The Economic and Social Review, Vol. 30, No. 2, April, 1999, pp. 175-196

Efficiency in Northern Ireland Hospitals: A Non-parametric Analysis*

DONAL G. McKILLOP
Queen's University of Belfast

J. COLIN GLASS
CHRISTINE A. KERR
GILLIAN McCALLION
University of Ulster

Abstract: The study examines the technical, scale and size efficiency of acute hospitals in Northern Ireland over the six-year period, 1986-1992. The efficiency estimates are used to investigate whether the empirical evidence supports the subsequent current rationalisation policy for hospital provision in Northern Ireland. Non-parametric analysis is used to measure the efficiency of larger and smaller hospitals relative to best practice. The results cautiously support the current policy of expanding larger hospitals and restructuring/closing smaller hospitals, but also indicate that the expansion of large hospitals may not yield substantial efficiency gains.

I INTRODUCTION

Since the mid-1970s successive governments in the UK have made continuous efforts to find ways of improving efficiency and curtailing expenditure in the National Health Service. New approaches to public sector management were introduced in the 1980s characterised, among other things, by contractual or quasi-market forms, increased delegation of resource decisions,

*The authors are indebted to the Northern Ireland Health and Social Services, Strategy and Intelligence Group, Analytical Support Branch, for providing data.

more explicit and measurable standards of performance, and weakening trust in professionals while strengthening the hand of managers (Jones and Dewing, 1997). The concept of an internal market was a further radical step in the pursuit of efficiency and cost reduction, whereby "competition with other hospitals where it is effective should also constrain costs" (Department of Health, 1989). In addition, the government felt that delegating responsibility as close as possible to the point of delivery would improve the performance of hospitals.

It is against this backdrop that the present study examines the levels of technical, scale and size efficiency of a sample of 23 Northern Ireland hospitals over the six year period, 1986-1992. The period under examination pre-dates the granting of trust status to certain hospitals in Northern Ireland. It also pre-dates the more recent strategy of concentrating hospital services in six large hospitals, supported by efficient medium-sized hospitals, with the associated closure of a number of smaller hospitals. Consequently, the current analysis offers an insight into the productive efficiency of Northern Ireland hospitals prior to the granting of trust status, as well as a platform to assess whether the ensuing re-organisation of hospital provision was warranted.

The paper employs the non-parametric approach of Färe *et al*. (1994) which utilises the observed inputs and outputs to construct the best practice reference units as a convex hull in input output space without estimating parameters. This permits us to measure the efficiency of input usage in producing a given level of outputs. A number of reasons can be put forward for the adoption of the non-parametric approach. First, it provides multi-input, multi-output efficiency measures that relate to best practice as opposed to average practice. Although on this count it must be noted that best practice is relative to the empirical sample of hospitals. Second, it does not require the use of a pre-specified functional form for technology nor distributional assumptions about error terms. This necessarily implies that inefficiency measurement and misspecification errors are not confounded. However it should be emphasised that being nonstochastic it makes no attempt to distinguish the effects of noise from those of inefficiency, so that inefficiency may incorporate some degree of noise. Third, the approach does not require the imposition of a particular behavioural assumption, such as cost minimisation, yet the results have a straightforward cost interpretation.

The non-parametric approach has been used to measure hospital efficiency in the recent studies by Hollingsworth and Parkin (1995), Ferrier and Valdmanis (1996), Parkin and Hollingsworth (1997) and Rosenman, Siddharthan and Ahern (1997). Hollingsworth and Parkin (1995) argue strongly that traditional techniques to measure efficiency through the use of performance indicators and efficiency indexes are fraught with methodological and practical difficulties.

In terms of the format of the discussion the following approach is taken. Section II outlines the non-parametric frontier methodology used, while Section

III discusses both the data and the empirical procedures employed in the study. Empirical results and concluding comments are presented in Sections III and IV, respectively.

II THE NON-PARAMETRIC APPROACH

The Construction of Best Practice Frontiers

To begin, let us consider firms (hospitals) producing m = 1,...,M outputs from n = 1,...,N inputs. The production possibilities set S^t, for each period t = 1,...,T, which models the transformation of input vectors $x^t \in R^N+$ into output vectors $y^t \in R^M+$ is then given by $S^t = \{(x^t,y^t) : x^t \text{ can produce } y^t\}$. If S^t and its boundary (the production frontier) are known, then efficiency computations could be made relative to this efficient or best practice frontier. However, given that S^t is never observed, this means that we need to construct or estimate an empirical production frontier at t from an observed set of production activities.

To do this, we first regard the production vectors of the hospitals in the empirical sample as the observed set of activities. We then need to determine which of these activities constitute best-observed practice so as to enable the empirical production frontier to be formed as a combination of these best practice activities. In other words, in data envelopment analysis (DEA) terminology, the empirical best practice frontier is constructed as a (piecewise linear) envelopment of the observed data. Once this is done, the efficiency of each producer can then be measured relative to this empirical frontier, with producers operating on (and determining) this frontier being termed *efficient* and producers not operating on the frontier being termed *inefficient*.

In proceeding to clarify what is involved in the above steps, note that the technology S^t can also be modelled by the input requirement set

$$L^t(y^t) = \{x^t : (x^t,y^t) \varepsilon S^t\} \qquad t = 1,...,T \tag{1}$$

where the input requirement set $L^t(y^t)$ denotes the collection of all input vectors x^t that yield at least output vector y^t during period t. Since the empirical estimation of $L^t(y^t)$ is equivalent to estimation of S^t, the empirical best practice frontier can thus be formulated as the lower boundary of the estimated $L^t(y^t)$.

To see how $L^t(y^t)$ is empirically estimated, let us assume that we observe j = 1,...,J producers each using n = 1,...,N inputs, x^t_{jn}, in each period t = 1,...,T to produce m = 1,...,M outputs, y^t_{jm}, in each period t = 1,...,T. Given this panel data set and notation, an activity analysis model is then employed to estimate $L^t(y^t)$ from the observed inputs and outputs. Hence, the empirical construction of the piecewise linear envelopment of the input requirement set is thus given by

$$L^t(y^t) = \{x^t : y^t_m \leq \Sigma^J_{j=1} z_j y^t_{jm}, \quad m = 1,2,\ldots,M,$$
$$\Sigma^J_{j=1} z_j x^t_{jn} \leq x^t_n, \quad n = 1,2,\ldots,N, \tag{2}$$
$$z_j \geq 0, \quad j = 1,2,\ldots,J\},$$

where z_j is an intensity variable from activity analysis denoting the intensity levels at which each of the J activities or hospitals are (or might conceivably be) operated.

