Ann Oper Res (2006) 145:265–280
DOI 10.1007/s10479-006-0041-8

Pricing public health care services using DEA: Methodology versus politics

Paul Rouse · Robert Swales

Published online: 11 July 2006

Abstract The New Zealand public health sector has used DEA since 1997 to identify efficient expenditure levels to set prices for hospital services at the DRG level. Given the size of the expenditure (NZ$ 2.6 billion), considerable robustness was required for the results and sophistication of the models/process. While the model development and application appeared to be successful, politics overturned the results in the short run. In the longer term, the results have been shown to be reasonably robust and have become a base-line reference for future developments. As such, this paper reports a relatively successful transfer of theory into practice.

Keywords Data Envelopment Analysis · Healthcare · Pricing

It's not what you do, it's the way that you do it!

So goes the old adage, often occurring more often than proponents of more rational approaches would wish. This paper describes the use of Data Envelopment Analysis (DEA) in setting the health funding allocations to New Zealand Health Providers by means of a National Pricing Framework. The total pool of funds to which DEA was applied total about NZ$ 2.6 billion, making this one of the larger (in dollar terms) applications of DEA to date.

DEA aficionados would not usually expect their method to evoke the kind of extremely negative and emotive reactions to DEA, expressed by participants in the NZ Health sector. The manner in which it was employed resulted in "efficient" and "inefficient" health providers i.e., "winners and losers," provoking strong resistance to DEA at times during the process. What also made this work different from many commercial projects, was the rigour and adherence by the participants to "scientific method" and evidence based modelling. It is unusual to

P. Rouse (✉)
Department of Accounting & Finance, The University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: p.rouse@auckland.ac.nz

R. Swales
Director, Meridien Consulting Limited, Auckland, New Zealand

find a sizable group of middle to senior managers in a public sector organisation discussing "hypotheses, acceptance criteria, mathematical modelling and regression results" as a regular part of their meetings. However, this was a regular occurrence during the development of the NZ National Pricing Framework in 1999 which forms the focus of this paper.

The paper proceeds as follows. The first section surveys some of the relevant literature and describes the NZ health sector and background. The national pricing framework and its use of DEA are outlined next with illustrative results. Section 3 discusses the interim stages and the attempts to discredit DEA and the framework. Section 4 describes some of the enhancements made to the model and the testing and validation processes performed. The outcome of the process is contained in Section 5 with a postscript for events subsequent.

1. Literature review and the New Zealand health sector background

During the period 1978–2001, over 3235 books and papers on DEA were produced (Tavares, in press) covering an extensive range of topics (1253) originating from 42 countries; this international flavour has been recently addressed in Cooper et al. (2004). While many published articles tend to be studies by academic researchers, there have been relatively fewer articles documenting the use of DEA in practice. Some areas where there has been ongoing investigation and use of DEA include deregulation of the electricity sector (Edvardsen and Forsund, 2003; Roos, 2003), banking (Asmild et al., 2004; Barr et al., 2002; Sherman and Ladino, 1995; Berger and Humphrey, 1997), airlines (Rouse, Putterill, and Ryan, 2002), site evaluation (Takamura and Tone, 2003) and health (Hollingsworth, 2003). The Australian Productivity Commission (1997) has produced a compilation of case studies describing the application of DEA in various public service areas such as hospitals, prisons, police patrols and schools. Recently several works have been or are in the process of being produced specifically targeted at practitioners e.g. Cooper. Seiford and Zhu (2004) and Coelli et al. (2003).

Focusing on health, Norlander and Roos (1998) describe an ongoing study in Sweden to implement Malmquist productivity indexes for pharmacies. Commencing in 1989, this study must rate as one of the longest running DEA applications and the authors note a number of outcomes to date. First, the Malmquist indexes have proven to be "very difficult to manipulate and they provide correct incentives for productivity, costs, quality and profitability" (*ibid)*. This provides some evidence for the analytical argument by Agrell and Bogetoft (2001) that the super-efficient DEA model overcomes most problems in implementing incentive plans. Second, considerable efforts were made in providing presentations of results, communicating with pharmacies and training managers. This appears to have been a successful implementation strategy: "implementation was not a directive issued by the management in order to meet external demands for greater goodwill or profit. Instead, it was the Corporation's own organisation, more particularly its controller unit, that called for better monitoring and control instruments, a wish supported by the management" (Norlander and Roos, 1998).

A recent report by the World Health Organisation (2000) examined the measurement of efficiency of health services across the globe. After considering various methods of efficiency estimation including DEA and stochastic frontier analysis (SFA), the WHO dismissed these in favour of a fixed effects model. A number of authors responded to this. Hollingsworth and Wildman (2003) used DEA and SFA to obtain new insights not provided by the WHO report while Richardson, Wildman, and Robertson (2003) produced a detailed critique of several major features of the WHO approach. Lauer et al. (forthcoming) applied DEA to the WHO data exploring alternative weight constraining combinations, concluding that allowing countries limited freedom to choose their weights on the metrics, went some way to respecting their varying circumstances.

The above is indicative of the growing use of DEA in the "real world" but more stories are needed of successful (and unsuccessful) transfers of this technology from the academic world into practice. This provides the motivation for this paper and NZ health setting to which we turn next.

NZ has had a predominantly public funded health system since 1936 after a century of gradual progression to dominant national provision. Secondary care provision (i.e. hospitals) is essentially free, but primary care, while subsidised, is provided at a cost to the patient. As elsewhere, NZ has experienced significantly increased health expenditure due to increased scope and complexity of services, as well as high levels of demand especially in elective surgery (e.g. hip replacements).

