The United Nations University

WIDER

World Institute for Development Economics Research

Working Papers No. 187

June 2000

Hospital Efficiency in Sub-Saharan Africa

Evidence from South Africa

Eyob Zere

UNU World Institute for
Development Economics Research
(UNU/WIDER)

Working Papers No. 187

Hospital Efficiency in Sub-Saharan Africa

Evidence from South Africa

Eyob Zere

June 2000

This study has been prepared within the UNU/WIDER project on Underdevelopment, Transition and Reconstruction in Sub-Saharan Africa, which is directed by Professor Tony Addison, UNU/WIDER.

UNU/WIDER gratefully acknowledges the financial contributions to the project by the Government of Italy (Directorate General for Development Cooperation), the Government of Sweden (Swedish International Development Cooperation Agency-Sida), and the Government of the United Kingdom (Department for International Development). These agencies can accept no responsibility for any information provided or views expressed.

Eyob Zere is affiliated with the Health Economics Unit, University of Cape Town.

UNU World Institute for Development Economics Research (UNU/WIDER)
A research and training centre of the United Nations University

The Board of UNU/WIDER

Nora Lustig
Harris Mutio Mule
Jukka Pekkarinen, Vice Chairperson
Ruben Yevstigneyev
Masaru Yoshitomi

Ex Officio

Hans J. A. van Ginkel, Rector of UNU
Matti Pohjola, Officer-in-Charge, a.i., UNU/WIDER

UNU World Institute for Development Economics Research (UNU/WIDER) was established by the United Nations University as its first research and training centre and started work in Helsinki, Finland in 1985. The purpose of the Institute is to undertake applied research and policy analysis on structural changes affecting the developing and transitional economies, to provide a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and to promote capacity strengthening and training in the field of economic and social policy making. Its work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

UNU World Institute for Development Economics Research (UNU/WIDER)
Katajanokanlaituri 6 B
00160 Helsinki, Finland

Copyright © UNU/WIDER 2000

Camera-ready typescript prepared by Janis Vehmaan-Kreula at UNU/WIDER
Printed at Pikapaino Paatelainen Oy, Helsinki

The views expressed in this publication are those of the author(s). Publication does not imply endorsement by the Institute or the United Nations University, nor by the programme/project sponsors, of any of the views expressed.

ISSN 0782-8233
ISBN 952-455-064-4 (printed publication)
ISBN 952-455-065-2 (internet publication)

CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

I am profoundly indebted to Professors Tony Addison and Matti Pohjola of UNU/WIDER and Professor Diane McIntyre of the Health Economics Unit, University of Cape Town, for their helpful comments. I gratefully acknowledge UNU/WIDER for its financial assistance and hospitality during my stay in Helsinki and thank its director, Professor Giovanni Andrea Cornia. Finally, I thank the University of Cape Town for its financial assistance. The usual disclaimer applies.

ABSTRACT

This study evaluates the technical efficiency and productivity of a sample of public sector hospitals in three provinces of South Africa using the non-parametric techniques of *data envelopment analysis* (DEA) and DEA-based *Malmquist productivity index* (MPI). A tobit regression is also estimated to identify some factors that may be associated with (in)efficiency. The sample consists of 86 hospitals classified into three levels: community hospitals with emergency services only (Level I), community hospitals with outpatient services (Level II) and non-academic secondary and tertiary hospitals (Level III). Recurrent expenditure and bed-size are used as inputs (but only the first one in Level III hospitals because of their small number). Outputs include inpatient days and outpatient visits – these are expected to capture the bulk of the activity of non-academic hospitals.

The findings indicate that there is a marked variation of performance among hospitals within each group. An average overall technical efficiency of 0.74, 0.68 and 0.70 is computed for Level I, II and III hospitals respectively. This implies that when compared to their peers on the frontier, the inefficient hospitals on average consume 35-47 per cent more resources. The number of hospitals on their respective group's frontier is 6 in Level I (n=55), 3 in Level II (n=19) and 2 in Level III (n=12). Most hospitals operate at a non-optimal scale, with decreasing returns to scale dominating in Level II and III hospitals.

If the inefficient hospitals were to operate as efficiently as their peers on the frontier, efficiency gains in terms of reduction in recurrent expenditure would amount to about R 279 million (about US$ 47 million) – an amount which can cover the costs of constructing a sizeable number of clinics or upgrading service quality where necessary. This would offset the need to raise user charges, and would potentially be more equitable. Furthermore, the number of hospital beds could also be reduced significantly.

The occupancy rate affects the technical efficiency positively in all three Levels of hospitals. The average length of stay, however, seems to have an adverse effect only in Level III hospitals. The tobit estimates for Level I hospitals indicate that the number of outpatient visits as a proportion of inpatient days affects technical efficiency positively. This might be an indication of the presence of economies scope between the two outputs.

The MPI for a sample of Western Cape hospitals exhibits a decrease in total factor productivity of about 12.5 per cent for the period 1992/93-

1997/98. This is a result of both a decline in efficiency and a technical regress.

These results indicate the potential to improve access and/or quality of care without injecting additional resources into the health sector. This is important given the financial constraints on social sector investment in South Africa. It is also concluded that inefficiency levels of such a magnitude are likely to undermine the government's initiatives to redress inequity in a sustainable manner. Finally, it is desirable to replicate this study on a large scale, covering all types of hospitals (public and private) so as to assess the gravity of the problem and its causes, and thereby maximize possible efficiency savings.

LIST OF ACRONYMS

AIDS	Acquired Immuno-deficiency Syndrome
ALS	Average Length of Stay
ANC	African National Congress
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
DRG	Diagnosis-Related Groups
DRS	Decreasing Returns to Scale
FTE	Full-time Equivalent
HIV	Human Immuno-deficiency Virus
IRS	Increasing returns to Scale
LP	Linear Programming
MPI	Malmquist Productivity Index
PHC	Primary Health Care
RTS	Returns to Scale
SA	South Africa
SSA	Sub-Saharan Africa
TE	Technical Efficiency
TFP	Total Factor Productivity
VRS	Variable Returns to Scale

PROJECT SUMMARY

Underdevelopment, Transition and Reconstruction in Sub-Saharan Africa

This UNU/WIDER project focuses on Angola, Eritrea, Ethiopia, Guinea-Bissau, Mozambique, and Somalia. These countries share a common history; state socialism was the ideology of their early development strategies and liberation movements, and economic failure together with the politics of the Cold War led to intense, and often recurring, conflict. Conflict erupted again in Angola, Eritrea, Ethiopia, and Guinea-Bissau during 1998 and 1999 while Somalia remains highly unstable.

Communities need help to reconstruct, private sectors must revitalize themselves, and state capacities must be built. This is an exceptionally demanding set of tasks given the scarcity of financial resources and skills. Moreover, economic reform is on the agenda for all of these countries, but progress is at best hesitant and, in many cases, stalled. The relationship between reconstruction and reform is also an uneasy one; the two agendas should be mutually supporting but this is often far from the case and in consequence opportunities to accelerate growth and poverty reduction are missed. These failures reinforce the already high vulnerability of these countries to conflict.

Further details of the project and its research outputs can be found on the UNU/WIDER website (www.wider.unu.edu) and/or by contacting the project director, Professor Tony Addison (email: addison@wider.unu.edu).

1. INTRODUCTION

The health care systems of Sub-Saharan African countries (SSA) including South Africa (SA) increasingly face critical resource constraints in their efforts to extend health services of acceptable quality to the vast majority of people. This severe shortage of health care resources is accounted for by a host of factors including, poor macroeconomic performance, cutbacks in public spending, rapid population growth, the AIDS epidemic and an upsurge in diseases such as malaria.

The hospital sector is a large consumer of scarce health care resources. Although the actual percentage varies from country to country, hospitals in developing countries consume an average of 50-80 per cent of the public sector health resources (Barnum and Kutzin 1993). In 1992/93, hospitals in SA consumed about 89 per cent of the total public sector expenditure on health (McIntyre *et al.* 1995a, Castro-Leal *et al.* 1999). The efficiency of hospitals merits close attention and scrutiny because of their enormous consumption of resources.

It is common knowledge that the health care system (especially hospitals) in developing countries is inefficient. The World Bank's policy study on *Financing Health Services in Developing Countries* (Akin *et al.* 1987) indicates that one of the major problems of African health care systems is the inefficiency of government health programmes, the others being problems of allocation and inequity.

Even though efficiency is accorded a central place in the health policies of most countries, in practice much remains to be done. The dearth of literature on hospital efficiency studies in SSA may, perhaps, indicate that in practice not much attention is given to efficiency by health care administrators. Much of the attention of policymakers, donors and health systems researchers seems to be focused on health sector reforms, prominent of which is the mobilization of additional resources for health care through user fees and other modalities of financing.

In the presence of inefficiency, costs of service delivery are inflated. This undermines the cost-recovery ratio and any other stated benefits of cost-sharing schemes. Furthermore, given the economic realities of SSA countries, the task of redressing inequalities in access to health care cannot be achieved without a concomitant improvement in efficiency. Inefficiency is more likely to breed further inequity.

South Africa is a country still grappling with the legacy of the apartheid system. There is a glaring gap in health indicators and access to health care between the most and the least privileged population groups. Given the macroeconomic and socio-demographic realities of the country, there is an urgent need to assess the efficiency and productivity of hospitals, and their determinants.

This study seeks to examine the technical efficiency and productivity of a sample of hospitals in South Africa. The findings will help deepen the understanding of the magnitude of inefficiency and its causes in SSA. Its specific objectives are to:

- evaluate the technical and scale efficiency of non-academic acute care hospitals in the Eastern, Northern and Western Cape provinces of South Africa

- identify some of the factors that are likely to influence the (in)efficiency of hospitals

- assess changes in the productivity of acute care hospitals in the Western Cape province

To date, there are very few published studies of hospital efficiency in SSA that have made use of the frontier approach to efficiency estimation. Of those few studies, none have used the method of *data envelopment analysis* (DEA). This study will therefore, demonstrate the advantages of this method in informing policy, planning and management.

The paper is organized as follows: section two presents background information on South African health sector; section three discusses the analytical and conceptual framework and introduces the method of data envelopment analysis; section four explains the data and methods; section five presents the empirical results; section six discusses the results; and section seven concludes by drawing out some policy implications.