Measuring Productive Efficiency Relative to the Constructed Frontier

The input-based technical efficiency of any individual producer can be obtained for period t as

$$F_i^t(y^t,x^t) = \min\{\lambda : \lambda x^t \varepsilon L^t(y^t)\} \tag{3}$$

where $0 < F_i^t(y^t,x^t) \leq 1$.

Given the input measure of technical efficiency (3), a producer's input vector x^t (used in producing y^t) will be located on the efficient frontier when (3) has a value of one. If (3) has a value less than one, the producer is classified as an inefficient producer relative to best-observed practice. Measure (3) for producer (hospital) j is empirically calculated as the solution to the linear programming problem

$$F_i^t(y^t,x^t) = \min_{\lambda,z} \lambda$$

$$\text{st} \quad \Sigma^J_{j=1} z_j y^t_{jm} \geq y^t_{jm}, \quad m = 1,\ldots,M,$$
$$\Sigma^J_{j=1} z_j x^t_{jn} \leq \lambda x^t_{jn}, \quad n = 1,\ldots,N,$$
$$z_j \geq 0, \quad j = 1,\ldots,J \tag{4}$$

where the restrictions on the z variables imply that constant returns to scale is imposed on the reference technology. Variable returns can be readily imposed by restricting the sum of the z solution values to be unity. The ratio of the latter two measures yields a scale efficiency measure, with scale efficiency (inefficiency) obtaining when this ratio equals one (less than one). Where scale inefficiency obtains, further analysis permits this to be classified as either increasing or decreasing returns.

As noted above the ratio of the constant returns to scale and the variable returns to scale efficiency levels of a hospital measures its scale efficiency. Indeed, the inverse of this ratio shows how much the ray average productivity of inputs would increase if the hospital altered its observed output bundle to move to a different point on the production frontier. Maindiratta (1990), however, argues that in many cases it is impractical to make such alterations as producers are often in the situation of meeting a pre-specified output task. For example, a hospital may have to provide for a certain number of in-patients and out-patients during the financial year and, even though technically efficient, this level of provision might leave it on the decreasing returns to scale segment of the production frontier. In such circumstances it would be difficult to advise on a reduction in provision in order for the hospital to attain scale efficiency.

Maindiratta poses a different, though related, question: would there be greater economy in input use if the target output level were to be produced together by several smaller producers of identical size than if it were produced by a single producer? If the answer is yes, the producer under consideration is too large and, instead of operating as a single unit, it should be broken up into several producers of a smaller size. In this context he introduced the concept of size efficiency. Following the Maindiratta (1990) approach the size efficiency of hospital j in period t may be assessed by implementing the mixed integer programming model

$$\begin{aligned} &\min_{\delta,\theta,K} \ \delta \\ \text{st} \quad &\Sigma^{J}_{j=1}\, \theta_j y^t_{jm} \geq y^t_{jm}, \quad m = 1,\ldots,M, \\ &\Sigma^{J}_{j=1}\, \theta_j x^t_{jn} \leq \delta x^t_{jn}, \quad n = 1,\ldots,N, \\ &\Sigma^{J}_{j=1}\, \theta_j = K, \\ &\theta_j \geq 0, \quad j = 1,\ldots,J, \\ &K \text{ is a positive integer,} \end{aligned} \tag{5}$$

where the activity variable is defined as $\theta_j = Kz_j$ with $\Sigma^{J}_{j=1}\, \theta_j = K$. The solution to problem (5) yields an optimal value of δ (denoted δ^*, with $\delta^* \varepsilon (0,1]$) which indicates whether or not radial contraction of hospital j's observed input vector is possible (as shown by $\delta^* < 1$ or $\delta^* = 1$, respectively). Note, however, by incorporating K as a positive integer, with $\theta_j = Kz_j$, problem (5) assesses radial input contraction for hospital j in the context of permitting K *identical* (in terms of their input-output bundles) smaller hospitals to cumulatively produce at least the output task of hospital j. Hence the solution to (5) also yields an optimal (integer) value for K. It should also be noted that permitting the smaller hospitals to be identical is not the severe restriction that it first appears. Maindiratta (1990) has demonstrated that if we first construct a version of (5) that allows the K hospitals to have any (cumulatively feasible) non-identical input-output configuration, there exists a set of equal weights which enables us to transform this situation into K identical hospitals and obtain the same optimal solution (δ^*). Consequently, the K hospitals are not a priori restricted to being identical. Also, since the solution to (5) yields a combined size and technical efficiency measure, dividing this value by the variable returns measure of technical efficiency yields the Maindiratta size efficiency measure. Size efficiency (inefficiency) obtains when the latter measure yields a value of one (less than one).

In utilising the Maindiratta analysis, (5) will yield an optimal value of $K = 1$ for scale-efficient hospitals (as these are also size efficient). Problem (5) will also yield optimal $K = 1$ for scale-inefficient hospitals where the input scale inefficiency is due to increasing returns to scale. However, for scale-inefficient hospitals where the input scale inefficiency is due to decreasing returns to scale, problem (5) may yield either $K = 1$ (indicating a size-efficient hospital) or $K > 1$ (indicating a size-inefficient hospital).

III DISCUSSION OF DATA AND EMPIRICAL PROCEDURES

Northern Ireland Hospital Sector

The study was prompted by an interest in examining the performance of hospitals in Northern Ireland over the period leading up to the 1993 reforms of health and social care in the region, given the 1987-1992 Regional Strategy for Health (DHSS, 1987) goal of ensuring maximum benefit from the use of resources in the hospital sector. Indeed, the efficient use of resources had prime importance in maintaining hospital services in an environment where there was pressure to divert resources from secondary to primary and community care. For example, between 1987 and 1992 the share of secondary care in the total revenue expenditure of the Health Boards fell from 54.4 per cent to 50.6 per cent (Northern Ireland Audit Office, 1993). In addition, efforts were also made to rationalise the number of hospitals in the region, though this strategy has often been met by local opposition (Northern Ireland Economic Council, 1995, p. 39). However, between 1987 and 1992 the number of beds in the main acute specialties fell from 6,457 to 5,514 (Northern Ireland Audit Office, 1993).