Other sources of funding apart from Government have emerged in the form of private health insurers, and the past decade has seen the emergence of a number of private providers for niche markets. This same period saw major reforms imposed on the public sector, central to which was a strong user-pays doctrine. Although health may not have been a prime Government focus, the NZ publicly funded percent of total health expenditure fell from 85.6 percent in 1988 to 77 percent in 1998 (c/f 84.2 percent to 84.9 percent in the UK). Nonetheless, the public system remains as the dominant provider.

In 1986 the Health Benefits Review recommended the creation of Regional Health Authorities to identify the health needs of their region but that these needs should be met by public, private or voluntary agencies. Each regional health authority would contract for health services on the basis of need, quality and value for money.

"Under our proposal area health boards would become more like the boards of public companies. They would be able to concentrate on running efficient services, helped greatly by a payment system which would give them an objective value of the services they provide. These prices would drive signals through the system, causing rapid improvements in resource use and clinical practice." (Gibbs, Fraser, and Scott, 1988).

Price information was a key plank in the NZ reform platform: "Unlike most other enterprises in the New Zealand economy, public hospitals do not know the value of their output. Thus they have only one side of the necessary management information," (Gibbs, Fraser, and Scott, 1988).

At this time, the NZ public sector was changing from line item budgeting to the broader accountability framework encompassing outcome, output, input distinctions articulated by Ramanathan (1985) and embodied in the Public Finance Act 1989. The emphasis on service delivery is patent in the proposed reforms: "no major improvements in our system can be achieved without paying hospitals and other health providers for the specific treatment they provide instead of reimbursing them for what they spend" (Gibbs, Fraser, and Scott, 1988).

In the early 1990s NZ embarked on internal market reforms of its national healthcare system by restructuring the public health sector around funding and providing roles. Four regional health authorities (RHAs) were created to act as purchasers of healthcare from public or private providers and most public providers of health services were reorganised into 23 Crown Health Enterprises (CHEs) that would operate as commercial entities.

Limited co-ordination among the RHAs led to pricing disparities and conflicting policy decisions. At the same time, many of the CHEs were experiencing substantial deficits. In 1996 the decision was made to merge all four RHA purchasers into a single health funding authority (HFA) and to implement a national price framework. CHEs were subsequently renamed HHS (Hospital and Health Services).

In 1997 the Government signalled their intention to 'cap' HHS spending levels leaving the sector to face the problem of how to handle the accumulated deficits and ongoing situation. The fundamental question faced by the HFA was *what should be the level of expenditure for*

the volumes contracted with the HHS's. Efficiency and effectiveness were bywords of the accountability regime prevalent in the public sector, so it was natural that questions were also raised concerning the level of efficiency/inefficiency in the HHS system. The solution proposed by the HFA to the question posed above was that *the level of expenditure ought to be what an efficient provider would incur.* This raised the problem of how the HFA was to identify an efficient provider.

Within the HFA, a small group of individuals was successful in persuading the HFA and other key stakeholders including Treasury that DEA was the appropriate tool to identify efficient and inefficient providers. They proposed two steps: first identify the level of inefficiency in the hospital sector for the 1996/1997 year and second, transfer an additional amount from the Capital Vote and use it to fund HHS over the following three years to enable HHS's to improve their efficiency and move to eliminating their deficit positions.

The model was built around a single input (total expenditure) and outputs relevant to the specific service category, of which there were five: medical/surgery, pregnancy/childbirth, community health, disability support services and mental health. Each service category had its own DEA model and a level of efficient expenditure was determined for each category. For example, if the total actual expenditure for medical/surgery was $1 billion and efficiency was calculated as 95 percent, then the inefficient amount would be $50 million.

The amount to be transferred was calculated as three years of the DEA annual estimated inefficiency. For example, if the total annual inefficiency across the five categories summed to $90 million, the transfer would be $270 million. This came to be known as the 'deficit switch'.

This initial project came under severe criticism. Much of this was valid and constructive, especially that concerning process and planning issues. Some of the key issues were as follows:

- Only contracted volumes were used. Generally there were material differences between contract and actual volumes.
- There was no systematic selection of outputs and anecdotal evidence suggests that some of this was ad hoc.
- The technical work was done by a small team of three to four people. Most of the sector were not involved (except to supply data) and did not understand DEA.
- The DEA models were run using ordinary LP software with a sole focus on efficiency scores. No information was extracted concerning peer units or targets.
- The case or cost weights used in the model had been adapted from Australia and major differences existed in NZ that should have been taken into consideration.

Noting these criticisms, the HFA decided to continue the project the following year but this time to focus on the development of a national pricing framework to provide prices for the 1999/2000 financial year. Since the deficit switch had used DEA to determine an efficient level of expenditure for each service category, it appears that the next logical step was to use this to set "best practice" prices.

2. The national pricing framework and evidence based model

The 1998/1999 pricing work was organised around three main goals:

1. A general strategy to improve health sector performance.
2. The application of scientific methodology and methods to the decision analysis.

3. More risk management processes to assure sustainability, feasibility and flexibility in setting final prices.

The creation of a national pricing framework was a major element in the HFA's strategy. As a sole purchaser, it wanted to ensure that the sector responded to price signals and directed resources to areas according to economic rationale as well as clinical influence. The period of operating deficits had convinced the HFA that HHS should adhere to contracted volumes, refraining from over-production and changes in case mix volumes. There was a strong belief in price consistency so for example, the treatment price for a broken leg should be the same in a South Island hospital as in North Island hospitals. Given that this was a monopsonistic market, consistent national prices based on benchmarking "efficient providers" were believed to be as close to market conditions as could be obtained.

This form of benchmarking for improvement using "efficient provider" prices accords well with other perspectives concerned with continuous improvement such as target costing. The target or allowable cost is calculated as the market selling price less desired profit: "We no longer look to the development of a total unit cost in order to help determine a selling price. Instead we use the selling price we believe the market will allow to help us determine the cost that the market will allow. Peter Drucker has referred to this concept as price-led costing as opposed to cost-led pricing." (Ferrara, 1995).