2. SOUTH AFRICA'S HEALTH SECTOR

2.1 Overview of the economic and socio-demographic profile

South Africa, a country of 38 million people, has recently emerged from decades of apartheid rule. The gap in health status between the priviledged minority and the underpriviledged majority is wide, and the subject of much discussion. With the population growing at 2.5 per cent (World Bank 1999), there is a growing demand for health care. To improve the health status of the mass of the population is an urgent priority for the government, but one that is difficult to meet given the scarcity of resources. The economy is growing slowly and thus the growth of public revenues with which to finance health care investment is very slow. Therefore the Government faces hard choices in allocating its limited resources.

Table 1 presents the basic health and development indicators of SA juxtapositioned with those of SSA and other upper middle income countries with which SA's economic performance is comparable.

Prior to 1994, SA had a recially structured political and administrative system. The system favoured the '*whites*' who accounted for about 13.2 per cent of the total population. The disadvantaged groups included, '*Africans*' that constituted about 75.6 per cent, '*coloureds*' (8.6 per cent) and '*Asians*' (2.6 per cent) (McIntyre *et al*. 1995a).

Income distribution is highly unequal. In the period 1986-95 the share of the richest quintile was about 63.3 per cent of the country's income, whereas that of the poorest quintile was only 3.3 per cent (World Bank 1998b). In 1993, about 23.7 per cent of the population lived below US$ 1 a day and about 50.2 per cent below US$ 2 a day (World Bank 1998a)[1]. The Gini index[2] for South Africa (58.4) is one of the highest in the world (*ibid*). Its Human Development Index (HDI)[3] for 1999 is 0.695 (classified as medium HDI) and has an HDI rank of 101 out of a total of 174 countries (UNDP 1999). The gap in living standards and human development indicators between the various racial groups is so large that SA harbours characteristics of both the developed and developing world.

---

1 Population below US$ 1 a day and US$ 2 a day are the percentages of the population living on less than US$ 1 a day and US$ 2 a day at 1985 international prices, adjusted for purchasing power parity.

2 The Gini index measures the extent to which the distribution of income among individuals or households deviates from a perfectly equal distribution. A Gini index of zero measures perfect equality while an index of 100 implies perfect inequality.

3 A composite index calculated from three indicators: life expectancy at birth, educational attainment and real GDP per capita.

TABLE 1
HEALTH AND DEVELOPMENT INDICATORS: SA, SSA AND OTHER UPPER MIDDLE INCOME COUNTRIES[4]

Indicator	SA	Other SSA	Other upper midddle income countries
GNP per capita, 1996 (US$)	3,520	490	4,600
Average annual growth of GNP per capita, 1995-96	1.0	1.9	5.0
Life expectancy at birth, 1996	65	52	70
Population with access to sanitation facilities, 1995 (%)	46	37	64
Population with access to safe water, 1995 (%)	70	45	76
Adult illiteracy rate[5], 1995 (M/F)	18/18	34/53	13/16
Infant mortality rate per 1000 live births, 1996	49	91	30
Under-five mortality rate per 1000 (1996)	66	147	36

Source: World Bank (1998a).

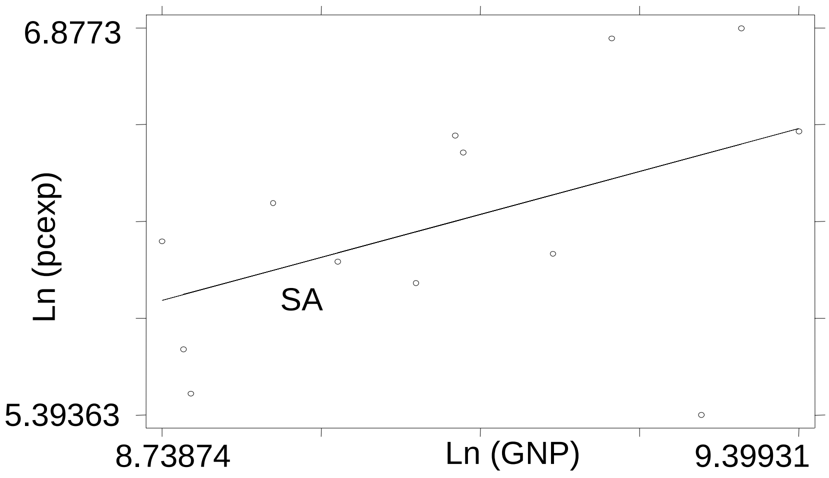
2.2 The health sector

Health care expenditure for the period 1990-95 amounted to about 7.9 per cent of GDP (3.6 per cent public and 4.3 per cent private expenditure). This translates into a health expenditure per capita of US$ 257. The corresponding average for SSA was 2.9 per cent with a per capita expenditure of US$ 55 (World Bank 1998a). A comparison of SA's per capita expenditure on health with those of other middle income countries shows that South Africa's per capita expenditure is a little short of the average for the upper middle income countries (Figure 1).

4 Upper middle income countries defined as those between incomes US$ 3,116 to US$ 9,635.

5 Age 15 years and above.

FIGURE 1
PER CAPITA GNP VERSUS PER CAPITA EXPENDITURE ON HEALTH, SELECTED UPPER MIDDLE INCOME COUNTRIES



Source of data: World Bank (1998a, 1999).

SA's health indicators are worse than those of countries that spend similar amounts on health care (McIntyre *et al.* 1998). This, among other things, is accounted for by the skewed distribution of health care resources between the public and private sectors, and between different geographical areas and between levels of care (*ibid*). The spatial distribution of health care resources is presented in Table 2.

South Africa's epidemiological profile follows the social stratification of its population. There are different causes of morbidity and mortality between the most and the least advantaged groups. Infectious and parasitic diseases – the curse of the poor – account for about 14 per cent of deaths amongst Africans, but only 2 per cent amongst whites. On the other hand, about 40 per cent of deaths amongst whites are attributable to cardio-vascular diseases, the classic disease of wealth. The corresponding figure for Africans is 12 per cent (ANC 1994).

TABLE 2
HEALTH FACILITIES AND PERSONNEL IN SA PROVINCES, 1992/93

Province	Hospital beds per 1,000 population	Doctors per 100,000 population	Nurses per 100,000 population	Pharmacists per 100,000 population
Eastern Cape	3.5	30.7	321.3	20.1
Eastern Transval	2.1	28.3	265.8	23.1
Gauteng	6.0	127.4	618.4	109.8
Kwazulu-Natal	3.8	53.5	431.9	28.7
Northern Cape	4.0	37.6	432.3	28.5
Northern Transvaal	2.5	15.5	293.2	7.8
North-West	3.3	22.7	273.5	22.8
Orange Free State	4.1	46.5	382.3	38.8
Western Cape	5.4	143.8	686.3	79.8
Total	4.0	60.2	421.5	42.6

Source: McIntyre *et al.* (1995a).

Tuberculosis is one of the country's major health problems. It consumes R[6]200 million in resources every year (Van Rensburg *et al.* 1992). The HIV/AIDS epidemic is also assuming alarming dimensions and threatens to reverse whatever gains the current Government may have scored in improving the quality of life of the people. South Africa has the fastest-growing AIDS epidemic in the world (UNDP 1999).

The Government's health policy emphasizes the principles of efficiency and equity as its prime objectives. Given the gross unfairness of the apartheid-era health policy, the Govenment intends to totally transform the system to ensure access for all South Africans. The Government's policy is based on a primary health care strategy that emphasizes equitable socio-economic development (ANC 1994).

---

6 R (Rand) is South African currency; approximately R 6 = US$ 1.

3. ANALYTICAL AND CONCEPTUAL FRAMEWORK

3.1 Efficiency: concept and measurement

The measurement of efficiency in the health care sector is complicated by the nature of the production process. Measurement of the ideal output – improved health status – is difficult, both conceptually and empirically (Grosskopf and Valdmanis 1987). Complications arise from the fact that health status is a function of many variables, many of which are exogeneous to the health sector – for example household income, education, and intra-household decisions.

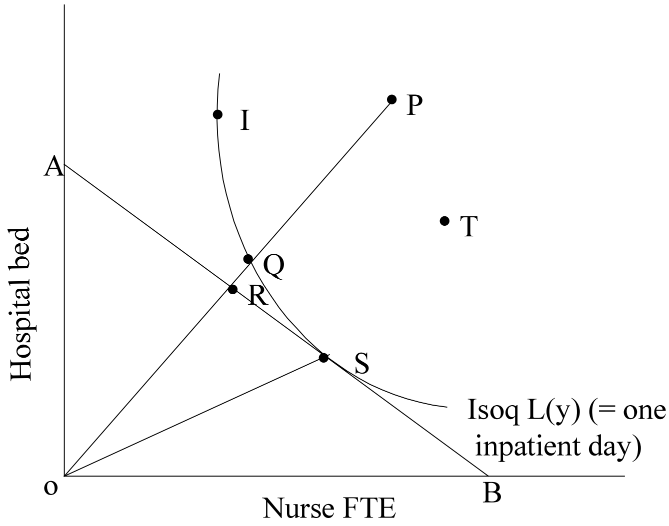
Farrell (1957), drawing upon the work of Debreu (1951) and Koopmans (1951), introduced a measure of productive efficiency that avoids the problems associated with traditonal average productivity measures (ratios). He refuted the idea of an absolute measure of efficiency and proposed that efficiency be measured relative to a best-performance frontier determined by a representative peer group. Furthermore, he provided the definitions and computational framework for technical and allocative (in)efficiency.

In the Farrell framework, a firm's efficiency is measured relative to the efficiency of all other firms in the industry, subject to the restriction that all firms are on or below the frontier. A firm is regarded as technically efficient if it is operating on the best-practice production frontier in the industry. The basic ideas underlying Farrell's concepts of technical and allocative efficiencies (under the assumption of constant returns to scale) are illustrated in Figure 2.

In the illustration below, a hospital produces its output (one inpatient day) using a combination of two inputs (nurse FTE and hospital bed). A technically efficient hospital is one that is *'located on an isoquant'*, that is on the frontier. Thus, hospitals operating at points *I*, *Q* and *S* are technically efficient. Hospitals operating at points *P* and *T* are technically inefficient. For the hospital operating at point *P*, the measure of technical efficiency (*TE*) is given as:

$$TE_p = \frac{OQ}{OP} \qquad (1)$$

FIGURE 2
TECHNICAL AND ALLOCATIVE EFFICIENCIES



This denotes the ratio of the minimal input required to the actual input use, given the input mix used by *P*. The ratio *QP/OP* represents the percentage by which all inputs could be reduced without a reduction in output. If the hospital producing at point 'P' is to be efficient, it has to relocate itself at point *Q*. Technical efficiency takes values between zero and one ($0 \leq TE_i \leq 1$). Technically inefficient production units have a TE_i value less than one, while the efficient ones have a TE_i value of 1.