The sample chosen for our study includes the (23) hospitals in the region which were classed as "acute" or "mainly acute" for the years 1986/1987-1991/1992. Of these 23 hospitals, most are non-teaching hospitals (only four are teaching hospitals) and 18 are located outside the two main urban areas. However, there is considerable variation in terms of hospital size. Over each of the six years of the period 1986-1992, these 23 hospitals consistently fell into the same three size classes (namely, 5 "large", 10 "medium" and 8 "small". Thus, for example in 1989, 5 were "large" (total inpatients and outpatients, TIO, ranging from 527,000 down to 120,000), 10 were "medium" (TIO ranging from 100,000 down to 53,000) and 8 were "small" (TIO ranging from 46,000 down to

14,000). Clearly, with the large variation in size and the relatively small total number of hospitals, it would be inappropriate to use 23 (hospital) observations to construct a reference technology for a given year. In addition, in attempting to utilise a meaningful disaggregation of inputs and outputs, the DEA would classify a number of hospitals as technically efficient simply as a result of the sum of the number of outputs and inputs being too large relative to the number of observations (see Nunamaker, 1985).

Window Analysis to Relieve Degrees-of-Freedom Pressure

Faced with a degrees-of-freedom problem, a standard solution is to run the DEA on pooled data (see Lovell, 1993) so as to ensure that the number of observations used in constructing the best practice frontier is sufficiently large relative to the sum of the number of outputs and inputs in the empirical model. Thus, if we use (say) years t–1 and t+1 in addition to year t, the frontier is then constructed from (3)(23)=69 observations, which is not problematic for the 4-output, 5-input empirical model. Each of the three observations for a given hospital (relating to its performance in t–1, t and t+1) is given an efficiency score by doing an individual DEA run for each observation relative to the frontier constructed from the 69 observations. An average of these three scores is then taken as the efficiency measure for that hospital in that three-year period. Moreover, since the six-year panel data set for the 23 hospitals over the period 1 April 1986–31 March 1992 can be divided into four three-year overlapping sets, 1986-1989, 1987-1990, 1988-1991 and 1989-1992, the pooling procedure also permits us to use the DEA "window analysis" technique for investigating efficiency change over time in a panel data context.

The pooling procedure solves the degrees-of-freedom problem, but it also gives rise to a further question. The three-year period gives 69 observations for frontier construction – but why not use 2-, 4-, 5-, or 6-year periods to get 46, 92, 115 or 138 observations, respectively? The answer to this question is essentially determined by how the empirical analysis attempts to avoid problems arising from differences in hospital size. For example, to cope with the hospital-size-mix problem, the study groups the 5 large- and 10 medium-sized hospitals together into one set termed the "larger" hospitals. This grouping of larger hospitals is particularly important in that policy makers have decided that future hospital services will be concentrated on this set of hospitals. As noted in the introduction, the rationalisation of hospital services will result in these services being concentrated in large hospitals supported by medium-sized hospitals. To obtain a "smaller" hospital-size grouping for comparative purposes, the 10 medium- and 8 small-sized hospitals were grouped into one set termed the "smaller" hospitals. Hence, if we use three-year pooling, we have 45 observations (per window) for frontier construction in the larger hospital set and 54 observations in the smaller hospital set. Since the use of two-year pooling would only give 30 and 36 observations, respectively, this was judged to be too small relative to the degrees-of-freedom problem.

What about the possibility of using more than three years in the pooling procedure? In answer to this, we need to note that the pooling procedure treats the annual observations of an individual hospital as distinct production units in the DEA analysis. This means that the relative performance of a hospital is empirically assessed by comparing it to its own performance in another year as well as to the performance of other hospitals. In doing this we are recognising that the efficiency of a producer can improve or deteriorate over time. Hence, since the frontier technology is formed as a combination of (observed) best practice activities, it is reasonable to include several realised input-output situations of a given hospital in the set of activities from which the best practice activities are to be determined. However, if some of the observations of each particular hospital are too widely dispersed in time, outlier problems and output-mix problems may emerge (see Newhouse, 1994). Consequently, in the pooling procedure, the study includes just three consecutive annual observations of each hospital within the set of activities available for constructing the piecewise linear envelopment of the data. By doing this for both the "larger" and "smaller" hospital sets, best practice frontiers were generated for each of the four three-year data windows corresponding to these sets.

The Issues of Hospital Output Quality and Hospital Case Mix

In addition to dealing with the problem of limited observations, the study also has to face the problem of specifying the outputs and inputs to be used in the empirical model. This, in turn, immediately raises the important issues of hospital output quality and hospital case mix. In assessing the relative efficiency of hospitals, it is clearly desirable that the resultant efficiency scores are not biased by a failure to adequately capture differences in output quality and case mix between hospitals. However, as discussed below, obtaining adequate measures of the quality of hospital services is not only extremely difficult but also requires the existence of a sophisticated, standardised data collection system across hospitals. Also, as noted below, to the extent that the quality of care issue and the case mix issue are interrelated, adequate adjustment for case mix differences via weighting can likewise quickly become a complicated matter. The following discussion elaborates with regard to these difficulties and also indicates the relevant checks that were available to the study with regard to the issues of output quality and case mix differences.

As recent studies indicate, adequate adjustment for output quality is an exceedingly complex matter, requiring very detailed standardised measures across hospitals. For example, as Thanassoulis *et al.* (1998) have noted, two aspects of the quality of a hospital's services need to be considered: (1) the quality of the various medical outcomes; and (2) the quality of the services as perceived by patients. Each of these immediately raises complex measurement problems. Ideally under (1) we need to know: (a) the exact medical condition, and accompanying health risk, of each patient immediately prior to their hospital treatment; and (b) the exact quality of medical outcome after this treatment as compared to the appropriate, expected best-practice quality of medical outcome on the "continuous spectrum ranging from a totally clean bill of health to death". Ideally under (2) we need to know patients perceptions of each aspect of the hospital service, as compared to a "reasonable" perception of best-practice service quality based on a "reasonable" understanding/expectation of hospital services and their associated waiting lists.