The HFA also wanted to follow some guiding principles: transparency, objectivity and openness especially in contract negotiations. While some degree of political leverage would always exist in contract negotiations, there was a desire to relegate this as far as possible to arguments at the margin and at the final stages.

In addition, the pricing framework had to meet conditions for sustainability, feasibility and flexibility. The HFA did not want HHS to return to the previous situation of deficits throughout the sector. At the same time, public needs had to be met (without translating into increased waiting lists) at both a local and national level. A major complication was to provide for particular differences among HHS, chief of which was "tertiariness". Given NZ's small physical size, specialist services are usually located within particular providers who needed to be compensated for the more complex nature of the tertiary services provided.

Hospital services were split into five service categories shown in Table 1 with their respective 1997/1998 expenditure. A number of items were excluded from the DEA application (e.g. iron lung services, liver transplants) totalling $165 million.

The DEA models

The production model for each service category consisted of a single input (expenditure) and two or more outputs corresponding to each category's major activities. Table 2 lists the main outputs for each service category together with their goodness of fit (r^2) from an OLS regression of the single expenditure (input) on the set of outputs under each major service

Table 1 Annual expenditure in 1997/1998 for the major service categories (abbreviations shown in brackets) provided across 23 HHS

Service Category	$ Million 97/98
Medical and Surgical (Med/Surg)	1,610
Pregnancy and Childbirth (Preg/Child)	221
Community Health	150
Mental Health	334
Disability Support Services (DSS)	260
Totals	$2,575

Table 2 Outputs for the five service categories

Medical & surgical	Pregnancy & childbirth	Community health	Mental health	Disability support services
(i) cost weighted discharges	(i) cost weighted discharges	(i) district nursing and home visits	(i) forensic bed days	(i) assessment treatment and rehabilitation bed days
(ii) cost-weighted non-DRG volumes	(ii) number of outpatient attendances	(ii) school dental treatments	(ii) clinical rehab/ in-patient detoxification bed days	(ii) number of needs assessments
(iii) specialist treatments attendances		(iii) population 0–15 years served by well child service	(iii) non-acute bed days	(iii) aged continuing care bed days
			(iv) contacts mental health services	
			(v) contacts drug and alcohol services	
$r^2 = 99.2\%$	$r^2 = 97.9\%$	$r^2 = 74.1\%$	$r^2 = 80.42\%$	$r^2 = 93.7\%$

category. The first two service categories had good statistical results for goodness of fit with lesser results for the remaining categories. As the work proceeded, a decision rule was adopted that a service category would be included only if its *r*-squared exceeded eighty percent. This resulted in community health being excluded from the analysis.

As is common in health organizations, volumes are adjusted for case severity typically using cost weights. Various names such as Purchase Units, DRGs (diagnostic related groups) or HRGs (health resource groups) are used for health "products" which tend to be nationally (or state) defined and (using the UK NHS description) "group together treatments that are clinically similar, consume similar quantities of resources and are likely to be similar in cost". The origins of case mix costing appear to lie in the US with the introduction of the prospective payment system by Medicare in 1983 (Hill, 2000). This was phased in over a five year period with reimbursement based on a combination of hospital-specific and federal rates times the relevant DRG weights.

The determination of case weights is typically performed using case mix costing systems as described by Jackson (2001): "Each 'job' requires a unique assemblage of materials and thus costs, with the 'job' being each patient's care... Rather than a 'top-down' allocation of costs to DRGs, clinical costing systems use a 'bottom-up' approach. Each time resources are used on behalf of a patient (an X-ray, a day of stay), the utilisation and related costs are 'tagged' to the patient's medical record number. When the episode is completed and the DRG assigned, these costs can be aggregated for analysis at the DRG level". Relative differences in cost at DRG levels inform the calculation of case weights.

Each output in the NZ National Pricing Framework therefore comprises purchase unit volumes aggregated using cost weights. A very different approach is described by Olesen and Petersen (2002) where the cost weights are used to construct assurance regions to determine a cost function based on 483 outputs each corresponding to a DRG.

Most New Zealand hospitals do not have detailed costs at DRG levels. Instead, the Sector relies on the Australian cost weights and in particular, the State of Victoria WIES (weighted inlier equivalent separations) weights. Data was collected from NZ HHS that had detailed costs to enable comparisons with the Australian WIES weights. The project team doing this comparison reported a high correlation (0.95) between the WIES weights and the NZ cost

data. One major difference, however, that had to be corrected for was the cost of blood, which is free for Australian hospitals but is charged for in NZ by the NZ blood service. Another problem was the unavailability of cost weights for most non-inpatient purchase units, particularly in mental health, community health and DSS. In the absence of cost weights for these purchase units, the default was unity, which raises at least two problems. First, the absence of cost weights means that although several of these capture volume of the activity, they do not reflect the intensity of the service provided. Second, each episode may comprise a number of activities that have varying intensity.

While these were (and still are) problems when trying to do this type of analysis, there may be mitigating conditions that overcome these problems.

1. Intensity may not be such an issue if most volumes are of average intensity and low variability. Or if each episode has low resource use or uses a constant interval of time e.g. half hour slot.
2. The process is managed so that variability is balanced and a constant rate of volume is achieved on an individual work cycle, e.g. district nurses may be scheduled to see ten patients per day, each of which may have varying requirements. In order to achieve this, variability must be balanced.

Furthermore, the relatively high "fit" for four of the five service categories was considered to provide a sufficient basis to proceed with the pricing model.

The pricing model was based around a ratio of efficient expenditure to cost weighted volumes and is described in Fig. 1 under four stages:

Part A setting the numerator: The data for a single expenditure input and multiple outputs for the 1997/1998 year were collected for each HHS for each service group. DEA was used to calculate the individual efficiency scores for each HHS in each service group. These efficiency scores were applied to the actual 97/98 funding for each HHS to obtain the overall efficient level of funding for each service group. The HFA described this as the *aggregate historical base funding.* The numerator can be expressed for each service category k as:

$$\sum_{h=1}^{H} Eff_{kh} \times Exp_{kh}$$

where *Eff* denotes the DEA efficiency for each HHS h multiplied by its expenditure.