Given input prices, the *isocost line* AB represents the minimum cost of producing one unit of output. Allocative efficiency demands that production take place at the point where the *isoquant* is *tangential* to the *isocost line*. Given this definition, hospitals producing at points *I* and *Q*, which are regarded as technically efficient, are *allocatively inefficient*. Only the hospital operating at point *S* is both *technically and allocatively efficient*. The allocative efficiency for the hospital at point *P* is given as:

$$AE_p = \frac{OR}{OQ} \qquad (2)$$

The ratio *RQ/OQ* represents the percentage reduction in production costs that would occur if production were to occur in the allocatively efficient point *S*.

Farrell proposed that *overall efficiency*[7] *(EE)* be measured as:

$$EE_p = \frac{OR}{OP} \tag{3}$$

The *overall (economic) efficiency (EE)* has the advantage that it easily decomposes into technical and allocative efficiencies:

$$\frac{OR}{OP} = \frac{OQ}{OP} \times \frac{OR}{OQ} \tag{4}$$

That is, $EE = TE \times AE$

The above measures represent *input-oriented*, *radial* measures of efficiency. They are input-oriented, as their focus is on the measurement of variations in input use between different hospitals for a standardized output. The measures are radial as they are taken along a ray from the origin in the input-output space. This implies that the current input-output mix determines the firm's technology and any possible increase in efficiency will be achieved if inputs are reduced proportionally, with output proportions held constant. The radial nature of the efficiency measures allows comparison of hospitals with similar input-output mixes. Furthermore, each input and output can be measured in its natural physical unit without having to resort to a weighting system, to express the different units in a common denominator such as price (Valdmanis 1992).

Output–oriented measures can also be illustrated on the input-output space by taking an example of a production process involving one input and two outputs. The two measures of efficiency are equivalent under the assumption of constant returns to scale (CRS).

As seen in the foregoing discussion, empirical estimates of efficiency measures involve two steps: (i) estimation of the frontier and (ii) calculation of the individual hospital deviations from the frontier.

---

7 Economic efficiency is the term that has replaced it in the recent literature (Coelli *et al.* 1998).

Currently, there are two approaches used in estimating frontiers (Seiford and Thrall 1990, Coelli 1998). These are the *parametric* approach, which employs *econometric* methods, and the *non-parametric* approach, which involves *linear programming* techniques. The non-parametric method known as *data envelopment analysis* is used in this study. An exposition of this technique follows.

3.2 Data envelopment analysis (DEA)

Building on Farrell's seminal work, Charness *et al.* (1978) proposed the non-parametric technique of *DEA* for measuring the relative efficiencies of decision making units (DMUs)[8] such as schools, post offices and hospitals. *DEA* uses linear programming methods to establish the frontier from sample data. The efficiency of a DMU is then measured relative to the efficiency of all others in the group, subject to the restriction that all DMUs lie on or below the frontier (Bjurek *et al.* 1990, Seiford and Thrall 1990, Coelli *et al.* 1998). This is performed by solving a series of LP problems.

DEA is the preferred method of efficiency analysis in the non-profit sector where (Coelli *et al.* 1998):

(i) random noise is less of a problem;

(ii) multiple-output production is relevant;

(iii) price data is difficult to find; and

(iv) setting behavioural assumptions such as profit (cost) maximization (minimization) is difficult.

There are two major drawbacks to this method (Lovell 1993, Coelli *et al.* 1998). First, DEA is *non-stochastic*. It does not capture random noise (e.g. epidemics, weather, strike). Any deviation from the estimated frontier is interpreted as being due to inefficiency. Second, it is *non-statistical*, in the sense that it is not possible to conduct statistical tests of hypothesis regarding the inefficiency and the structure of the production technology.

Ferrier and Valdmanis (1996), however, argue that, these drawbacks may not be as serious as they initially seem. First, as there is no *a priori*

---

8 Intended to emphasize an orientation toward managed entities in the public and /or not-for-profit sectors.

specification of the functional form of the technology, specification error that might show up as a noise is ruled out. Secondly, as inputs and outputs are measured in their natural physical units, a measurement error is most unlikely.

Under the assumption of constant returns to scale (CRS), the efficiency of hospital j_0 can be obtained by solving the following model (Charnes *et al.*, 1978):

$$Max\ h_0 = \sum_{r=1}^{s} u_r y_{rj_0} \qquad (5)$$

Subject to

$$\sum_{i=1}^{m} v_i x_{ij_0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \qquad j = 1,.., N$$

$$u_r v_i \geq 0$$

Where:

y_{rj} ($r = 1,\ldots, s$) = observed amount of output r from hospital j

x_{ij} ($i = 1,\ldots, m$) = observed level of input i used by hospital j

u_r = weight given to output r

v_i = weight given to input i

The first constraint indicates that the weighted sum of inputs for the particular hospital equals one. The second implies that all hospitals are on or below the frontier, that is, the efficiencies of all hospitals have an upper bound of one. The weights u_r and v_i are treated as unknowns, and their weights are obtained in the linear programming solution.

The CRS assumption is only appropriate if all hospitals are operating at an optimal scale. When hospitals are not operating at an optimal scale, the *TE* can be decomposed into *pure technical efficiency* and *scale efficiency*. Thus in a situation where the CRS assumption does not hold, the *TE* measure is mixed with scale efficiency. To disentangle the effects of scale efficiency it is necessary to use a DEA model with a *variable returns to scale* (VRS) assumption. To this end Banker *et al.* (1984), developed an extension of the original CRS model. The LP problem to be solved is:

$$Max\ h_0 = \sum_{r=1}^{s} u_r y_{rj_0} + u_0 \qquad (6)$$

Subject to

$$\sum v_i x_{ij_0} = 1$$

$$\sum u_r y_{rj} - \sum v_i x_{ij} + u_0 \leq 0, \qquad j = 1,\ldots,n$$

$$u_r, v_i \geq 0$$

$$u_0 \gtreqless 0$$

Where the notations are as given in (5). The additional term corresponds to an intercept (Bjurek *et al.* 1990) and is unconstrained in sign. The sign of u_0 determines the returns to scale, where $u_0 < 0$ indicates increasing returns to scale, $u_0 = 0$ is for constant returns, and $u_0 > 0$ is for decreasing returns to scale.

3.3 The Malmquist productivity index

The Malmquist productivity index that was proposed by Caves *et al.* (1982), measures total factor productivity (TFP)[9] change between two data points in terms of ratios of distance functions. A Malmquist index greater than one indicates growth in productivity, while a value of less than one indicates a decline. The Malmquist index approach requires neither *a priori* behavioural assumption about the production technology nor input and output price data. These characteristics make it more appealing for measuring productivity in the public sector.

9 It is the average product of all inputs (in contrast, partial factor productivity is the average product of a single input; e.g. child deliveries per midwife).

Following Färe *et al*. (1994) output-oriented Malmquist total factor productivity change between periods t and t+1 is defined as

$$M_o^{t,t+1}\left(y^t, x^t, y^{t+1}, x^{t+1}\right) = \left[\frac{D_o^t\left(y^{t+1}, x^{t+1}\right)}{D_o^t\left(y^t, x^t\right)} \times \frac{D_o^t\left(y^{t+1}, x^{t+1}\right)}{D_o^{t+1}\left(y^t, x^t\right)}\right]^{1/2} \quad (7)$$

Färe *et al*. (1994) further decomposed the MPI into two parts: one measuring efficiency change and another measuring technologic change as follows:

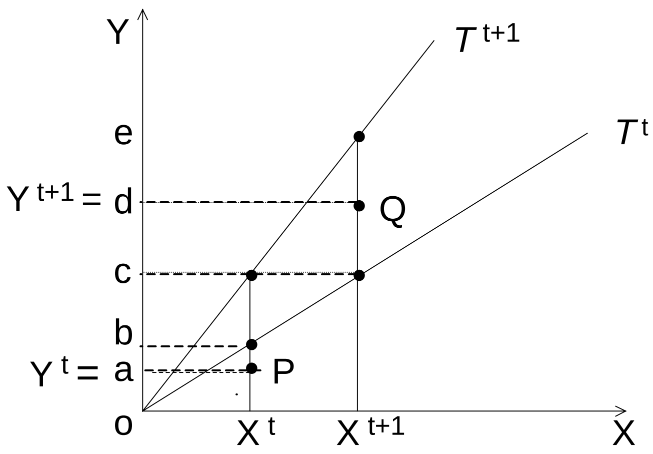
$$M_0^t\left(y^t, x^t, y^{t+1}, x^{t+1}\right) = \left[\frac{D_o^{t+1}\left(y^{t+1}, x^{t+1}\right)}{D_o^t\left(y^t, x^t\right)}\right] \times \left[\frac{D_o^t\left(y^{t+1}, x^{t+1}\right)}{D_o^{t+1}\left(y^{t+1}, x^{t+1}\right)} \times \frac{D_o^t\left(y^t, x^t\right)}{D_o^{t+1}\left(y^t, x^t\right)}\right]^{1/2}$$

(8)

The first term on the right hand side measures efficiency change and the second measures technical change. Färe *et al*. (1994) listed a number of different methods to calculate the Malmquist productivity index. However, the most preferred of these methods is the one that uses DEA-like linear programming techniques. Four linear programming problems are solved for each hospital to compute four distance functions to measure the total factor productivity change between two periods under a constant returns to scale technology (*ibid*). The technical efficiency change can further be decomposed into pure efficiency-change component and scale-change component by solving two additional linear programming problems under variable returns to scale technology (Coelli *et al*. 1998).

The definition and measurement of *MPI* is illustrated in Figure 3.

FIGURE 3
OUTPUT-BASED MALMQUIST PRODUCTIVITY INDEX



T^{t} and T^{t+1} represent the production technology in two periods, t and $t+1$. The hospital produces at point P in period t and at point Q in period $t+1$. Using formula (8), the decomposition of the MPI from the above figure is given as:

$$Efficiency\ \ change\ =\ \frac{0d/0e}{0a/0b} \qquad (9)$$

That is, the efficiency change is the ratio of the Farrell technical efficiency in period $t+1$ to that in period t.