In addition, even if one had such ideal, detailed information, there is still the difficult problem of what weights to give (1) and (2), both in relation to a particular hospital service and between such services. Finally, to compound the complexity, there is also the problem of measuring input quality so as to be able to assess a hospital's relative efficiency, given both the quantity and quality of its inputs.

Faced with this complexity regarding adequate quality measures of hospital services, and a lack of appropriate, standardised data across hospitals, policy makers and researchers have been forced to largely rely upon output quantity measures complemented by various, partial checks on output quality. In the Northern Ireland case, for the sample period, the latter checks relate specifically to clinical quality rather than to both clinical and patient-perceived service quality. (Unfortunately, for the period studied, there is no data available for waiting lists at the hospital level that could serve as an indicator for patient satisfaction). Although recognisably incomplete, auditing reports of peer monitoring of comparable medical procedures across hospitals have found no evidence of major differences in clinical quality. Such reports indicate that while hospitals did differ with regard to, for example, the degree of consultant cover, they found no evidence of instances where the quality of care had been prejudiced (e.g. Northern Ireland Audit Office, 1993, p. 36). Consequently, with no better measure of the quality of hospital services being available for the sample period, and with no evidence of major differences in the quality of output, the current study uses quantity measures only (as do Söderlund *et al.* (1997), for similar reasons).

In addition to the issue of output quality, any measurement of the relative efficiency of hospitals must try to take account of differences in hospital case mix. Efficiency measurement inevitably involves some degree of aggregation of outputs (and of inputs) in order to have sufficient degrees of freedom for meaningful empirical analysis. This means that the resultant outputs are aggregate measures of heterogeneous cases whose mix differs both by hospital and over time. Clearly such variation in case mix could bias the results of relative efficiency measurement.

Ideally case mix should be addressed by using clinical specialty groupings combined with estimated weights. For example, weights may reflect the relative cost of a patient within a particular specialty, with the weights being the ratio of average cost per specialty to average cost of all specialties. Provided hospitals are using standardised cost allocation mechanisms within and across specialties, and provided cost measures are not biased by differences in clinical quality, such weights will provide a reasonable way of addressing case mix. Unfortunately, such detailed cost data is not available for Northern Ireland hospitals. Moreover, although measures provided by the National Casemix Office have recently enabled studies on English hospitals to address the case mix issue, there are still no plans to implement such measures (and such data collection) for Northern Ireland hospitals.

Faced with this data deficiency problem, the study implemented an important check to ascertain whether particular efficiency scores were likely to be affected by case-mix differences when utilising data that is not adjusted by some case-mix weighting procedure. To grasp what is involved in this check, note that if inefficient hospitals in each of our two hospital size groupings are being internally (to the respective group) evaluated mainly against efficient hospitals which are comparable in terms of case mix and size, this gives more confidence in using the efficiency scores. Hence, to check this, we used the non-zero optimal z values in (4). Since these represent basic solutions, indicating which hospitals are used to form the part of the reference frontier against which a given observation is being evaluated, we can ascertain which hospitals are being frequently used as efficient producers for peer comparison in the various DEA runs.

This check revealed that comparable hospitals were indeed being mainly used in the various (internal) comparisons. Not only does this result increase confidence about the reliability of the efficiency scores; it also has relevance to the quality of care issue. Hospitals, which are comparable in size and case mix have reasonably similar specialist cover which, hopefully, should ensure a similar quality of care in this relative efficiency measurement context. The importance of this point should not be overlooked. It is one thing to argue in isolation that neglect of output heterogeneity and quality may affect the efficiency measure of an individual hospital. However, it is much more difficult to argue that such neglect will create a systematic bias when comparable hospitals are being used as peers when obtaining the relative efficiency scores. Hence, while the implementation of this check can not be regarded as a perfect substitute for some case-mix weighting procedure, it nonetheless permits best-practice empirical analysis of the policy-makers' decision to rationalise Northern Ireland hospitals on the basis of similar data.

The above check, however, still leaves the possibility that some hospitals are not efficient in the true sense, but are only members of the efficient subset because they are outlier observations. If so, they would have an undue influence on the efficiency results and hence on the (external) comparison between size groupings. To check for this possibility, we followed the procedure outlined by Andersen and Petersen (1993) and Wilson (1995) for examining the observations located on the efficient frontier. This check gave us no reason to suspect that any of these observations were not efficient in the true sense. (In selecting the sample hospitals we deliberately omitted certain specialist hospitals so as to avoid this possibility.)

Model Specification Issues

As to which outputs and inputs to include in the empirical model, it is important to test the sensitivity of the efficiency results to changes in the input-output specification, even though one may have strong reasons for favouring a particular specification in the context of the available data. To do this, we ran a series of DEAs, with a gradual increase in the number of outputs and inputs used. The rationale for this approach, also used by Parkin and Hollingsworth (1997), was to determine whether there was any consistency between the different results. This sensitivity analysis revealed a considerable degree of stability in the efficiency scores (particularly with regard to which hospitals were on the efficiency frontier) when using different input-output specifications around the 4-output, 5-input specification eventually decided upon. Note that these different specifications involved different classifications of outputs and inputs (including capital and drugs expenditures) in various combinations.

As regards outputs, the empirical analysis aggregates the multiplicity of hospital outputs into four major categories, familiar in Northern Ireland hospital administration: general surgery; general medical; maternity; accident and emergency. In each category, output is measured by the total number of inpatients and outpatients (with regard to inpatients, we utilised discharges rather than length of stay in order to avoid confounding apparent differences in efficiency with occupancy rates). On the input side, inputs are aggregated into the following five categories: nursing staff; administrative staff; ancillary staff; specialists; and bed complement. The first three inputs are measured by full-time equivalent staff members. The specialists' input is measured by (annual) expenditure on specialists (deflated relative to base year 1990). Given the practice of sharing specialists between hospitals, made easier by the geographical proximity of hospitals in Northern Ireland, this expenditure measure is a more accurate gauge of the specialists input than the number of specialists. The bed complement is measured by the number of beds with the latter input being viewed as a proxy for physical plant, given the unavailability of net assets data for each hospital (see also Grosskopf and Valdmanis, 1987). Descriptive statistics of the input and output measures for the larger and smaller hospital groups are given in Table 1. For ease of exposition the information is only provided for the final time period 1989-1992.