The aggregate historical base funding is obtained from the year for which data is most recently available (typically two years preceding the target funding year). In the case of 1999/2000 this was the 1997/1998 financial year. Due to the two-year delay, adjustments were made to the aggregate historical base funding to take into account price inflation relevant to the health sector and changes in capital cost.

Part B setting the denominator: HFA purchase units (PUs) were adjusted for cost weights (where available) to obtain a total number of equivalent units. In setting prices, the NZ health sector also had several proposed pricing adjusters to take into account certain environmental factors that potentially affect individual HHS service delivery and cost structure. These included diseconomies of scale, tertiary adjusters, special systems and clinical developments, and rurality. The model assumed a medium-sized secondary hospital as the default base with

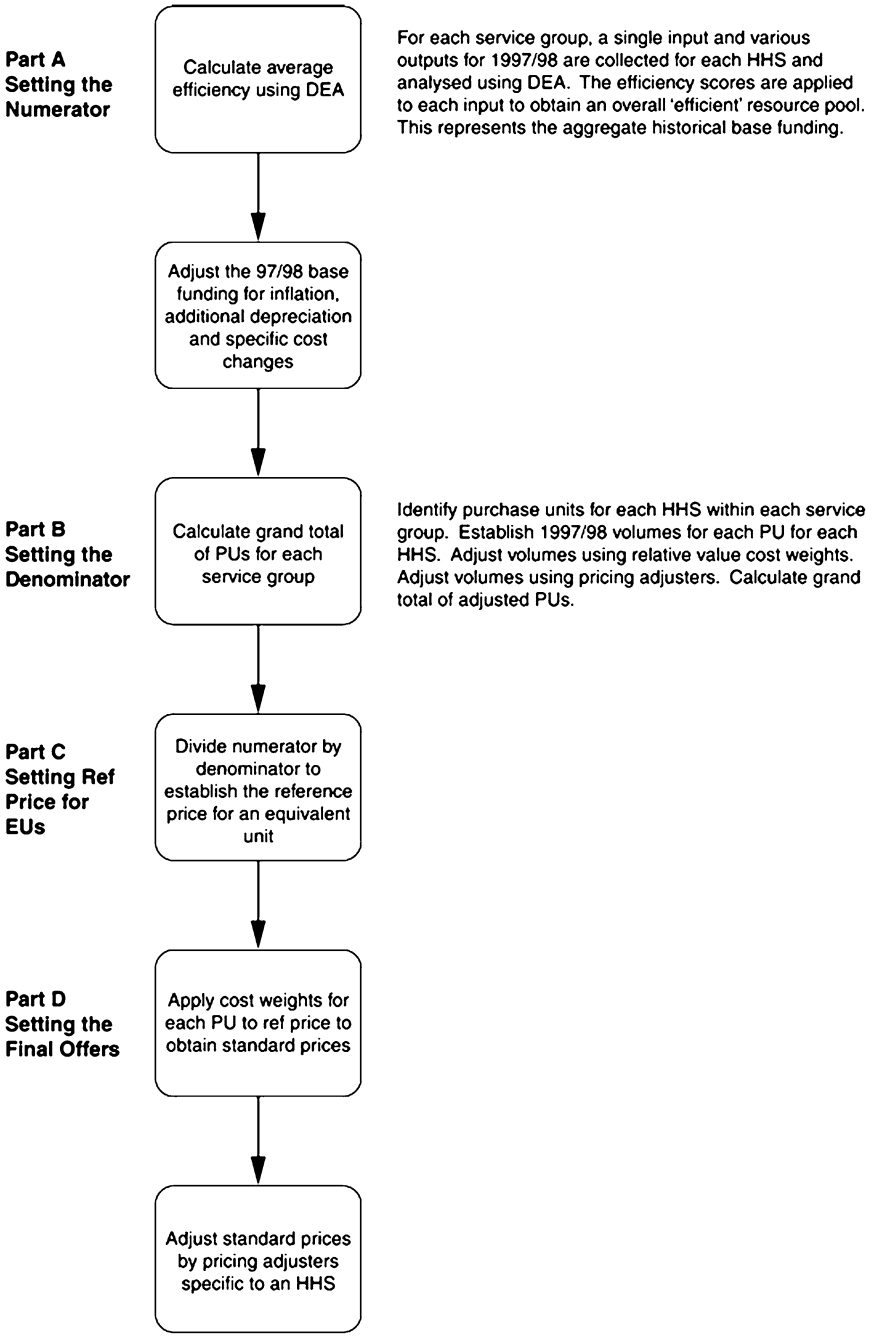

Fig. 1 Schematic of process for setting national prices

adjustments to be made for deviations from this standard. The denominator can be expressed for each service category k as:

$$\sum_{h=1}^{H}\sum_{r=1}^{s} y_{khr}$$

$$\forall y_r \quad s.t.$$

$$y_r = \sum_{i=1}^{I} y_{ri} cw_i$$

where y_r denotes the outputs ($r = 1, \ldots, s$) and $y_{ri}cw_i$ denotes the cost weighted volumes of the purchase units i for each output r.

Part C setting the reference price for the Equivalent Units: The numerator is divided by the denominator to calculate a base price per equivalent unit for each service category. This is multiplied by the cost weight per purchase unit to obtain its respective price.

$$\frac{\sum_{h=1}^{H} Eff_{kh} \times Exp_{kh}}{\sum_{h=1}^{H}\sum_{r=1}^{s} y_{khr}} = bpeu \quad \text{(base price per equivalent unit)}$$

Price per purchase unit $i = bpeu \times cw_i$

Part D setting the final offers: For each individual HHS, the offer for the 1999/2000 contract funding was to comprise the product of contract volumes and the purchase unit reference prices with adjustments for any environmental factors particular to that HHS.