The technical change is the geometric mean of the shift in technology evaluated at x^{t+1} and the shift in technology evaluated at x^{t}.

$$Technical\ change = \left[\frac{0d/0c}{0d/0e} \times \frac{0a/0b}{0a/0c}\right]^{1/2} \qquad (10)$$

4. DATA AND METHODS

4.1 Source of data

Data are mainly obtained from the annual statistical publications of the department of health, provincial administration of the Western Cape, which includes the Eastern, Northern and Western Cape Provinces. The data covers the years 1992/93-1997/98. The selection of the time period covered was dictated by the availability and completeness of the data. Thus, while efficiency scores will be computed on cross-sectional data for hospitals of the three provinces for the year 1992/93, Malmquist productivity indices will be calculated only for the Western Cape province hospitals.

4.2 Data description

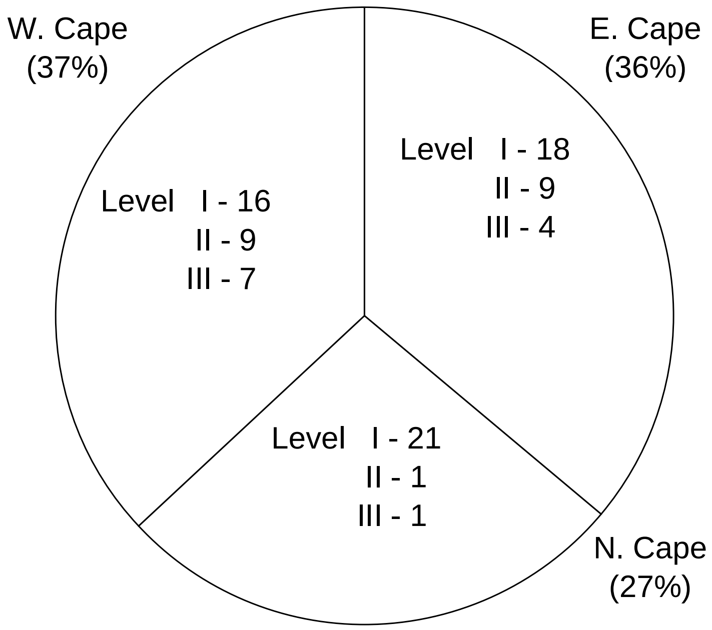
The statistical reports include data on inputs, outputs and other relevant hospital service statistics. For the purpose of this study, on the basis of their size and scope of activity the hospitals are classified into three groups. These in ascending order of their size and complexity are Level I (n=55), Level II (n=19) and Level III (n=12).

Recurrent inputs are expressed in monetary terms as total recurrent expenditure. The number of beds is provided as a measure of the capital input. Outputs are expressed as outpatient visits and inpatient days. For assessing productivity growth, data of Western cape Provincial hospitals (n=10) is used. The provincial distribution of the hospitals is presented in Figure 4.

4.3 Inputs and outputs

The selection of inputs and outputs for a DEA study requires careful thought as the distribution of efficiency is likely to be affected by the definition of outputs and the number of inputs and outputs included (Magnussen 1996). Conceptually, improved health status is the ultimate output of hospitals or the health care system at large. However, health status measurement poses difficulties because health is multi-dimensional and there is subjectivity involved in assessing the quality of life of patients (Clewer and Perkins 1998). Due to the difficulty in accurately measuring improvements in health status, hospital output is measured as an array of

FIGURE 4
PROVINCIAL DISTRIBUTION OF HOSPITALS STUDIED



intermediate outputs (health services) that supposedly improve health status (Grosskopf and Valdmanis 1987). Butler (1995) classifies hospital output into four broad categories: inpatient treatment, outpatient treatment, teaching and research.

Measuring hospital output by such variables as inpatient days or outpatient visits, does not capture the case-mix and the quality of service rendered. Even though the use of *Diagnosis-related Groups* (DRGs) may handle the problem of hospital case-mix, the absence of such data makes its use limited in most developing countries. Within the context of this study, it is assumed that stratifying hospitals according to their level may to some degree take account of the case-mix and factors such as staffing pattern and medical technology used that are likely to affect the quality of care delivered.

Inputs in hospital production are classified as labour, capital and supplies. The labour input can be disaggregated into the various professional groups such as physician, nurse and administrative staff. In most studies, capital is proxied by the number of hospital beds.

In this study, in Level I and II hospitals, input variables used are the *annual total recurrent expenditure* (in Rand) and the *bed-size*. The aggregate total recurrent expenditure includes the salaries of personnel, expenditure on drugs and other supplies. In Level III hospitals the input variable considered is only the recurrent expenditure, because of a small sample size. The outputs considered are *inpatient days* and *outpatient visits*. As the hospitals are non-academic, the teaching function is non-existent and a significant research output is not expected. These two outputs are most likely to capture the bulk of the activities performed by these hospitals.

4.4 The empirical DEA model

Input-oriented, constant and variable returns to scale DEA models are used in computing the efficiency scores. The choice between input-/output-oriented DEA models is made according to the flexibility of inputs or outputs. An input-oriented model is preferred in this study, because hospital managers are unlikely to have control of the demand side factors which are determined partly by the health-care seeking behaviour of the public.

A DEA-based *Malmquist productivity index* is computed to assess changes in the productivity of the Western Cape hospitals over a period of six years. This covers the period 1992/93-1997/98. The efficiency scores and the Malmquist productivity indices are computed using data envelopment analysis programme, version 2.1 (DEAP 2.1) (Coelli 1996).

5. RESULTS

5.1 General characteristics

The three levels of hospitals are found to have different sizes as measured by the bed-size. Level I community hospitals are the smallest. Those of Level II are about twice as large and Level III hospitals have the largest size, which is about eight times that of Level I.

There is also a marked gap in the activity levels of the three groups of hospitals. The volume of output in terms of admissions, outpatient visits and inpatient days is the highest in Level III hospitals. The mean recurrent expenditure in Level III hospitals is about twenty three times those of Level I hospitals. The summary statistics is presented in Appendix A.

5.2 Technical efficiency

The DEA models estimated for the three groups of hospitals indicate the presence of a marked deviation of the efficiency scores from the respective best-practice frontiers. Level I hospitals have the highest mean efficiency score. A summary of the efficiency scores is presented in Table 3.

TABLE 3
TECHNICAL EFFICIENCY SCORES

Technical efficiency measure	Mean	SD	Min	Max	Hospitals on frontier
			Level I Hospital		
CRS	0.740	0.124	0.518	1	6
VRS	0.828	0.174	0.468	1	17
Scale	0.900	0.124	0.518	1	6
			Level II Hospitals		
CRS	0.681	0.204	0.283	1	3
VRS	0.825	0.192	0.442	1	8
Scale	0.825	0.147	0.508	1	3
			Level III Hospitals		
CRS	0.695	0.162	0.516	1	2
VRS	0.820	0.125	0.671	1	3
Scale	0.845	0.140	0.641	1	2

The overall level of technical inefficiency in the three groups of hospitals is in the range 35.1 to 46.8 per cent. This is a combined inefficiency due to operations at a non-optimal scale (inappropriate hospital size) and pure technical inefficiency. It implies that on average hospitals use about 35.1 to 46.8 per cent more of resources than what is required for the given output levels. The decomposition of the overall level of inefficiency[10] into *scale* and *pure technical* inefficiency components is given as follows:

[10] Inefficiency is calculated as (1/efficiency score)-1.

TABLE 4
DECOMPOSITION OF OVERALL TECHNICAL INEFFICIENCY

Hospital	Pure technical inefficiency (%)	Scale inefficiency (%)
Level I	20.8	11.1
Level II	21.2	21.2
Level III	22.0	18.3

The magnitude of pure technical inefficiency seems more or less similar across the three Levels of hospitals. However, the scale inefficiency of Level I hospitals is much lower than those of the other two, suggesting that they are operating relatively closer to the optimal scale than the higher level hospitals. The above figures also indicate that more than 50 per cent of the overall inefficiency in the industry is attributed to pure technical inefficiency.

A significant proportion of hospitals in Level II and III operate at a non-optimal scale. Only about 58 per cent of Level II and Level III hospitals operate at a scale efficiency of 0.8 or more. The corresponding proportion for Level I hospitals is 84 per cent. About 75 per cent of Level I hospitals have an overall efficiency score of more than 0.6. On the other hand, only 50 per cent of the hospitals in the other two categories have a similar score. The following boxplots illuminate the distribution of the CRS efficiency scores across the three hospital groups.

The boxplots indicate that the median CRS technical efficiency score of Level I hospitals is higher than the other two Levels, which have about equal medians.

5.3 Returns to scale

Most of the hospitals in the three groups operate at variable returns to scale. Decreasing returns to scale is predominant in Level II and Level III hospitals, while increasing returns to scale is more prevalent in Level I hospitals. The figure below presents the distribution of RTS in the three groups of hospitals.