Table 1: *Descriptive Statistics of the Output and Input Measures*

Measure	*Larger Hospitals, 1989-1992*			*Smaller Hospitals, 1989-1992*		
	Mean	*Min*	*Max*	*Mean*	*Min*	*Max*
General Surgery[a]	28,502	9,357	242,866	12,408	3,722	35,479
General Medicine[a]	22,627	7,036	119,590	9,971	2,712	24,253
Maternity[a]	12,460	3,425	59,828	6,182	1,204	59,828
A&E[a]	32,977	12,717	117,341	4,868	0	42,038
Nursing[b]	405	221	999	195	32	426
Administration[b]	67	33	378	29	4	116
Ancillary[b]	98	19	254	54	7	160
Specialists[c]	1,362	541	11,163	502	25	1,400
Beds	365	201	1,334	180	38	362

Note: a = total outpatients and inpatients; b = whole-time equivalents; c = £'000 deflated relative to 1990.

IV EMPIRICAL RESULTS

Before looking at the results in detail, we first highlight the major finding that the "larger" hospitals manifested higher technical and scale efficiency than the "smaller" hospitals throughout all data windows (as shown in Tables 2 and 3). Also, the "smaller" hospitals were found to have experienced a substantial deterioration in both technical and scale efficiency over the sample period.

"Larger" Hospitals — Technical and Scale Efficiency Results

Table 2 focuses upon "larger" hospitals and presents information on the geometric means, standard deviations and minimum values for the overall technical, pure technical and scale efficiency measures. This information is detailed for each of the four time periods under consideration. A number of points of note emerge from an examination of Table 2. In the first instance it is clear that a degree of consistency emerges across the respective time periods for each of the efficiency measures. For example, differences *vis à vis* the geometric mean measures, where they emerge, are of the order one per cent or less. This consistency is also reflected in the standard deviations and minimum values. Second, only in the case of the variable returns measure does a distinct trend emerge. For the period 1986-1989, relative to the constructed frontier, this average technical efficiency value for all "larger" hospitals in the sample was

Table 2: *Scale and Technical Efficiency Measures: "Larger" Hospitals*

Efficiency Measure	*Geometric Mean*	*Standard Deviation*	*Minimum*
	Technical efficiency (constant returns)		
1986-1989	0.938*	0.0846	0.7496
1987-1990	0.9328*	0.0897	0.7379
1988-1991	0.951*	0.0743	0.7558
1989-1992	0.9393*	0.0843	0.6799
	Technical efficiency (variable returns)		
1986-1989	0.969*	0.0614	0.8008
1987-1990	0.9718*	0.0603	0.7835
1988-1991	0.986	0.0367	0.8169
1989-1992	0.991	0.0517	0.7701
	Scale efficiency		
1986-1989	0.968*	0.0504	0.7822
1987-1990	0.9599*	0.0616	0.7694
1988-1991	0.965*	0.0594	0.7701
1989-1992	0.949*	0.0759	0.7113

*Statistically different from one at the 5 per cent significance level.

Table 3: *Scale and Technical Efficiency Measures: "Smaller" Hospitals*

Efficiency Measure	*Geometric Mean*	*Standard Deviation*	*Minimum*
	Technical efficiency (constant returns)		
1986-1989	0.9089*	0.1201	0.515
1987-1990	0.8793*	0.1303	0.463
1988-1991	0.8463*	0.1401	0.5438
1989-1992	0.8423*	0.1488	0.485
	Technical efficiency (variable returns)		
1986-1989	0.9443*	0.0893	0.6179
1987-1990	0.9343*	0.0967	0.6359
1988-1991	0.9292*	0.1052	0.6203
1989-1992	0.926*	0.1123	0.5416
	Scale efficiency		
1986-1989	0.9625*	0.0583	0.7031
1987-1990	0.9411*	0.0729	0.6465
1988-1991	0.9108*	0.0891	0.6739
1989-1992	0.9132*	0.0911	0.6732

*Statistically different from one at the 5 per cent significance level.

0.969. Average pure technical efficiency then steadily increases over the sample periods and achieves a value of 0.991 for the 1989-1992 time period. The implication of the results are that while over the period 1986-1989 "larger" hospitals, on average, could have produced the same level of outputs with approximately 3.1 per cent fewer resources than they actually employed, this potential efficiency gain had been reduced, by 1989-1992, to a mere 0.9 per cent. Consequently by 1989-1992 the scope for further efficiency gains by "larger" hospitals, at least from a pure technical efficiency perspective, had all but disappeared with actual inputs being almost equal to the minimum feasible inputs in producing observed outputs.

The third point of note to emerge from an examination of Table 2 relates to scale efficiency. Although the window analysis does not reveal the existence of a distinct trend there is, nevertheless, evidence of a fall in the scale efficiency of the "larger" hospitals. The average scale efficiency is 0.968 in 1986-1989 which indicates that a 3.2 per cent proportionate reduction in all inputs beyond what is achieved by eliminating pure technical inefficiency would be feasible if the input and output bundles were suitably altered. (Reduced or increased as needed.) For the period 1989-1992 the average scale measure had fallen to 0.949 which suggests that scale inefficiency had risen to 5.1 per cent (with this being offset by the above fall in pure technical inefficiency). At a later stage in this section we will investigate further the source of this increase in scale inefficiency.

"Smaller" Hospitals — Technical and Scale Efficiency Results

For the "smaller" hospitals subgroup, the efficiency measures show a much greater spread, as evidenced by the standard deviations and minimum values, but perhaps more importantly the window analysis reveals a trending deterioration for each of the measures.

Take overall technical efficiency as a case in point. In 1986-1989 average overall technical efficiency was 0.9089 but declined to 0.8423 for the time frame 1989-1992. This means that if overall technical inefficiency had been eliminated in the later period, an equiproportionate reduction in all inputs of 15.8 per cent, on average, could have been achieved compared to 9.1 per cent in the 1986-1989 period. These results indicate that the set of "smaller" hospitals not only commenced with a lower average level of overall technical efficiency in 1986-1989, but also experienced a greater decline in this average level over the time frame 1986-1989 to 1989-1992, relative to the set of "larger" hospitals.