While this multi-stage framework appears straightforward, parts of it are relatively complicated requiring a number of refinements. To better understand these refinements and the debate that surrounded each, it is important to appreciate that this was essentially a resource allocation process whereby the pricing framework when applied to the contracted volumes would determine how much each HHS received from the HFA. Consequently, there were incentives for the HHS to increase the numerator (aggregate historical base funding) as much as possible with an opposite incentive for the HFA.

In order to better manage these potential conflicts, a pricing project group was established with representatives from the HFA, the HHS and other stakeholders being the Ministry of Health and the NZ Treasury. This was a large group (between 40 to 50 in size) with several smaller work groups to focus on specific tasks e.g. pricing adjuster, the DEA model. The DEA work was carried out by a small technical team (2 to 4 persons) assisted by the authors acting as technical advisors.[1]

3. Refinements to the model

In the earlier work for the deficit switch, variable returns to scale were assumed in determining the efficiencies for each service category. For this round, the HFA were strongly of the view

[1] We did not become involved as technical advisors until the pricing work was well under way at the end of 1998.

that constant returns to scale was the appropriate model for the med/surg and preg/child service categories. Given that there were already separate funding pools for diseconomies and hospitals in rural areas, the HFA believed it would be paying twice for the same issue if variable returns to scale were used. Furthermore, there was a strong wish to treat diseconomies and rurality as potential price adjusters to be evaluated in the technical work. The second major reason was based on the results of testing by the technical team who set certain criteria for determining constant returns to scale. These included visual inspection of scatter plots of outputs to input as well as a number of regression results such as constant near to zero and insignificant, high *r*-squared and high model significance. Med/surg and preg/child met these criteria and constant returns to scale was held to be the most appropriate model for the DEA work in those categories.

The intention with price adjusters was to recognise that a level playing field did not exist among HHS and that differences between individual HHS environments affected their service delivery, configurations and resource consumption. Hence a price adjuster can be viewed as the unfavourable impact on a sub-group of HHS providers of a feature or influence that is uncontrollable (in the short to medium term) and arises from their environment or service configuration mandated or agreed with Government, where the impact is demonstrably different from the overall population of HHS providers. Potential price adjusters agreed by the HFA and HHS were ethnicity in respect of Maori and Pacific Islanders, diseconomies, rurality and tertiary.

Several work groups were established to devise suitable proxies for each pricing adjuster and evaluate any data collection and measurement problems. Measures of ethnicity were relatively straightforward and consisted of the relative populations of each ethnic group within an HHS area. Attempts to determine suitable proxies for diseconomies were unsuccessful and eventually abandoned in favour of the argument that the other pricing adjusters were to a large extent, capturing the increasing returns to scale (rurality) and decreasing returns to scale (tertiary) diseconomies.

The argument for a rurality adjuster is based around a provider incurring higher costs due to a widely dispersed population. This relative isolation invokes a number of problems in risk management and meeting the demands of professional and Ministerial standards that affects maintenance of capability and service indivisibilities, e.g. the provision of 24 h emergency services. The proxy for this was calculated using mesh blocks and measuring the distance from the centre of each mesh block to the nearest health facility, weighted by the population.

NZ is not alone in distinguishing tertiary services from secondary. Scandinavian countries and Canada also identify this as a separate feature. The Joint Policy and Planning Committee of the Ontario Ministry of Health and Long Term Care defines tertiary as follows: "hospital services provided to patients requiring complex treatment. Tertiary care frequently involves a wide range of services, equipment, or techniques that are specialised and expensive. Hospitals that provide tertiary level of care services are believed to have higher cost per weighted case. The extra costs incurred by tertiary centres are associated with the variable utilisation of specialised programmes and equipment; the high proportion of transfer cases; and the treatment of more complex cases".

Two proxies were investigated for the tertiary price adjuster: clinical complexity level (CCL) and out-of-catchment tertiary cases. The first proxy (CCL) proved unsatisfactory and produced results that did not align with the sector's overall perception of which hospitals were tertiary and which were not. The second proxy tries to capture the movement of complex cases between hospitals. The assumption is that patients needing complex and intensive treatment are transferred to a higher-level facility. As complex cases are more expensive to treat than non-complex cases, the argument is made that a hospital receiving a large number of complex

cases should receive extra funding to cover additional costs. It was decided that this proxy would be used to set the tertiary pricing adjuster, with adjustments made for local complex cases and holidaymakers.

The process adopted was to test for significance using ordinary least squares (OLS) regression analysis to gauge whether a price adjuster could legitimately be used in the price setting process. Initially the technical team tried to regress expenditure (for med/surg and preg/child service categories) on outputs plus the proxies for candidate price adjusters. Results were insignificant for all price adjusters. An alternative approach adapted from Fried et al. (2002) proved more successful whereby the radial slack from an initial DEA run for the two service categories was regressed on the price adjusters only. The tertiary price adjuster was highly significant with an r-square of close to sixty percent. None of the other pricing adjusters were significant with the exception of the proxy for the Pacific Island population in each region.

At this point the work diverged from Fried et al. where the results of the second stage regression are used to modify the inputs/outputs to rerun in a third stage. Instead, the pricing group decided that the tertiary proxy was to be used as an *additional* output in med/surg and preg/child. Since most of the tertiary HHS had obtained DEA scores of less than 100 percent under constant returns to scale, the effect of the additional tertiary output was to increase their efficiency, in some cases improving it to 100 percent. The increase in the aggregate historical funding (i.e. the numerator) from the additional output was to be attributed to a tertiary pool which would be paid to tertiary hospitals as a lump sum over and above their revenue stream based on priced volumes. Of the 23 HHS, only seven qualified with positive values for the tertiary proxy (measured by out-of-catchment flows).

