 0.04. Similar considerations may be relevant for the other efficient practices reported in Table 2. However, Practice 3 is the only one that performs best from both an efficiency and appropriateness point of view.

The results for the sample are presented in Figure 1, in which practices are dichotomised by their level of efficiency and effectiveness. With regard to the quadrants in the figure, practices with efficiency scores of 1 and effectiveness scores ≥ 0.90 were regarded as the best performers, whereas those with efficiency scores < 1 and effectiveness scores < 0.90 were classified as poor performers. Those with scores in only one area (i.e. efficiency or effectiveness) were categorised in quadrants that indicated their need to improve in the dimension in which they performed poorly.

As illustrated, Practices 3 and 24 had the best performance in regard to efficiency and appropriateness. Thirty-three practices needed to improve on appropriateness, but were efficient. The remaining practices (61) were poor performers in both efficiency and appropriateness scores. There were no practices in the lower right quadrant, where a practice may have had high appropriateness and low efficiency.

Additional information can be drawn from Table 3 with regard to the restricted models. In general, the restrictions were more severe for RM1 than for RM2; this means that it was more difficult to avoid hospital admissions by prescribing drugs (RM1) than by

Table 2 Descriptive summary of efficiency scores, input and output variables

Efficiency of physician practices ($n = 96$)	Mean (SD)
Input oriented, VRS efficiency score	0.86 (0.15)
Input oriented, VRS efficiency score – bias corrected	0.78 (0.12)
Input oriented, VRS efficiency score – RM1 model[a]	0.66 (2.32)
Input oriented, VRS efficiency score – RM2 model2[b]	0.80 (0.17)
Input oriented, VRS efficiency score – RM3 model3[c]	0.52 (0.26)
Inputs	
Visits	969.61 (600.76)
Admissions	27.63 (33.31)
Drugs 1 (metformin and insulin)	40.38 (18.06)
Drugs 2 (other)	22.45 (12.09)
Outputs	
Low severity	23.26 (15.45)
Medium severity	47.88 (21.49)
High severity	28.20 (18.42)

RM = restricted model; VRS = variable returns to scale. [a] Drugs 1 preferred over admissions. [b] Visits preferred over admissions. [c] Drugs 1 and visits preferred over admissions.

Table 3 Full performance assessment comparing all efficiency and appropriateness scores

Practice	Efficiency standard model	Efficiency RM1	Efficiency RM2	Efficiency RM3	Appropriateness score
1	1	0.64	1	0.53	0.17
2	0.66	0.44	0.65	0.19	0.79
3	1	0.56	1	0.52	1
4	1	1	1	1	0.11
5	0.97	0.17	0.78	0.62	0.44
6	1	1	1	1	0.15
7	0.92	0.92	0.91	0.91	0.59
8	0.88	0.63	0.87	0.49	0.56
9	1	1	1	1	0.04
10	0.58	0.35	0.58	0.27	0.02
11	1	0.77	0.73	0.55	0.49
12	0.98	0.72	0.94	0.65	0.01
Average for all 96 practices	0.86	0.66	0.80	0.52	0.36

RM, restricted model.

		Effectiveness (Appropriateness of the Clinical Pathway)	
		Low < 90	High = 0.90
Efficiency	High = 1.0	Improvement Need On Appropriateness P65, P37, P30, P63, P87, P9, P73, P46, P18, P44, P4, P6, P19, P21, P1, P26, P20, P31, P34, P64, P60, P79, P75, P80, P11, P28, P95, P94, P86, P74, P15, P23, P16	Best Performance P3, P24
	Low <1.0	Poor Performance P48, P83, P10, P29, P41, P78, P50, P71, P49, P66, P17, P2, P47, P68, P89, P82, P81, P25, P92, P69, P27, P42, P96, P77, P93, P45, P13, P33, P53, P32, P91, P85, P84, P61, P38, P39, P55, P67, P57, P72, P14, P70, P62, P43, P8, P90, P56, P54, P40, P7, P51, P88, P52, P35, P59, P36, P5, P22, P12, P76, P58	Improvement Need on Efficiency