Furthermore unlike the situation for "larger" hospitals, where (over time), scale efficiency and pure technical efficiency move in opposite directions, "smaller" hospitals have been subject to deterioration in both pure technical efficiency and scale efficiency. For the period 1986-1989 the mean level of pure technical efficiency under variable returns to scale was 0.9443 which implies that, on average, every input could be reduced by 5.6 per cent without reducing any output of a hospital. For the same period the sample average of scale efficiency was 0.9625 suggesting that a 3.75 per cent proportionate reduction in all inputs beyond what is achieved by eliminating pure technical inefficiency would be feasible if the input and output bundles were suitably altered. Comparable figures for 1989-1992 were 0.926 for pure technical efficiency (representing an equiproportionate input reduction of 7.4 per cent) and 0.9132 for scale efficiency (representing an 8.7 per cent reduction in all inputs).

The window analysis reveals "smaller" hospitals to have a lower initial (1986-1989) level of average overall technical efficiency relative to their "larger" counterparts and, over the sample period, to have experienced a pronounced decline in average overall technical efficiency in contrast to the marginal improvement achieved by "larger" hospitals. Given such a scenario the policy pursued by the health authorities in Northern Ireland, that of switching resources from "small" to "large" hospitals and either closing the "small" hospitals or turning them into specialist service providers appears justified.

In this respect, it is interesting to check whether specific hospitals, which have been subsequently closed did manifest low efficiency over the sample period. Examining the results for the (small) Moyle hospital, its overall level of technical efficiency in all of the time frames considered never rose above 0.689. From an efficiency perspective closure of this hospital can clearly be justified. In contrast to this case, however the small hospital at Banbridge was also subsequently closed even though the analysis revealed that this hospital was highly efficient over the sample period. Indeed Banbridge was located on the frontier on five separate occasions in the course of the window analysis. The average overall technical efficiency score for Banbridge, across the respective time frames, was 0.937 (with a minimum value of 0.705 and a maximum value of 1). This highlights that the closure of Banbridge was more to do with the overall rationalisation of the Northern Ireland hospital sector and the allied issue of patient accessibility than with efficiency. The Craigavon hospital, one of the six large hospitals targeted for additional resources, is within the catchment area for Banbridge and with the expansion of the Craigavon hospital, Banbridge became surplus to requirements.

Before leaving this discussion of hospital efficiency estimates, it is interesting to note how these results relate to the health service indicators used by hospital managers and policy makers. Since the latter use indicators such as the average length of inpatient stay, occupancy, and the medical and nursing resources per case, to assess hospital performance, we can check whether these indicators are correlated with the DEA estimates of overall efficiency. Given that the latter estimates are bounded, we undertook a Tobit analysis (which permitted both linear and non-linear relationships). The Tobit analysis confirmed that lower average stay, higher occupancy, and smaller utilisation of medical and nursing resources per case, were associated with greater efficiency. This finding not only demonstrates the plausibility of the overall technical efficiency results, but it also suggests that the DEA best-practice measures can provide a useful complement to health service indicators in assessing hospital performance.

Returns to Scale Results

From examination of the information detailed in Table 2 and Table 3, it is apparent that the major cause of overall technical inefficiency for "larger" hospitals is scale inefficiency, while for "smaller" hospitals both pure technical inefficiency and scale inefficiency are almost of equal blame. Therefore the next stage in the analysis is to explore the source of the scale inefficiency for each of the hospital subsets and for each of the data windows under consideration. This information is presented in Table 4 for "larger" hospitals and in Table 5 for the "smaller" hospitals.

Table 4: *Returns to Scale Characterisation: "Larger" Hospitals*

Time Period	*Constant Returns to Scale (No.)*	*Increasing Returns to Scale (No.)*	*Decreasing Returns to Scale (No.)*
1986-1989	25	9	11
1987-1990	21	13	11
1988-1991	24	13	8
1989-1992	24	15	6

Table 5: *Returns to Scale Characterisation: "Smaller" Hospitals*

Time Period	*Constant Returns to Scale (No.)*	*Increasing Returns to Scale (No.)*	*Decreasing Returns to Scale (No.)*
1986-1989	23	0	31
1987-1990	19	1	34
1988-1991	15	0	39
1989-1992	14	3	37

The window analysis detailed in Table 4 reveals that over the period approximately 50 per cent of "larger" hospitals exhibit constant returns to scale which of course implies that the other 50 per cent of the sample experience scale inefficiency. Of those classed as scale inefficient a marginally greater proportion of hospitals are operating with increasing returns to scale. When the results are scrutinised on an individual hospital basis the most notable findings centre on the hospitals targeted for significant expansion. Our sample set contains five of the six targeted hospitals (the other, the mid-Antrim hospital, did not become operational until after 1992). For two of the five hospitals — the Royal and the City — the window analysis reveals that throughout the period, almost without exception, these two hospitals were operating at close to the minimum efficient scale. In contrast, for the majority of the period, the other three hospitals have been subject to decreasing returns to scale. In the case of Altnagelvin decreasing returns to scale held in all instances while for the Ulster hospital and Craigavon hospital (although decreasing returns was the dominating result) there was some evidence, particularly for the analysis conducted for the 1989-1992 window, of both hospitals gravitating towards minimum efficient scale.

These findings provide only limited support for the current policy of concentration in health service provision. It is clear from the discussion that the dictates of efficiency in resource utilisation would militate against a further expansion of provision at, for example, Altnagelvin hospital, which is the major provider of hospital services in the north-west of Northern Ireland. More generally, none of the five identified hospitals are in a position to achieve efficiency gains through a further expansion in scale.

In sounding this note of caution about the expected gains in efficiency from increased concentration in hospital services in Northern Ireland, it is interesting to record that our empirical findings accord with other recent empirical studies elsewhere. For example, from a survey of empirical studies, Ferguson *et al.* (1997) noted that whereas the optimal size for an acute hospital lies in the 200-300 bed range, diseconomies of scale generally begin to appear in the 300-600 bed range. Likewise, recent evidence presented by Dranove (1998) suggests that 270 beds is an optimal size for an acute hospital. In the current study, the scale-efficient larger hospitals in the latest period were located in the 222-358 bed range.

It should, however be emphasised that post-1992 another major hospital became operational in mid-Antrim and there are plans put forward for a further new hospital in the north-west of Northern Ireland (this hospital will cover some of the catchment area currently provided for by Altnagelvin). The introduction of these new hospitals, which to some degree will ease pressure on existing providers, may go some way towards pushing the larger hospitals towards minimum efficient scale.