There was some disagreement over the lack of significance for the Maori pricing adjuster as there was a general belief that Maori (and Pacific Islanders) presented themselves to secondary hospitals at advanced stages of illness, thereby requiring more treatment and consequently increased resources. This remained unresolved with suggestions that alternative approaches should be adopted following more detailed case studies.

Rurality was also a contentious issue for small rural HHS who argued that the statistical analysis was too coarse. Notwithstanding, the pricing group agreed (in the majority) to follow a self-imposed methodology where if the null hypothesis could not be rejected, the pricing adjuster would not enter into the pricing model. This excluded all the pricing adjusters except for the tertiary adjuster.

The pricing adjusters are interesting because they provide better insights into so-called scale diseconomy effects. The tertiary adjuster was interpreted to explain decreasing returns to scale since all the tertiary HHS were located at the high end of the size scale. The effect of adding the additional tertiary output was essentially to close the distance between the constant returns to scale frontier and non-increasing returns to scale. Conversely, if rurality had been accepted and treated the same way (i.e. as an output),[2] this could have been interpreted as explaining part of the increasing returns to scale diseconomy. Figure 2 shows these effects for both increasing and decreasing returns to scale.

As the technical work progressed, concerns arose over the issue of weight flexibility. Several members of the pricing group were unhappy with an HHS being able to ignore (i.e. zero or epsilon weight assigned to) an output. The technical team experimented initially with virtual weights but then decided to use price ratios constructed around the HFA's knowledge of cost variability for different purchase units within a service category. For example, the

[2] This is conjectural since the rurality proxy was not found to be significant and therefore failed the acceptance criteria.

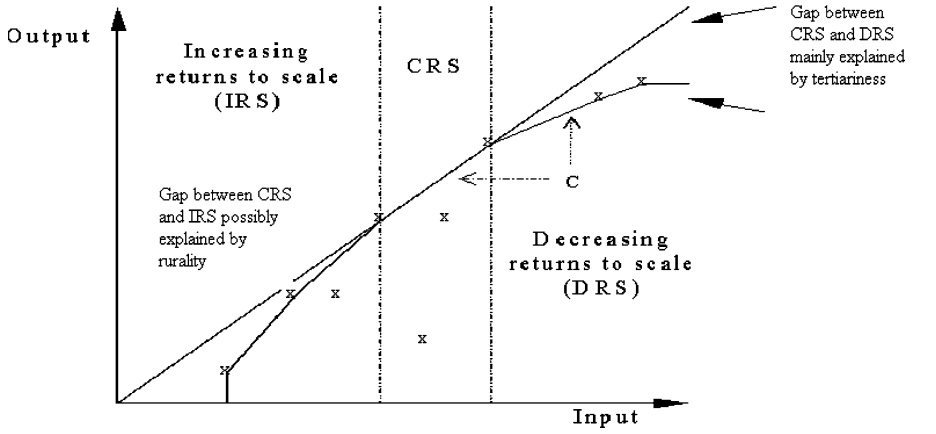

Fig. 2 Rurality and tertiariness as potential explanations of increasing and decreasing returns to scale

weights for the inpatient output in med/surg was set at a lower bound of 1 and upper bound of 5 to reflect the known range of cost variation within that activity. The effects of the weight constraints were very slight (less than 0.5 percent) but still had an impact in dollar terms as noted next.

The sheer size of the expenditure amounts imposed strict requirements on data validity and testing. While most researchers are not usually concerned over one or two percent differences in efficiency, each percentage change in this application could make a difference of millions of dollars to the efficient expenditure pool. Consequently both HHS and the HFA were keen to ensure that data quality and process quality were as high as could be obtained and that opportunities for gaming were either prevented or minimized. Some of the more obvious data checks included reconciliations to the annual financial statements, scatter plots of outputs to medical FTE staff and investigation of outliers. Those familiar with cost weights would be aware of the practical problems in assigning the correct weights to DRGs and their intermediation using length of stay. All of these procedures required active checking.

The DEA technical team used the Warwick software package and this output was compared with alternative DEA software and linear programming packages for particular checks (e.g. price ratios).[3] The Warwick software performed very well and all results were satisfactorily confirmed. Following the same procedure used in the deficit switch, sensitivity analysis was performed by varying the input and outputs by factors of five to ten percent. While this had some effect on individual HHS efficiency, the effect on the efficient pool was less than 1 to 2 percent, which was regarded as within acceptable limits. Personally we believed that this did not achieve very much but the pricing group was content with this test. A new refinement was the use of the super-efficiency modification to (i) identify any outliers in the peer reference group and (ii) signal whether any gaming was taking place by HHS shifting expenditure across service categories. With regard to the second reason, if an HHS knew that it was more efficient than the other HHS in a service category (e.g. preg/child), there was a possibility that they could shift expenditure out of another service category (e.g. mental health) in which they knew they would be inefficient. Provided the transfer did not make them inefficient in preg/child, the effect would be to increase the efficiency in mental health thereby increasing

[3] The DEA packages were IDEAS and software developed by the authors. The LP software used was STORM.

the efficient expenditure pool overall. While this might sound hypothetical, it must be borne in mind that the analysis was carried out in several stages, with results reported at each stage. The super-efficient results were therefore used to track changes in HHS efficiency at the different stages across the five service categories. No evidence was found of cost shifting but one HHS whose super-efficiency score exceeded 120 percent was treated as an outlier and excluded from the preg/child category. The effect of this was to increase the average efficiency and therefore raise the level of efficient expenditure for this service category. It could be argued that the HHS in question was in fact "best practice" since it provided the largest volume of maternity services in the country. Alternatively, it could be argued that it was unrepresentative of the general provision for preg/child across the country because of its extremely high volume of service provision. In this case, the HFA agreed that it would be treated as an outlier.