Figure 1 Standard efficiency and effectiveness quadrants

visiting the patient personally. Of course, when both restrictions were introduced (RM3) the average score was lower (0.52). It is also interesting to compare all the scores between the models. For instance, Practices 3 and 4 had high efficiency scores in RM2 also (1 and 0.98, respectively). This meant that they also used a correct mix of inputs (i.e. they did not resort to hospital admissions and, instead, took care of their patients at home, by visiting them). Among the 22 efficient practices in RM2, only one (Practice 3) had an appropriateness score of 1 – incidentally, this was also the only practice that had an appropriateness score of 1 among all the practices evaluated. Within the efficient practices in RM1, none had an appropriateness index score higher than 0.70; amongst the efficient practices in RM3, none had an appropriateness score

higher than 0.52. Interestingly, these 11 practices had efficiency scores of 1 in the standard efficiency model, RM1, and RM2 as well. These practices, on average, tended to spend less effort with appropriateness programmes and had more patients of low and medium severity.

Discussion

The availability of individual clinical pathway data and collaboration with primary care physicians provided a more complete performance assessment of family practices than previously reported in the literature, which has generally been limited to efficiency perspectives.

From an epidemiological point of view, the detected prevalence of diabetes of 7.9% was higher than the 4.9% reported on a national basis by the Istituto Superiore di Sanità (www.iss.it). This may indicate how diabetes might evolve in the future, since in the Ligurian population aging has already reached a level that will be reached in other Italian regions in the next 20 years.

The second important result is the existence of great variability among family practices, with respect to the efficiency scores, and greater still in input mix and appropriateness scores. From the restricted models, it appeared that most physicians generally preferred to focus on medication than on visits and this leads to more hospital admissions.

The most surprising result, however, comes from inspection of Table 2. It is apparent that most of the inefficient practices were also inappropriate, and that in no case was high appropriateness accompanied by low efficiency. This seems to contradict the conventional notion that increasing appropriateness (and therefore quality) in healthcare delivery requires greater use of resources.

From the results, it also appeared that appropriateness and efficiency should be evaluated and assessed together, since healthcare delivery performance cannot be limited to efficiency alone. This requires that performance should be assessed in strict collaboration between economists and clinicians, to know both the appropriate input mix and the appropriate clinical pathway to compare with current practice behaviours. Furthermore, it would be interesting to include notions of clinical appropriateness from the patients' perspective (e.g. quality of life measures) as well. Such a measure could help gauge the value of appropriate clinical pathways in relation to their impact on patients.

Assessing the performance of family practices shows that improvement is still needed. It is recommended that future work might include socioeconomic conditions as external variables to make differences in the performance of primary care practices more apparent, as practices facing challenges in their population may have greater difficulty in achieving good performance, due to external factors.[15] Also, political intervention may be implemented to promote improvements. One of the first issues explored could be to promote the so-called medicine initiative, which is to give the physician an active role in intervention, especially where patient adherence to medicines is unsatisfactory.

DEA can also give provide insight into performance assessment in primary care practices, such as helping decision makers to detect problem practices and plan appropriate strategies for improvement fine-tuned case by case. This assessment is particularly timely because of the current debate taking place in Italy regarding the possible reform of primary care. It is recommended that reforms promote a shift in resources from hospitals to primary care. This would leave primary care practices operating continuously 24 hours a day, seven days a week, making physicians group together. Different dimensions of practices do not seem, however, to induce more efficiency.[21]

The issue of how to achieve improvement remains unresolved. Most relevant in this case seems to be the incentive embodied in payment design. Currently, Italian family physicians are reimbursed by capitation, but this could be replaced by a mixed system, in which a quota based on the appropriateness of prescriptions is added to the fixed compensation.[22] Of course, in order to implement incentive schemes such as 'pay for performance', it is necessary to share with physicians appropriate methods for assessing the efficiency and the quality of their activity, which is what was attempted in this paper.

ACKNOWLEDGEMENTS

Data were collected thanks to the collaboration with GP-LIGUR.net (Dr Pier Claudio Brasesco).