FIGURE 5
CRS TECHNICAL EFFICIENCY BY HOSPITAL LEVEL

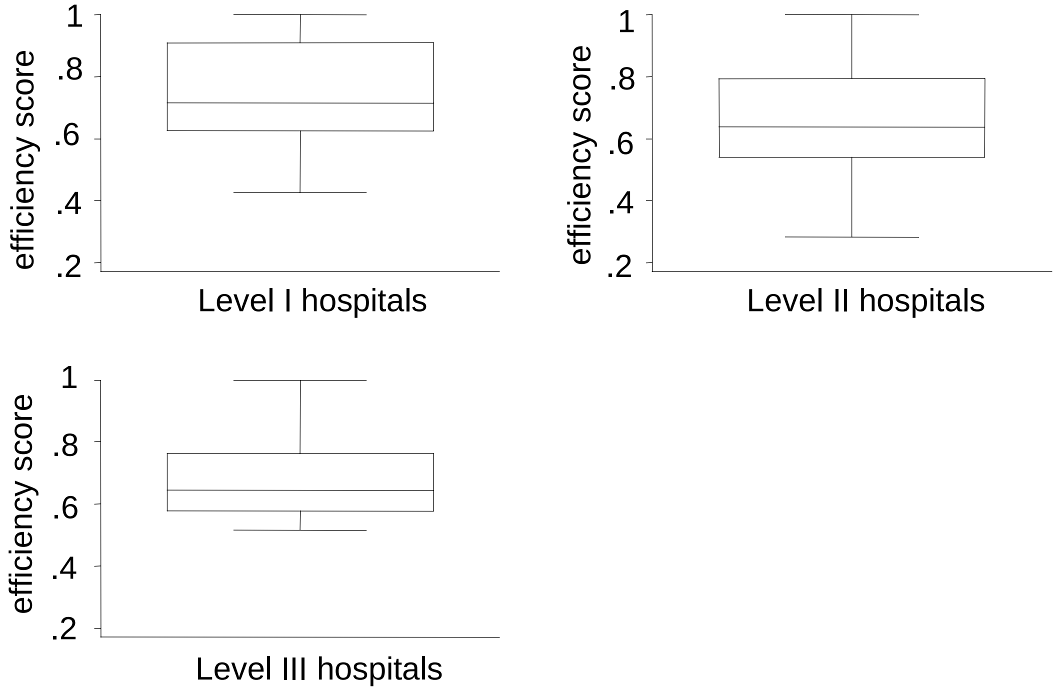
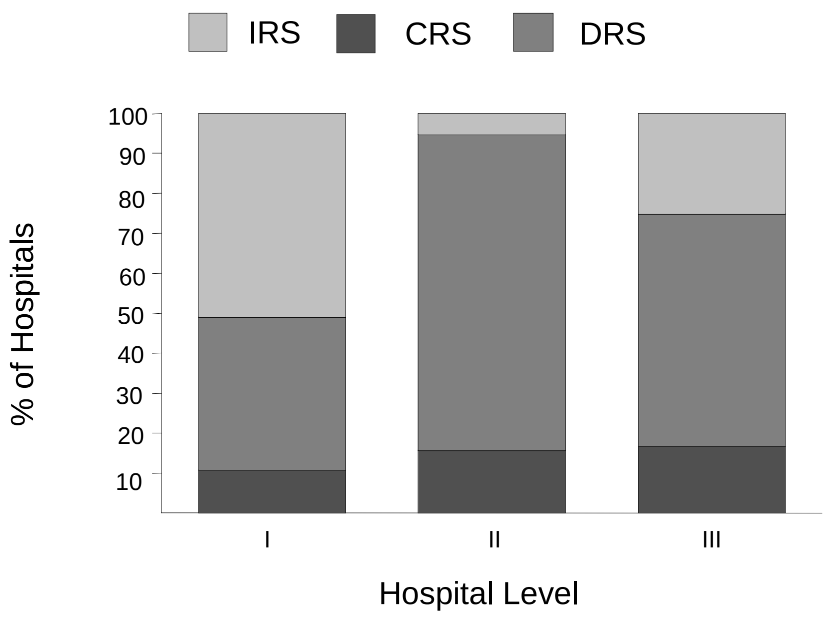


FIGURE 6
RETURNS TO SCALE BY HOSPITAL LEVEL



About half of the hospitals had decreasing returns to scale. On the cost side, this implies that half of them experience diseconomies of scale. Only about 13 per cent (11/86) of all the hospitals in the sample operate at an optimal scale.

5.4 Input savings

A. Level I

If the relatively inefficient hospitals operate as efficiently as their peers on the frontier, the average reduction in recurrent expenditure would be about R 601,748. This amounts to a total of R 29.5 million saving from this group as a whole. Similarly, the number of beds in the relatively inefficient hospitals could be cut by an average of 18 beds. The efficient bed size for this group of hospitals is projected to be about 44 beds – a decrease of about 29.5 per cent from a mean level of 61 beds.

B. Level II

In Level II hospitals, the reduction in recurrent expenditure as a result of pure technical and scale efficiency gains amounts to an average of R 3.5 million. In the group as a whole, this amounts to a saving of more than R 56 million. The number of beds could also be cut by an average of about 53 beds. This brings down the required average number of beds in the group from the current size of 112 to 68.

C. Level III

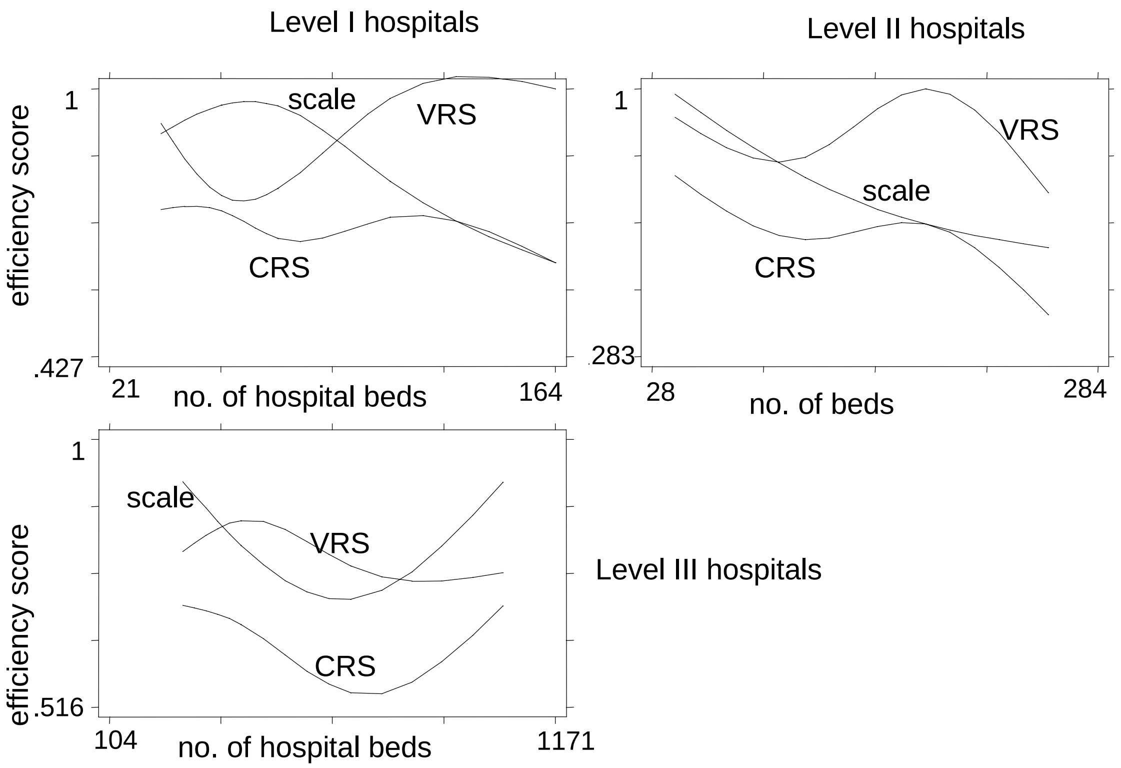
The average saving in recurrent expenditure in Level III hospitals is about R 19.4 million. This amounts to a total saving of about R 194 million. With respect to the number of beds, an average cut of 193 beds is expected if the hospitals operate on the technically and scale efficient frontier. This brings down the efficient bed-size to an average of 321.

5.5 Technical efficiency and bed-size

In relation to the number of beds, the overall technical efficiency in Level I hospitals reaches its highest mean value of 0.795 when the bed size is between 40-60. This is when the scale efficiency is on its increasing side and the pure technical efficiency declines. Efficient scale for these group of hospitals seems to be located within a bed size range of 60-80. Pure technical efficiency declines progressively until about a bed size of 80 and picks up thereafter. Thus the smaller (< 40 beds) and larger (> 80 beds) hospitals of Level I seem to have a higher degree of pure technical efficiency compared with their peers having a bed-size in the middle. However, the rise in technical efficiency at the higher levels of bed size is overshadowed by the steeply declining scale efficiency. Thus overall technical efficiency decreases when the increasing returns to scale at the smaller bed size levels are exhausted.

Out of 31 hospitals of Level I that had a bed size of less than 60, about 71 per cent experienced increasing returns to scale, indicating the existence of economies of scale. The proportion of those that experienced decreasing and constant returns to scale respectively was 9.7 and 19.3 per cent. The relationship between bed-size and technical efficiency is presented in Figure 7.

FIGURE 7
TECHNICAL EFFICIENCY AND HOSPITAL BED-SIZE



As in the case of Level I hospitals, the CRS technical efficiency in Level II hospitals also exhibits a trend of decline with increasing number of beds.

The pure technical efficiency seems relatively better with small (< 50) and large (> 150) bed sizes. The scale efficiency for this category of hospitals shows a progressive rate of decline, with the largest drop in efficiency (of about 26 per cent) occurring when the number of beds increases to over 150. Constant returns to scale is observed in only three hospitals (3/19) with bed-size of less than 50. Fifteen of the hospitals (15/19) with a bed-size of more than 50 have decreasing returns to scale, which in other words implies that they experience diseconomies of scale.

In Level III hospitals the CRS and scale efficiency plots assume a U-shape. Both Levels reach a relative minimum level at a bed-size of 600-800. After this minimum level there is a sustained increase in efficiency. This is contrary to Levels I and II where the minimal increase in CRS efficiency is not maintained as a result of a sharp drop in scale efficiency at higher levels of bed size. As can be seen from Figure 6, the pure technical efficiency behaves in a manner which is diametrically opposed to those of Levels I and II. It has a positive slope at the smaller bed size levels (< 400 beds) followed by various rates of decline as the bed size increases. Thus, whereas the pure technical efficiency increases at higher bed size levels in Levels I ad II, in Level III hospitals, it decreases tremendously. The decrease in pure technical efficiency is, however, more than offset by a greater increase in scale efficiency.