The above scale economies results indicate that policy makers need to pay attention to the possible negative effects of increasing concentration in hospital services in Northern Ireland. As well as noting the danger of scale diseconomies, it is essential that policy makers try to obtain better information on the quality of services by size of hospital, given their preference for larger size. This is particularly so given the contention, by Ferguson *et al.* (1997), that much of the research which attempts to investigate the relationship between patient volumes and outcomes in hospitals may overestimate the claimed positive impact of the volume of activity on the quality of care. Additional concerns for policy makers, with respect to more concentrated health services, relate to the questions of equity of access and the possibility of shifting costs to patients. For example, when these questions are directed to the provision of maternity services, there is some evidence that the quality of service and patient satisfaction is higher in smaller hospitals. Hence, although our scale efficiency findings provide certain support for the current policy of concentrating hospital services in Northern Ireland, policy makers need to be aware that concentration has the potential to generate costs as well as benefits.

In turning to the returns to scale results for the "smaller" hospitals, the window analysis, detailed in Table 5, reveals a steady decline in the number of "smaller" hospitals operating at the minimum efficient scale. In the period 1986-1989 approximately 40 per cent of "small" hospitals were operating at constant returns to scale, by 1989-1992 this percentage share had dropped to 25 per cent.

The scale inefficiency of the "small" hospital subset was almost exclusively due to decreasing returns to scale. In each time frame in excess of 55 per cent of small hospitals were classed as being subject to decreasing returns to scale. This suggests that, within this "smaller" hospital subset, economies of scale are fully exploited at relatively low levels of output. In contrast, in the "larger" hospital set, diseconomies of scale only begin to appear at much higher levels of outputs. This result provides further support for the current policy of concentrating acute hospital provision in large/medium-sized hospitals and only using very small hospitals for specialist provision (such as geriatric provision).

Size Efficiency Results

Further insight into the whole question of scale inefficiency can be obtained by measuring the size efficiency of Northern Ireland hospitals. As indicated in Section II size efficiency is concerned with measuring the potential input saving, additional to that achieved by eliminating pure technical efficiency, that could be achieved by optimally apportioning a hospital's output to a number of smaller hospitals (or hospital divisions). In doing this we are not implying that government policy is or should be directed towards breaking up certain hospitals into smaller units. Rather, we are attempting to gain further insight into potential efficiency gains in the context of hospitals producing under decreasing returns to scale. The size efficiency results for "larger" hospitals are presented in Table 6 with those for "smaller" hospitals detailed in Table 7.

Table 6 reveals that of those "larger" hospitals which are scale-inefficient due to decreasing returns to scale, a number in each of the time frames could

Table 6: *Size Efficiency: "Larger" Hospitals*[a]

Time Period	*Size Efficient (K=1)*	*Size Inefficient (K>1)*	*Maximum Value of K*
1986-1989	5	6	2
1987-1990	7	4	2
1988-1991	4	4	2
1989-1992	4	2	2

[a] The figures in column 2 and column 3 refer to those hospitals characterised as being subject to decreasing returns to scale.

potentially achieve input saving by a reorganisation into smaller units. For example, for the time frame 1986-1989, of the 11 hospitals classed as being subject to decreasing returns to scale, a total of six are size inefficient. For the remaining five hospitals, even though they are operating under decreasing returns to scale, it would be best to operate them in a purely technically efficient manner as single hospitals, if their output bundles were to be left unchanged. Comparable figures for the 1989-1992 time period indicate that six are subject to decreasing returns to scale with two of that number size inefficient. The final column of Table 6 identifies the highest value of K recorded in each of the time periods. For this subset of "larger" hospitals radical results were not apparent with the maximum optimal value of K being two. This suggests that in the worst case scenario it would be more efficient if the target output bundles were produced by two identical (in terms of their input-output bundles) smaller hospitals (or hospital divisions).

Size efficiency findings are reported in Table 7 for the sample of "smaller" hospitals which are scale inefficient due to decreasing returns to scale. From these results it is apparent that in each period a sizeable proportion of hospitals are classed as size inefficient. Furthermore, the final column of Table 7, which documents the highest value of K recorded, indicates that certain size inefficient hospitals would potentially benefit from a radical reorganisation. Table 8 for "larger" hospitals and Table 9 for "smaller" hospitals present details of the

Table 7: *Size Efficiency: "Smaller" Hospitals*[a]

Time Period	*Size Efficient (K=1)*	*Size Inefficient (K>1)*	*Maximum Value of K*
1986-1989	9	22	4
1987-1990	7	27	4
1988-1991	3	36	6
1989-1992	7	24	5

[a] The figures in column 2 and column 3 refer to those hospitals characterised as being subject to decreasing returns to scale.

potential reduction in all inputs that could be obtained if hospitals operated under pure technical efficiency. They also present details of the additional potential saving if hospitals operated under size efficiency. These tables draw together the information discussed earlier in this section. The stark conclusion from Table 8 is that the potential of "larger" hospitals for additional efficiency gain has been almost exhausted. Indeed, if it is accepted that Health Service policy in Northern Ireland is about channelling resources towards a smaller number of large hospitals, then size efficiency gains can be ruled out leaving only the potential for a meagre 0.09 per cent efficiency gain through the achievement of pure technical efficiency. Clearly, efficiency gains in large hospitals must come via scale efficiency as noted above. Table 9 highlights that "smaller" hospitals have substantial technical and size inefficiencies in addition to the substantial scale inefficiency noted earlier.

Table 8: *Potential Reduction in all Inputs via Technical and Size Efficiency: "Larger" Hospitals*

Time Period	*Pure Technical Efficiency (%)*	*Size Efficiency (%)*	*Total Potential Efficiency Gain (%)*
1986-1989	3.1	1.08	3.18
1987-1990	2.82	1.114	2.934
1988-1991	1.4	1.365	1.765
1989-1992	0.09	1.2	1.29

Table 9: *Potential Reduction in all Inputs Via Technical and Size Efficiency: "Smaller" Hospitals*

Time Period	*Pure Technical Efficiency (%)*	*Size Efficiency (%)*	*Total Potential Efficiency Gain (%)*
1986-1989	5.574	2.6	8.174
1987-1990	6.57	4.573	11.143
1988-1991	7.08	8.311	15.391
1989-1992	7.4	5.0	12.4

V CONCLUSION

The study examined the technical, scale and size efficiency of a sample of 23 Northern Ireland hospitals over the six-year period, 1986-1992. In computing these efficiency measures the study investigates whether the empirical results accord with the current rationalisation policy for hospital provision in Northern

Ireland. As such it investigates the efficiency of large hospitals, which are at present being expanded, and the efficiency of small hospitals which are currently being either downsized or closed.