REFERENCES

1 Chilingerian J and Sherman H. DEA and primary care physician report cards: deriving preferred practice cones from managed care service concepts and operating strategies. *Annals of Operational Research* 1997;73:35–66.
2 Ozcan Y. Physician benchmarking: measuring variation in practice behavior in treatment of otitis media. *Health Care Management Science* 1997;1:5–17.
3 Pai C-W, Ozcan YA and Jiang HJ. Regional variation in physician practice pattern: an examination of technical and cost efficiency for treating sinusitis. *Journal of Medical Systems* 2000;24:103–17.

4 Ozcan YA, Jiang HJ and Pai C-W. Physician efficiency in treatment of sinusitis: do primary care physicians or specialist provide more efficient care? *Health Services Management Research* 2000;13:90–6.
5 Nicholoson G and Hall M. Diabetes mellitus: new drugs for a new epidemic. *The British Journal of Anaesthesia* 2011;107:65–73.
6 Levetan C, Salas J, Wilets I and Zumoff B. Impact of endocrine and diabetes team consultation of hospital length of stay for patients with diabetes. *The American Journal of Medicine* 1995;99:22–8.
7 Diabete Italia. *I numeri del diabete*. nd. www.diabeteitalia.it/il-diabete/i-numeri-del-diabete/default.aspx (accessed 09/12).
8 Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science* 2003;6:203–18.
9 Giuffrida A and Gravelle H. Measuring performance in primary care: econometric analysis and DEA. *Applied Economics* 2001;33:163–75.
10 Sherman HD and Zhu J. *Service Productivity Management*. Springer: New York, 2006.
11 Ozcan Y. *Health Care Benchmarking and Performance Evaluation: an assessment using data envelopment analysis (DEA)*. Springer: Norwell, MA, 2008.
12 Giuffrida A, Gravelle H and Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 1999;319: 94–8.
13 Aguilar R. Evaluating treatment algorithms for the management of patients with type 2 diabetes mellitus: a perspective on the definition of treatment success. *Clinical Therapeutics* 2011;33:408–24.
14 De Blaser L, Depreitere R, De Waele K, Vanhaecht K, Vlayen J and Sermeus W. Defining pathways. *Journal of Nursing Management* 2006;14:553–63.
15 Amado C and Dyson R. Exploring the use of DEA for formative evaluation in primary diabetes care: an application to compare English practices. *Journal of the Operational Research Society* 2009;60:1469–82.
16 Farrel MJ. The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A* 1957; 120:253–90.
17 Charnes A, Cooper WW and Rhodes E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978;12:429–44.
18 Cooper W, Seiford L and Tone K. *Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-Solver software* (2e). Springer Science +Business Media, LLC: New York, NY, 2007.
19 Banker RD, Charnes A and Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 1984;30: 1078–92.
20 Simar L and Wilson P. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* 1998;44:49–61.
21 Milliken O, Devlin RA, Barham V, Hogg W, Dahrouge S and Russell G. *Comparative Efficiency Assessment of Primary Care Models Using Data Envelopment Analysis*. Cahiers de Recherche/Working Papers: University of Ottawa, 2008.
22 Gosden T, Forland F, Kristiansen IS *et al*. Capitation, salary, fee-for-service and mixed systems of payment: effect on the behavior of primary care physicians (Cochrane Review). *The Cochrane Library, Issue 3, 2006*. Update Software: Oxford.

FUNDING

This work has been supported by MIUR FIRB RBFR08IKSB project.

PEER REVIEW

Not commissioned; externally peer reviewed.

CONFLICTS OF INTEREST

None declared.