5.6 Technical efficiency and occupancy rate

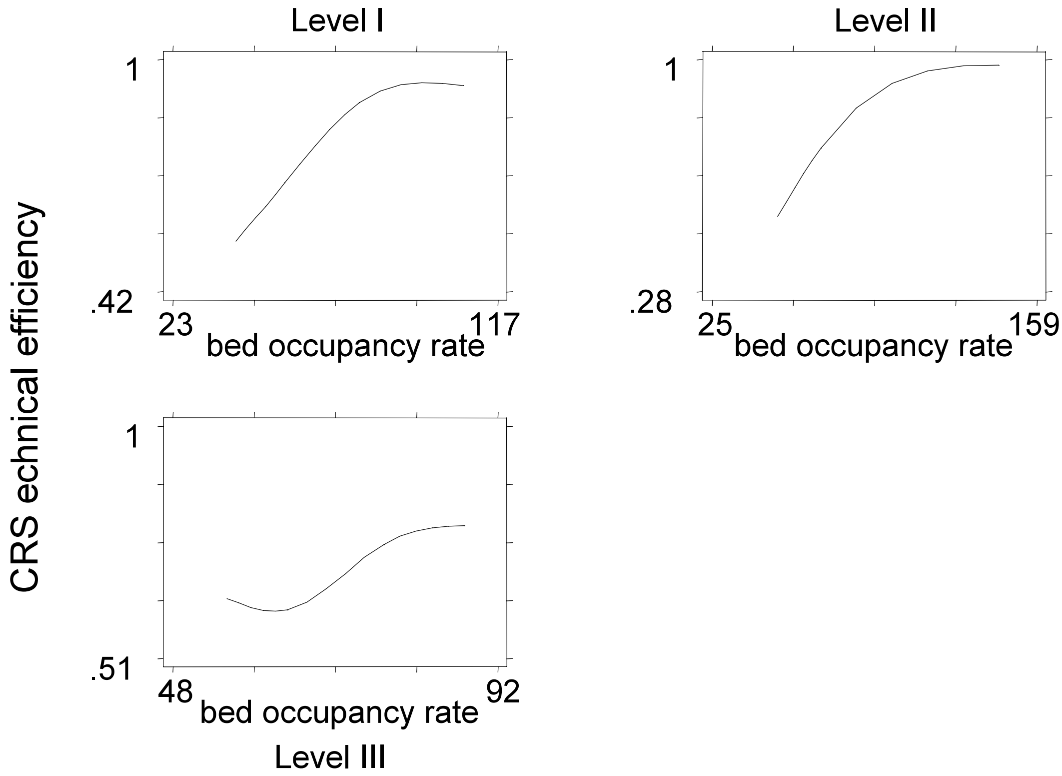
In all three types of hospitals, the overall technical efficiency increases with the increase in the occupancy rate. The highest levels of technical efficiency are seen in occupancy rates in the range of 80-100 per cent. Figure 8 depicts the relationship between efficiency scores and occupancy rate.

Technical efficiency tends to decline when the occupancy rate exceeds the 80-100 per cent mark.

5.7 Technical efficiency and average length of stay

The average length of stay does not seem to have a significant effect on the overall technical efficiency in Level I and Level II hospitals. However, in Level III hospitals, technical efficiency decreases by about 23 per cent as ALS increases from less than 5 days to more than 7 days. This can be easily discerned from figure 9.

FIGURE 8
THE EFFECT OF OCCUPANCY RATE ON TECHNICAL EFFICIENCY



As can be seen from Figure 9, the plots in Level I and Level II hospitals are almost flat, indicating that the impact of ALS on efficiency is very small. However, in Level III hospitals, the fitted curve is very steep. Thus we see a dramatic decrease in efficiency levels for small increases in ALS.

5.8 Provincial variations in technical efficiency

There is some degree of variation in efficiency levels among the three provinces. In Level I and Level III hospitals, the mean efficiency scores of the Western Cape province are higher than the other two provinces. However, in Level II hospitals those of the Eastern Cape Province are the highest. The distribution of the efficiency scores by province is given in Table 5 below.

In Level I and Level III hospitals, the mean efficiency score of hospitals in the Western Cape surpasses those of the Eastern Cape province by about 17 per cent. In contrast, in Level II hospitals, the mean efficiency score of hospitals in the Eastern Cape province outstrips that of hospitals in the Western Cape by about 12 per cent.

FIGURE 9
THE EFFECT OF ALS ON OVERALL TECHNICAL EFFICIENCY

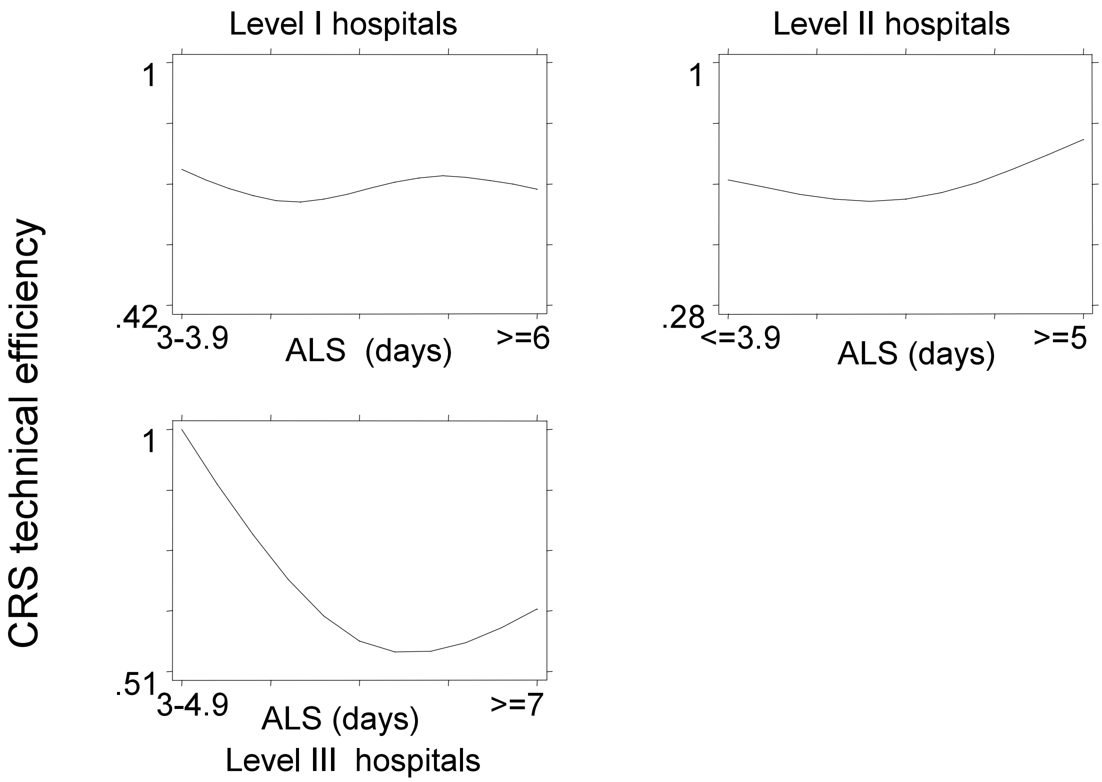


TABLE 5
DISTRIBUTION OF TECHNICAL EFFICIENCY SCORES BY PROVINCE

Province	Level I hospital			Level II hospital			Level III hospital		
	mean	SD	no.	mean	SD	no.	mean	SD	no.
Eastern Cape	0.69	0.18	18	0.73	0.25	9	0.64	0.09	4
Northern Cape	0.72	0.18	21	0.54	0	1	0.51	0	1
Western Cape	0.81	0.15	16	0.65	0.16	9	0.75	0.18	7

5.9 Econometric analysis of the determinants of inefficiency

Studies have shown that institutional factors at the discretion of the management as well as environmental factors beyond the control of the DMU affect a DMU's efficiency (e.g. Ferrier and Valdmanis 1996, Valdmanis 1992, Ozcan and Luke 1993, Rosko *et al*. 1995). Some of the factors that influence the efficiency of hospitals cited in the literature are:

ownership (profit versus not-for profit), location (urban/rural), teaching status (academic or non-academic), payment source, occupancy rate, quality *etc*.

The efficiency scores of Level I hospitals only are examined using a *censored tobit model* to identify factors influencing inefficiency. The other two types are excluded as their numbers are not sufficiently large to undertake a multivariate analysis.

In the tobit model, for computational convenience it is preferred to assume a censoring point at zero (Greene 1993). To this end, the DEA efficiency scores are transformed into inefficiency scores, left-censored at zero using the formula:

$$\textit{Inefficiency score} = (1/\textit{DEA score})-1 \qquad (11)$$

The model is defined as follows:

$$y_i^* = \beta_i x_i + u_i$$
$$y_i = y_i^* \text{ if } y_i^* > 0$$
$$y_i = 0 \text{ if } y_i^* \leq 0$$

Where $u_i \sim N(0,\sigma^2)$, and

y_i	is the observed inefficiency score
β_i	is a kx1 vector of unknown parameters
x_i	is a kx1 vector of explanatory variables

Due to data constraints, some important variables within the hospitals and their operating environment have been omitted. The empirical model, therefore, takes the following form:

$$INEFF = \alpha_0 + \beta_1 \text{OCC} + \beta_2 \text{ALS} + \beta_3 \text{OUTPRO} + \beta_4 \text{PROV1} + \beta_5 \text{PROV2} + \varepsilon_i \qquad (12)$$

Where:	INEFF	inefficiency score
	OCC	bed occupancy rate (%)
	ALS	Average length of stay (days)
	OUTPRO	outpatient visits as a proportion of inpatient days
	PROV1	location dummy:
		= 1 if Eastern Cape

	= 0 otherwise
PROV2	location dummy: = 1 if Northern Cape = 0 otherwise

The occupancy rate is a composite index that incorporates inpatient admissions, the average length of stay and the number of beds. However, multicollinearity is assumed to be less of a problem as the value of the occupancy rate is determined by the relative position of each of its components and not a single one. Furthermore, a simple correlation analysis has shown the absence of a strong relationship among the variables.

Statistical Analyses are performed using STATA 5 statistical software (Statacorp 1997).

The regression results are presented in Table 6.

TABLE 6
ESTIMATION RESULTS FOR TOBIT MODEL

Variable	Coefficient	t-ratio
constant	1.9112	8.216
OCC	-0.0171	-7.940
ALS	-0.0319	-1.289
OUTPRO	-1.5110	-5.407
PROV1	0.0282	0.292
PROV2	-0.0819	-0.842
$\chi^2_{(6)}$	56.72	

The bed occupancy variable has a sign consistent with our expectation. It is negatively related to inefficiency. This implies that higher occupancy levels are associated with higher level of efficiency. The coefficient of *ALS* has a negative sign in line with our *a priori* expectation, although it is not statistically significant. The number of outpatient visits as a proportion of inpatient days (*outpro*) has a very high statistical significance. In this category of hospitals, an increase in the number of outpatient visits relative to inpatient days is likely to result in an increase in overall efficiency levels. The location of a hospital (*prov1*, *prov2*) has no significant bearing on efficiency.