In addition, focus groups of about four to five people were formed to examine the results, model and financial reconciliations. These helped to improve the quality assurance of the work as well as extending some of the detailed knowledge through the sector since group members gained a greater understanding of the pricing model and its implications for policy objectives.

4. Criticism, resistance and failure

The technical work continued for almost three months and did not always proceed smoothly. One of the larger HHS appointed one of the Big Four Chartered Accounting firms as its own technical advisor. This same firm had performed a similar role at the time of the deficit switch and had been opposed to DEA. This view remained unchanged as evidenced by extracts from a report they provided half way through the work: "In our opinion, DEA is an unsuitable tool for performing efficiency analysis. Regression should replace DEA as the efficiency measurement tool". "Having to make a choice between CRS and VRS is a good example of why we consider DEA to be an inappropriate analysis tool". "It is our opinion that the process of generating prices based on so-called 'efficient cost' should be replaced with a process that is more aligned with accepted commercial principles". "Unfortunately the HFA have determined that efficient means 100 percent efficient provider rather than a more pragmatic definition of 'relatively efficient' ".

These comments were used by some to attempt to discredit the process, in particular DEA. A spirited defence succeeded in convincing the pricing group that the CA firm's report was seriously flawed, and could be ignored. Unfortunately, the damage had been done and a belief persisted in the sector for several years that DEA was defective. While Fig. 1 shows that DEA is only part of a complex model, the pricing work became known as the DEA model and the target for any criticism of subsequent funding allocations.

Having battled the critics and their advisors, the project continued to the final rollout of prices for each purchase unit and final checks to ensure that these prices made sense. The focus groups reported back, comparisons were made of the total revenue effects across all HHS using the prior year's total expenditure, there appeared to be general agreement about the process and results within the Pricing Group and everything was set for the release of the first NZ national pricing framework.

Overnight, a decision was made by a hastily convened meeting of the HHS chief executives and the HFA heads that using the National Prices for funding allocations was too risky. Instead they would merely roll over the previous year's allocations with a small increase for inflation. This was known as the CHA agreement (CHA being the Crown Health Association

comprising the CEOs of all the CHEs). The Pricing Project, which had involved over forty middle to senior health managers as well as several teams of consultants for almost four months was shelved, and replaced with a negotiated political settlement.

Postscript

Understandably, many members of the pricing group were surprised and to some extent concerned at this outcome, but life carried on. The following year a political squabble arose concerning the size of the tertiary allocation to one of the larger HHS (the same HHS that had engaged the CA firm and been active in obtaining the CHA rollover). A working group had been set up to examine the tertiary allocations rolled over from the previous year in line with the CHA agreement. The recommendations of this working group changed the basis for the tertiary allocation to this particular HHS, essentially halving the amount. In the ensuing argument, a review panel was established to adjudicate what the allocation should be for this particular HHS. Despite its previous opposition to DEA, one of the key pieces of evidence used by this HHS in support of its argument for a higher claim, was the tertiary amount determined by the previous year's DEA modelling! The decision of the review panel ultimately was based mainly upon the price model built upon DEA.

During the following three years, the national pricing framework was increasingly used in the Sector. In fact, to end on an optimistic note, the NZ Health Sector is in the process of redoing the National Pricing Framework and, having considered other alternatives, has decided to again use DEA as a key tool in estimating efficient expenditure levels. Due to major restructuring and a shift in fund allocation to population based funding, there is no longer the same tension between a central funder and individual hospital providers that previously existed and the new Pricing Framework is seen as both a benchmark for improvement and a means for pricing inter-district flows. Given this lessening of tension and greater collaborative atmosphere, it is to be hoped that the following anecdote would not be repeated.

> During the pricing work in 1999, a number of presentations were made at which DEA and the pricing models were explained. One of the presenters noticed that an attendee had scribbled throughout her notes the following:

DEAth

While researchers are always enthusiastic about the transfer of new technology and theory into practice, such transfers are often not without problems and resistance. There are real dangers if implementation of methods such as DEA is left in the hands of a small team of "technos" or analysts. Training and information workshops must be considered as part of the implementation process to reach both technical and general audiences. Making the "black box" transparent is never easy but essential if DEA is to succeed in adding real value to organisations outside of academia.

The New Zealand application of DEA has persisted since 1997 and there is a high probability that it will become a permanent part of the NZ health sector initiatives for improvement at both sector and hospital levels. Obtaining this level of acceptance has been an uphill struggle and the battle is by no means over. However, there are few within the NZ health sector who have not heard of DEA and a growing number who are considering using it for internal applications within their respective hospitals.

Reflecting about the environment in which the NZ model was constructed, the construction of the pricing model was a complex and at times, uncertain process. Although an emphasis was placed on following an evidence-based approach, unless the process is subject to agreed

and rigorously enforced control safeguards, there is the risk that the process is influenced by individual participants, special interest groups or key stakeholders. There is of course the possibility of political interference given that price setting processes in both commercial and governmental environments are based on a set of policy decisions some of which are explicit while others may be implicit or simply be in the form of un-discussable assumptions. It is evident that the construction of the pricing model was a major project for the sector and that project management processes played a significant part in shaping the quality of results produced and their use by participant organisations throughout the health sector. It is also important to note that the model was constructed in an environment of policy settings and constraints which was subject to change (and in fact did so following the change of government in 2000).

These observations highlight the need for rigour in both overall governance procedures and project management methods, particularly with respect to change control processes. In recent times there has been a shift in project management practice around the notion of locating projects within a "controlled environment" (Office of Government and Commerce, 2002) in which the rules for managing changes in project scope and the management of risk are planned for and actively managed as the project work is performed. Based on our experience with this project, this is an area where there is scope for improvement through the application of more structured processes.