ADDRESS FOR CORRESPONDENCE

Angela Testi, Department of Economics, University of Genoa, Via Vivaldi 5, 16126 Genoa, Italy. Tel: (+39) 010209 – 5271. Email: testi@economia.unige.it

Received 11 June 2013
Accepted 2 October 2013

Appendix

Table A1 Standard efficiency and bias-corrected efficiency scores with confidence intervals

Practice	Efficiency score	95% lower bound CI	95% upper bound CI	Bootstrapped efficiency score	95% lower bound CI	95% upper bound CI	Bias (efficiency score-bootsrapped efficiency score)	Appropriate-ness score
1	1.00	0.88	1.00	0.94	0.88	1.00	0.06	0.17
2	0.66	0.58	0.66	0.62	0.58	0.66	0.04	0.79
3	1.00	0.71	1.00	0.86	0.71	1.00	0.14	1.00
4	1.00	0.71	1.00	0.86	0.71	1.00	0.14	0.11
5	0.97	0.86	0.97	0.92	0.86	0.97	0.05	0.44
6	1.00	0.71	1.00	0.86	0.71	1.00	0.14	0.15
7	0.92	0.74	0.91	0.84	0.74	0.91	0.08	0.59
8	0.88	0.74	0.87	0.81	0.74	0.87	0.07	0.56
9	1.00	0.70	0.99	0.86	0.70	0.99	0.14	0.04
10	0.58	0.49	0.58	0.54	0.49	0.58	0.04	0.02
11	1.00	0.72	1.00	0.86	0.72	1.00	0.14	0.49
12	0.98	0.79	0.98	0.90	0.79	0.98	0.08	0.00
13	0.72	0.63	0.72	0.68	0.63	0.72	0.04	0.01
14	0.85	0.68	0.84	0.78	0.68	0.84	0.07	0.66
15	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.69
16	1.00	0.76	1.00	0.87	0.76	1.00	0.13	0.83
17	0.66	0.59	0.66	0.63	0.59	0.66	0.03	0.83
18	1.00	0.71	1.00	0.86	0.71	1.00	0.14	0.07
19	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.16
20	1.00	0.72	1.00	0.86	0.72	1.00	0.14	0.18
21	1.00	0.71	1.00	0.86	0.71	1.00	0.14	0.16
22	0.98	0.83	0.97	0.91	0.83	0.97	0.07	0.00
23	1.00	0.83	0.99	0.90	0.83	0.99	0.1	0.70
24	1.00	0.80	0.99	0.91	0.80	0.99	0.09	0.94
25	0.68	0.56	0.68	0.63	0.56	0.68	0.05	0.17
26	1.00	0.70	0.99	0.85	0.70	0.99	0.15	0.17
27	0.69	0.58	0.68	0.64	0.58	0.68	0.05	0.27
28	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.51
29	0.60	0.48	0.60	0.56	0.48	0.60	0.04	0.01
30	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.02

Table A1 Continued

Practice	Efficiency score	95% lower bound CI	95% upper bound CI	Bootstrapped efficiency score	95% lower bound CI	95% upper bound CI	Bias (efficiency score-bootsrapped efficiency score)	Appropriateness score
31	1.00	0.79	1.00	0.88	0.79	1.00	0.12	0.20
32	0.74	0.61	0.74	0.69	0.61	0.74	0.05	0.21
33	0.73	0.64	0.72	0.68	0.64	0.72	0.05	0.25
34	1.00	0.70	1.00	0.87	0.70	1.00	0.13	0.21
35	0.95	0.74	0.94	0.87	0.74	0.94	0.08	0.25
36	0.97	0.76	0.97	0.88	0.76	0.97	0.09	0.04
37	1.00	0.70	0.99	0.86	0.70	0.99	0.14	0.01
38	0.78	0.71	0.78	0.74	0.71	0.78	0.04	0.06
39	0.79	0.66	0.79	0.73	0.66	0.79	0.06	0.49
40	0.92	0.72	0.91	0.84	0.72	0.91	0.08	0.07
41	0.63	0.54	0.63	0.59	0.54	0.63	0.04	0.59
42	0.69	0.56	0.69	0.64	0.56	0.69	0.05	0.36
43	0.87	0.71	0.87	0.80	0.71	0.87	0.07	0.47
44	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.08
45	0.71	0.63	0.71	0.67	0.63	0.71	0.04	0.59
46	1.00	0.76	0.99	0.88	0.76	0.99	0.12	0.05
47	0.66	0.55	0.66	0.61	0.55	0.66	0.05	0.53
48	0.51	0.43	0.51	0.47	0.43	0.51	0.04	0.85
49	0.66	0.58	0.65	0.61	0.58	0.65	0.05	0.01
50	0.64	0.54	0.64	0.59	0.54	0.64	0.05	0.66
51	0.94	0.76	0.93	0.86	0.76	0.93	0.08	0.01
52	0.94	0.76	0.94	0.87	0.76	0.94	0.07	0.00
53	0.74	0.59	0.73	0.68	0.59	0.73	0.06	0.08
54	0.92	0.74	0.91	0.84	0.74	0.91	0.08	0.00
55	0.80	0.68	0.80	0.75	0.68	0.80	0.05	0.19
56	0.89	0.72	0.89	0.82	0.72	0.89	0.07	0.40
57	0.82	0.74	0.82	0.78	0.74	0.82	0.04	0.91
58	0.99	0.79	0.99	0.91	0.79	0.99	0.08	0.20
59	0.96	0.73	0.96	0.88	0.73	0.96	0.08	0.61
60	1.00	0.82	1.00	0.92	0.82	1.00	0.08	0.30
61	0.78	0.68	0.77	0.73	0.68	0.77	0.05	0.60
62	0.86	0.78	0.86	0.82	0.78	0.86	0.04	0.76