5.10 Productivity growth

Input-output data for the years 1992/93–1997/98 were used in the analysis. This interval includes a period of structural break in South African history. It encompasses both the apartheid and post-apartheid periods.

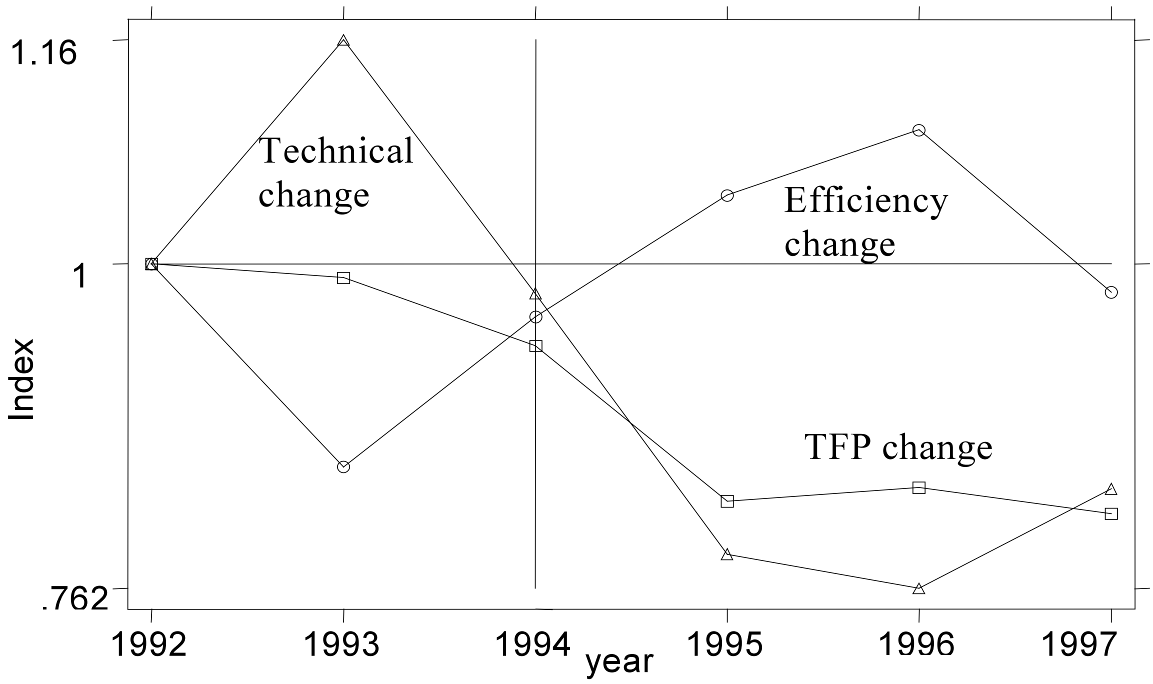
Over the sample period, total factor productivity (TFP) dropped by 12.1 per cent. This is largely due to a decline in technical progress. The drop in technical efficiency is marginal. Technical efficiency increased in the two immediate years after 1994/95. Over the years 1992/93-1997/98, technical efficiency dropped by an average of 2.1 percent, as opposed to a 16.5 percent decrease in technological growth. Efficiency change and technical change are observed to move in opposite directions (Figure 10). The MPI summary of the annual means is presented in Table 7.

TABLE 7
MPI SUMMARY OF ANNUAL MEANS

Year	Efficiency change	Technical change	TFP change
1992/93[11]	1	1	1
1993/94	0.851	1.164	0.990
1994/95	0.961	0.978	0.940
1995/96	1.050	0.787	0.826
1996/97	1.098	0.762	0.836
1997/98	0.979	0.835	0.817
Mean	**0.984**	**0.893**	**0.817**

[11] 1992/93 is the base year.

FIGURE 10
PRODUCTIVITY CHANGE, 1992/93-1997/98



6. DISCUSSION

The results of this study provide preliminary empirical evidence on the performance of public sector hospitals in three provinces of South Africa. The findings suggest that many of the hospitals operate at pure technical and scale efficiency levels well below a best-practice frontier that is determined by the relatively efficient ones from the group. Only about 12.8 per cent of the hospitals operate efficiently as compared to their peers.

It should be borne in mind that DEA is not a measure of absolute efficiency and that the efficiency scores only reflect the performance and production technology of the group. It is possible that the addition of more hospitals in the sample is likely to increase the number of inefficient hospitals.

A level of technical inefficiency in the range of 35 to 47 per cent indicates that significant amounts of health care resources are being wasted. This finding supports the commonly held view that Africa's health facilities (and, in particular, hospitals) operate at unacceptable levels of technical inefficiency. Technical inefficiency is present to varying degrees, in most hospitals in both the developed and developing countries (see for example, Wouters 1993, McMurchy 1996, Ersoy *et al.* 1997, Ferrier and Valdmanis 1996, Hao and Pegles 1994, Ozcan *et al.* 1996, Rosko and Chilingerian 1999). However, in Sub-Saharan Africa, not many efficiency studies have been conducted using the frontier models. Thus, there is no clear and quantifiable evidence on the type and degree of inefficiency.

The efficiency estimates of this study indicate that if the inefficient hospitals were to operate on the best-practice frontier, recurrent expenditure could be reduced by about R 279.5 million. This is more than the annual budget for tuberculosis. In 1992/93, public sector hospitals in South Africa collected an average of 9 per cent of their recurrent expenditure in fee revenue (McIntyre *et al.* 1995a). This implies that efficiency savings exceed by far the cost-recovery from user fees. Any benefits that user fee schemes may have, are likely to be compromised as a result of technical inefficiency. Thus it is necessary that health sector reforms give due emphasis to efficiency.

A study in South Africa estimated the cost of building 1,000 clinics at between R 400 million and R 1 billion and recommended the improvement of hospital efficiency as a possible source of funding (McIntyre *et al.* 1995b). The findings of our study seem to suggest that if current deviations from best-performance frontiers are minimized, efficiency savings from hospitals alone would possibly cover the amount of resources necessary to increase health care coverage. If we take the lower limit of the cost of clinic construction, then the efficiency savings from the hospitals included in this study alone could cover the cost of building about 698 clinics. Taking the upper limit would mean that the cost of about 279 clinics could be covered from resources released from these hospitals. This is of immense importance to the government in its initiatives to promote equity.

If the economy's poor performance continues then it will be very difficult to significantly increase the amount of resources devoted to health. This implies that it will be very difficult to meet the objective of health for all. Thus, raising the efficiency of hospitals is essential to releasing scarce resources for basic health care. However, not all the efficiency gain should be reallocated to basic health care facilities, since there is much to be done to raise the quality of care in hospitals to acceptable standards. Thus a portion of the efficiency gains should be retained by the hospitals themselves.

Bed-sizes in all three groups of hospitals appear to exceed what is required for the given output levels. But, a caveat is in order here. The fact that the number of beds is more than what is required does not imply that the number of beds exceeds the population's *need* for hospital services. The public sector hospital bed-population ratio for the country as a whole is low by international standards (McIntyre *et al.* 1995a). The utilization of existing hospital facilities, however, depends not only on supply-side factors, but also on demand-side factors. Therefore, given the current demand levels, it can be safely argued that the existing number of beds is in excess of what is required. The results of a number of studies on hospital efficiency in other countries indicate the redundancy of beds (Ozcan *et al.* 1996, Brownell and Roos 1995).

Increasing returns to scale is seen in about 37 per cent of the hospitals evaluated. Level I hospitals have the largest proportion of hospitals (50.9 per cent) with increasing returns to scale. In the presence of increasing returns to scale, expansion of outputs reduces unit costs. A hospital with increasing returns to scale will, therefore, benefit by augmenting its scale of operations. Economies of scale may occur as a result of staff being able to specialize in their areas of expertise, the ability to spread overhead costs over a larger number of output units, discounts from bulk-buying of supplies and the ability to use expensive diagnostic equipment at full capacity.

However, hospital outputs are not like other commodities, which can be stored for future use. Increasing the scale of operations requires an increase on the demand side, which is beyond the domain of the hospital management. Merger of hospitals that are in close proximity to one another may be a possible option in this case. This option may, however, pose some problems, especially in sparsely populated rural areas. If a few hospitals of a bigger size are to be established in centrally located places, residents of such areas may incur additional costs in travel expenditure and in delayed treatment of emergency cases. These problems may, to some extent, be minimized by establishing PHC units that have a link with the centrally located hospitals through an effective referral system. It should, however, be emphasized that any initiatives undertaken to reap economies of scale must be implemented only after a careful appraisal of the circumstances surrounding the operation of the hospital(s) under consideration.

About half of the hospitals have decreasing returns to scale. The greatest proportion of hospitals with decreasing returns to scale is found in those of Level II (78.9 per cent) followed by Level III (58.3 per cent). Decreasing returns to scale imply that a hospital is too large; its unit cost rises with each increment in output. Diseconomies of scale may arise as a result of problems such as red-tape, poor communications and poor labour relations which are often encountered in large organizations.

The presence of decreasing returns to scale implies that the scale of operations should be reduced so as to curb the unnecessarily inflated costs of production and render the inefficient hospitals as efficient as their peers. This is an option that has to be considered seriously as it is also likely to promote the government's move towards the re-allocation of resources away from hospital-based services to cost-effective non-hospital services in line with the Primary Health Care strategy.

In Level I and II hospitals overall technical efficiency shows a trend of decline with the increase in the number of beds. Efficiency is at its highest at bed-sizes in the first quartile. The small bed-size hospitals have a relatively higher overall efficiency score than the larger ones. In contrast, in Level III hospitals efficiency scores for the small and large bed-size hospitals are higher than those in the middle. This is in line with the findings of Ferrier and Valdmanis (1996) who in their study of a sample of rural US hospitals found the relationship between technical efficiency and size to be U-shaped. They attributed this variation partly to the fact that patients tend to bypass medium-sized hospitals in their vicinity in favour of larger hospitals with more advanced technologies or better facilities.