Another area concerns the issue of how to educate all participants concerning model construction and how the outputs of the model can be used in practice. It is noteworthy how the Swedish pharmacy study described earlier placed such an emphasis on presentations, communication and training. Such skill building is essential to remove the "black box" syndrome often associated with more sophisticated technology such as DEA.

In summary, attention to governance, rigorous project management, communication and training must not be neglected if we wish to make a successful transition of sophisticated technologies such as DEA from academia into practice.

Acknowledgments We would like to thank two anonymous referees for their constructive suggestions and Knox Lovell for his enthusiastic encouragement.

References

Agrell, P.J. and P. Bogetoft. (2001). "DEA-Based Regulation in Health Care Systems." Paper presented at the XXI Spanish Congress of Health Economics, Oviedo, June 6–8.

Asmild, M., J.C. Paradi, V. Aggarwall, and C. Schaffnit. (2004). "Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry." *Journal of Productivity Analysis* 21(1), 67–89.

Australian Productivity Commission. (1997). *Data Envelopment Analysis: A Technique for Measuring the Efficiency of Government Service Delivery.* Industry Commission, Steering Committee for the Review of Government Service Provision Reports, November.

Barr, R.S., K.A. Killgo, T.F. Siems, and S. Zimmel. (2002). "Evaluating the Productive Efficiency and Performance of U.S. Commercial Banks." *Managerial Finance* 28(9), 3–25.

Berger, A.N. and D.B. Humphrey. (1997). "Efficiency of Financial Institutions: International Survey and Directions for Future Research." *European Journal of Operational Research* 98(2), 175–212.

Coelli, T., A. Estache, S. Perelman, and L. Trujillo. (2003). *A Primer on Efficiency Measurement for Utilities and Transport Regulators*. World Bank.

Cooper, W.W., L.M Seiford, E. Thanassoulis, and S.H. Zanakis. (2004). "DEA and Its Uses in Different Countries." *European Journal of Operational Research* 154, 337–344.

Cooper, W.W., L.M. Seiford, and J. Zhu (eds). (2004). *Handbook on Data Envelopment Analysis*. Boston: Kluwer Academic Publishers.

Edvardsen, D.F. and F. Førsund. (2003). "International Benchmarking of Electricity Distribution Units." The 23rd Arne Ryde Symposium, The Nordic Electricity Market, October 3–4, Lund, Sweden.
Ferrara, W.L. (1995). "Cost/Management Accounting: The 21st Century Paradigm." *Management Accounting* (December), 30–36.
Fried, H.O., C.A.K. Lovell, S.S. Schmidt, and S. Yaisawarng. (2002). "Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis." *Journal of Productivity Analysis* 17(1/2), 157–174.
Gibbs, A., D. Fraser, and J. Scott. (1988). Unshackling the Hospitals: Report of the Hospitals and Related Services Taskforce. Government Print, Wellington.
Hill, H.T. (2000). "Adoption of Costing Systems in US Hospitals: An Event History Analysis 1980–1990." *Journal of Accounting and Public Policy*, 19, 41–71.
Hollingsworth, B. (2003). "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care." *Health Care Management Science* 6(4), 203–218.
Hollingsworth, B. and J. Wildman. (2003). "The Efficiency of Health Production: Re-Estimating the WHO Panel Data using Parametric and Non-Parametric Approaches to provide additional information." *Health Economics* 12, 493–504.
Jackson, T. (2001). "Using Computerised Patient-Level Costing Data for Setting DRG Weights: The Victorian (Australia) Cost Weight Studies." *Health Policy* 56, 149–163.
Lauer, J.A., C.A.K. Lovell, C.J. Murray, and D.B. Evans. (forthcoming). "World Health System Performance Revisited." *BMC Health Services Research*.
Norlander, N.O. and P. Roos. (1998). "Implementing the Malmquist Productivity Index: The Case of the National Corporation of Swedish Pharmacies." Essay 4. In R. Färe, S. Grosskopf, and R.R. Russell. (eds). *Index Numbers: Essays in Honour of Sten Malmquist*, Kluwer Academic Press.
Office of Government and Commerce. (2002). *Managing Successful Projects with PRINCE2*. HM Treasury, London, United Kingdom.
Olesen, O.B. and N.S.C. Petersen. (2002). "The Use of Data Envelopment Analysis with Probabilistic Assurance Regions for Measuring Hospital Efficiency." *Journal of Productivity Analysis*, 17, 83–109.
Ramanathan, K.V. (1985). "A Proposed Framework for Designing Management Control Systems for Not-For-Profit Organizations." *Financial Accountability & Management*, 1(1), 75–92.
Richardson, J., J. Wildman, and I.K. Robertson. (2003). "A Critique of the World Health Organisation's Evaluation of Health System Performance." *Health Economics*, 12, 355–366.
Roos, P. (2003). "Measurement of Output of Electricity Distribution." The 23rd Arne Ryde Symposium, The Nordic Electricity Market, October 3–4, Lund, Sweden.
Rouse, P., M. Putterill, and D. Ryan. (2002). "Integrated Performance Measurement Design: Insights from an Application in Aircraft Maintenance." *Management Accounting Research*, 13, 229–248.
Sherman, H.D. and G. Ladino. (1995). "Managing Bank Productivity using Data Envelopment Analysis." *Interfaces*, 25(2), 60–73.
Takamura, Y. and K. Tone. (2003). "A Comparative Site Evaluation Study for Relocating Japanese Government Agencies out of Tokyo." *Socio-Economic Planning Sciences*, 37, 85–102.
Tavares, G. (in press). "A Bibliography of Data Envelopment Analysis." *Socio- Economic Planning Sciences*.
World Health Organisation. (2000). *The World Health Report: Health Systems Improving Performance*. World Health Organisation, Geneva, Switzerland.