Table A1 Continued

63	1.00	0.80	1.00	0.90	0.80	1.00	0.1	0.02
64	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.24
65	1.00	0.69	1.00	0.86	0.69	1.00	0.14	0.00
66	0.66	0.54	0.65	0.61	0.54	0.65	0.05	0.34
67	0.81	0.67	0.81	0.75	0.67	0.81	0.06	0.25
68	0.67	0.58	0.66	0.62	0.58	0.66	0.05	0.00
69	0.69	0.60	0.68	0.64	0.60	0.68	0.05	0.58
70	0.86	0.74	0.86	0.80	0.74	0.86	0.06	0.19
71	0.65	0.54	0.64	0.59	0.54	0.64	0.06	0.42
72	0.84	0.72	0.84	0.78	0.72	0.84	0.06	0.95
73	1.00	0.70	1.00	0.85	0.70	1.00	0.15	0.04
74	1.00	0.74	1.00	0.88	0.74	1.00	0.12	0.67
75	1.00	0.74	1.00	0.87	0.74	1.00	0.13	0.43
76	0.99	0.81	0.99	0.92	0.81	0.99	0.07	0.13
77	0.71	0.65	0.70	0.67	0.65	0.70	0.04	0.63
78	0.64	0.54	0.63	0.59	0.54	0.63	0.05	0.89
79	1.00	0.71	1.00	0.85	0.71	1.00	0.15	0.40
80	1.00	0.77	1.00	0.88	0.77	1.00	0.12	0.44
81	0.68	0.61	0.68	0.64	0.61	0.68	0.04	0.19
82	0.67	0.57	0.67	0.62	0.57	0.67	0.05	0.59
83	0.53	0.47	0.53	0.50	0.47	0.53	0.03	0.75
84	0.76	0.65	0.76	0.71	0.65	0.76	0.05	0.45
85	0.76	0.68	0.76	0.73	0.68	0.76	0.03	0.05
86	1.00	0.70	1.00	0.86	0.70	1.00	0.14	0.58
87	1.00	0.80	0.99	0.89	0.80	0.99	0.11	0.03
88	0.94	0.85	0.94	0.89	0.85	0.94	0.05	0.03
89	0.67	0.53	0.66	0.61	0.53	0.66	0.06	0.13
90	0.89	0.80	0.88	0.84	0.80	0.88	0.05	0.76
91	0.75	0.63	0.75	0.70	0.63	0.75	0.05	0.23
92	0.69	0.62	0.68	0.65	0.62	0.68	0.04	0.91
93	0.71	0.62	0.70	0.66	0.62	0.70	0.05	0.83
94	1.00	0.71	0.99	0.86	0.71	0.99	0.14	0.56
95	1.00	0.72	1.00	0.86	0.72	1.00	0.14	0.52
96	0.70	0.60	0.70	0.65	0.60	0.70	0.05	0.55