In many ways, the results of the study cautiously accord with the rationalisation policy now being pursued. For example, our results highlight that the set of "smaller" hospital relative to "larger" hospitals not only commenced with a lower average level of overall technical efficiency but also experienced a much greater level of decline over the time frame under scrutiny. Furthermore, unlike the situation for "larger" hospitals where, over time, scale efficiency and pure technical efficiency move in opposite directions with the aggregate effect that of little change, "smaller" hospitals have been subject to a marked deterioration in both pure technical efficiency and scale efficiency. The discussion of size efficiency also yielded a stark contrast between large and small hospitals with a much greater proportion of the smaller hospitals classed as size inefficient.

The study has focused on efficiency in the provision of hospital services in Northern Ireland. Clearly larger hospitals have a comparative advantage over their smaller counterparts in this area. Efficiency in service provision does not, however, necessarily equate with quality of care. Although the analysis has sought to ensure a similar quality of care in relative efficiency measurement by focusing on intra-group movements that are based on hospitals of comparable size, we must stress that this study makes no statement as to where the best quality of care is to be obtained. As an ongoing research programme a reworking of the analysis based on a quality of care adjustment of the outputs would be the natural next step. Unfortunately, at the present time, data is not available to permit such a follow up study. Indeed, given the pace of the rationalisation programme such a study, even in the near future, may only have historical relevance for "smaller" hospitals.

REFERENCES

ANDERSEN, P., and N.C. PETERSEN, 1993. "A Procedure for Ranking Efficient Units in Data Envelopment Analysis", *Management Science*, Vol. 39, No. 10, pp. 1,261-1,264.

DEPARTMENT OF HEALTH, 1989. *Working for Patients: The Health Service — Caring for the 1990s*, London: HMSO, p. 24.

DEPARTMENT OF HEALTH AND SOCIAL SERVICES, 1987. *A Regional Strategy for Health and Social Services 1987-92*, Belfast: Department of Health and Social Services.

DRANOVE, D., 1998. "Economies of Scale in Non-revenue Producing Cost Centres: Implications for Hospital Mergers", *Journal of Health Economics*, Vol. 17, pp. 69-83.

FÄRE, R., S. GROSSKOPF, and C.A.K. LOVELL, 1994. *Production Frontiers*, Cambridge: Cambridge University Press.

FERGUSON, B., J. POSNETT, AND T. SHELDON, 1997. *Concentration and Choice in Healthcare,* London: The Royal Society of Medicine Press.

FERRIER, G.D., and V. VALDMANIS, 1996. "Rural Hospital Performance and its Correlates", *Journal of Productivity Analysis*, Vol. 7, No. 1, pp. 63-80.

GROSSKOPF, S., and V. VALDMANIS, 1987. "Measuring Hospital Performance", *Journal of Health Economics*, Vol. 6, pp. 89-107.

HOLLINGSWORTH, B., and D. PARKIN, 1995. "The Efficiency of Scottish Acute Hospitals: An Application of Data Envelopment Analysis", *Institute of Mathematics and its Applications: Journal of Mathematics Applied to Medicine and Biology,* Vol. 12, pp. 161-173.

JONES, C.S., and I.P. DEWING, 1997. "The Attitudes of NHS Clinicians and Medical Managers towards Changes in Accounting Controls", *Financial Accountability and Management*, Vol. 13, No. 3, pp. 261-280.

LOVELL, C.A.K., 1993. "Production Frontiers and Productive Efficiency", in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency,* New York, Oxford: Oxford University Press, pp. 3-67.

MAINDIRATTA, A., 1990. "Largest Size-efficient Scale and Size Efficiencies of Decision-making Units in Data Envelopment Analysis", *Journal of Econometrics*, Vol. 46, pp. 57-72.

NEWHOUSE, J.P., 1994. "Frontier Estimation: How Useful a Tool for Health Economics?" *Journal of Health Economics,* Vol. 13, pp. 317-322.

NORTHERN IRELAND AUDIT OFFICE, 1993. *Department of Health and Social Services: The Provision of Acute Hospital Services in Northern Ireland,* London: HMSO.

NORTHERN IRELAND ECONOMIC COUNCIL, 1994. *The Reform of Health and Social Care in Northern Ireland: An Introduction to the Economic Issues,* Report 110.

NORTHERN IRELAND ECONOMIC COUNCIL, 1995. *Health and Personal Social Services to the Millenium,* Occasional Paper 5.

NUNAMAKER, T.R., 1985. "Using Data Envelopment Analysis to Measure the Efficiency of Non-profit Organisations: A Critical Evaluation", *Managerial and Decision Economics*, Vol. 6, No. 1, pp. 50-58.

PARKIN, D. and B. HOLLINGSWORTH, 1997. "Measuring Production Efficiency of Acute Hospitals in Scotland, 1991-94: Validity Issues in Data Envelopment Analysis", *Applied Economics,* Vol. 29, pp. 1,425-1,433.

ROSENAMN, R., K. SIDDHARTHAN, and M. AHERN, 1997. "Output Efficiency of Health Maintenance Organisations in Florida", *Health Economics*, Vol. 6, pp. 295-302.

SÖDERLUND, N., I. CSABA, A. GRAY, R. MILNE, and J. RAFTERY, 1997. "Impact of the NHS Reforms on English Hospital Productivity: An Analysis of the First Three Years", *British Medical Journal*, Vol. 115, pp. 1,126-1,129.

THANASSOULIS, E., A. BOUSSOFIANE, and R.G. DYSON, 1998. *DEA Models for the Assessment of the Provision of Perinatal Care in England on the Basis of Quantity and Quality of Outputs*, University of Warwick, Coventry: Warwick Business School Research Paper No. 26.

WILSON, P.W., 1995. "Detecting Influential Observations in Data Envelopment Analysis", *Journal of Productivity Analysis,* Vol. 6, pp. 27-45.