Technical efficiency is observed to increase with increases in the occupancy rate in all three Levels of hospitals. This finding corroborates that of Ferrier and Valdmanis (1996). The average length of stay has a negative relationship with efficiency in Level III hospitals. An increase in the average length of stay is followed by a dramatic fall in the technical efficiency scores. This may perhaps be a result of an excess bed capacity which might not be constraining the clinicians and thus allowing them to keep patients for a long time. In this situation a high average length of stay is likely to lead to high average costs per admission (Barnum and Kutzin 1993).

There appears to be a slight variation in the efficiency scores among the provinces. The Tobit analysis of Level I hospitals, however, shows that it is not of statistical significance. Variation could result from differences in regulatory environments, demographic characteristics *etc.* (Ferrier and Valdmanis 1996). In present circumstances, the issue of differences in regulatory environments is ruled out, because all of the hospitals are public sector hospitals under the same department. Differences in the factors affecting the demand-side may have a greater impact as there are striking differences in the socio-economic and demographic characteristics of the provinces.

In Level I hospitals, the number of outpatients as a proportion of inpatient days seems to have a significant negative influence on inefficiency. Thus increasing the number of outpatient visits is likely to promote efficiency. This may indicate the existence of economies of scope between outpatient and inpatient care. Since this category of hospitals has a very limited outpatient activity, increasing the activities of the outpatient's department is likely to promote technical efficiency.

The MPI for a sample of hospitals in the Western Cape province shows a total factor productivity decline of about 12.1 per cent over the sample period. This change has important implications as it represents a change that has taken place in the immediate post-apartheid era, when the government is faced with the difficult task of increasing access to health care and redressing the inequalities of the past. The marked technical regress seen after the installation of the new government may, with all the necessary precautions, be taken as an indication of its initiatives to divert resources away from hospitals consistent with its policy of Primary Health Care. These findings are, however, indicative and for conclusive results it is recommended that the evaluation be undertaken on a sufficiently large data set.

7. CONCLUSION AND POLICY IMPLICATIONS

This study examined the technical efficiency and productivity of 86 public sector hospitals in three provinces of South Africa using DEA. The findings lend support to the commonly held view that public sector hospitals in developing countries are inefficient.

The results indicate the existence of a huge potential for improving access and/or quality of care without injecting additional resources into the health sector from any source, public or private. The evidence further suggests that efficiency savings exceed revenue from user fees. Efficiency savings could augment the gains from user fees in terms of mobilizing additional resources, and increase cost-recovery ratios. Alternatively, if the current user fee rates are deemed high, it is possible to reduce prices and achieve the current cost-recovery ratios, thus avoiding the possible untoward effects of user fees on the most vulnerable groups of the population.

The fact that bed-size in all three levels of hospitals is large and that decreasing returns to scale dominate the picture may justify downsizing most of the hospitals. Allocative efficiency may be improved if these resources are re-deployed for the promotion of Primary Health Care, which is at the core of the Government's health policy. It should, however, be noted that at this juncture the short-run costs of re-allocation might exceed the gains from efficiency savings (for example, substantial amounts of staff relocation costs may be incurred). A slashing of the current budget of the health sector or any proposed increase is not recommended, as there is a dire need to improve access and quality of care.

There is some evidence that suggests a decline in total factor productivity. This is an issue of measure concern, as it is likely to be a serious impediment to the Government's efforts to improve access and quality of health care. With such a degree of inefficiency, initiatives to redress the inequity that was inherited from the previous regime will be jeopardized. The issue of hospital efficiency, therefore needs to be addressed with due concern to equity if the Government's stated goals of Health Policy are to be realized.

APPENDIX

Summary Statistics

Variable	Mean	Standard deviation	Min	Max
Level I hospitals				
Bed	61	26	21	164
Admission	2,626	1,557	530	7,343
Outpatient visit	2,498	2,230	172	9,183
Inpatient day	13,066	6,799	2,572	40,969
Occupancy (%)	58.6	17.9	23.6	117.2
Average length of stay (days)	5.1	1.4	3.4	10.8
Bed turnover rate	42	15.4	12.3	81.9
Expenditure (Rand)	2,069,587	1,311,240	434,913	7,164,874
Level II hospitals				
Bed	112	67	28	284
Admission	5,775	3,115	1,312	13,854
Outpatient visit	37,390	35,986	2,057	146,524
Inpatient day	26,188	14,320	7,891	61,852
Occupancy (%)	71.7	29.2	25.8	159.6
Average length of stay (days)	4.7	1.4	2.6	9.2
Bed turnover rate	52	16	35	107
Expenditure (Rand)	8,987,729	8,056,606	879,665	36,500,000
Level III hospitals				
Bed	483	320	104	1,171
Admission	19,572	10,298	6,436	40,998
Outpatient visit	120,788	96,200	13,264	357,808
Inpatient day	118,994	85,625	35,025	328,806
Occupancy (%)	68.6	12.4	48.5	92.3
Average length of stay (days)	5.9	1.6	3.6	9
Bed turnover rate	46.3	17.7	20	84
Expenditure (Rand)	48,500,000	36,700,000	11,300,000	127,000,000

REFERENCES

Akin, J., N. Birdsall and D. deFerranti (1987). *Financing Health Services in Developing Countries: An Agenda for Reform*. Washington DC: The World Bank.

ANC (African National Congress) (1994). *A National Health Plan for South Africa*. Johannesburg: ANC.

Banker, R. D., A. Charness and W. W. Cooper (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(1): 1078-1092.

Barnum, H. and J. Kutzin (1993). *Public Hospitals in Developing Countries: Resource Use, Cost, Financing*. Baltimore: Johns Hopkins University Press for the World Bank.

Bjurek, H., L. Hjalmarsson and F. R. Forsund (1990). Deterministic parametric and non-parametric estimation of efficiency in service production: A comparison. *Journal of Econometrics*, 46(1/2): 213-227.

Brownell, M. D. and N. P. Roos (1995). Variations in length of stay as a measure of efficiency in Manitoba hospitals. *Canadian Medical Association Journal*, 152(5): 675-682.

Butler, J. R. (1995). *Hospital Cost Analysis*. Dordrecht: Kluwer Academic Publishers.

Castro-Leal, F., J. Dayton, L. Demery and K. Mehra (1999). Public social spending in Africa: Do the poor benefit? *The World Bank Research Observer*, 14(1): 49-72.

Caves, D., L. Christensen and E. Diewert (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica*, 50(6): 1393-1414.

Charness, A., W. W. Cooper and E. Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2: 429-444.

Clewer, A. and D. Perkins (1998). *Economics for Health Care Management*. London: Prentice Hall.

Coelli, T. J. (1996). *A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program.* CEPA Working Paper 96/8. Department of Econometrics, University of New England.

Coelli, T., D. S. P. Rao and G. Battese (1998). *An Introduction to Efficiency and Productivity Analysis*. Boston: Kluwer Academic Publishers.

Debreu, G. (1951). The coefficient of resource utilisation. *Econometrica*, 19, 273-292.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Ersoy, K., S. Kavuncubasi, Y. A. Ozcan and J. M. Harris 2nd (1997). Technical efficiencies of Turkish hospitals: DEA approach. *Journal of Medical Systems*, 21(2): 67-74.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1994). Productivity developments in Swedish hospitals: A Malmquist index approach, in Charness A., W. W. Cooper, A. Y. Lewin and L. S. Seiford (eds) *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston: Kluwer Academic Publishers, 253-272.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, 120(3): 283-281.

Ferrier, G. D. and V. Valdmanis (1996). Rural hospital performance and its correlates. *The Journal of Productivity Analysis*, 7: 63-80.

Greene,W. H. (1993). *Econometric Analysis* (2nd edition). New York: Macmillan.

Grosskopf, S. and V. Valdmanis (1987). Measuring hospital performance: A non-parametric approach. *Journal of Health Economics*, 6: 89-107.

Hao, S. and C. C. Pegles (1994). Evaluating relative efficiencies of veterans affairs medical centers using data envelopment analysis. *Journal of Medical Systems*, 18(2): 55-67.

Koopmans, T. C. (1951). An analysis of production as an efficient combination of activities, in Koopmans, T. C. (eds), *Activity Analysis of Production and Allocation*. New York: John Wiley and Sons Inc.

Lovell, C. A. K. (1993). Production frontiers and productive efficiency, in Fried H. O., C. A. K. Lovell and S. S. Schmidt (eds), *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.

Magnussen, J. (1996). Efficiency measurement and the operationalisation of hospital production. *Health Services Research*, 31(1): 21-37.

McIntyre, D., G. Bloom, J. Doherty and P. Brijlal (1995). *Health Expenditure and Finance in South Africa*. Durban: Health Systems Trust and the World Bank.

McIntyre, D., G. Bloom, J. Doherty and P. Brijlal (1995). *Key Findings of the Review of Health Expenditure and Finance in South Africa*. Durban: Health Systems Trust and the World Bank.

McIntyre, D., L. Gilson, N. Valentine and N. Söderlund (1998). *Equity of Health Sector Revenue Generation and Allocation: A South African Case Study*. Major Applied Research 3, Working Paper 3, Bethesda, MD: Partnerships for Health Reform Project, Abt Associates Inc.

McMurchy, D. (1996). *Efficiency Studies of Public Sector Hospitals in the Eastern, Northern and Western Cape Provinces*. HEU Working Paper No. 6. Cape Town: Health Economics Unit, University of Cape Town.

Ozcan, Y. A. and R. D. Luke (1993). A national study of the efficiency of hospitals in urban markets. *Health Services Research*, 27(6): 719-739.

Ozcan, Y. A., M. J. McCue, and A. A. Okasha (1996). Measuring the technical efficiency of psychiatric hospitals. *Journal of Medical Systems*, 20(3): 141-50.

Rosko, M. D. and J. A. Chilingerian (1999). Estimating hospital inefficiency: Does case-mix matter? *Journal of Medical Systems*, 23(1): 57-71.

Rosko, M. D., J. A. Chilingerian, J. S. Zinn and W. E. Aaronson (1995). The effects of ownership, operating environment, and strategic choices on nursing efficiency. *Medical Care*, 33(10): 1001-1021.

Seiford, L. M. and R. M. Thrall (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics*, 46: 7-38.

Statacorp (1997). *Stata Statistical Software: Release 5.0*, College Station, TX: Stata Corporation.

UNDP (United Nations Development Programme) (1999). *Human Development Report 1999*. New York: Oxford University Press.

Valdmanis, V. (1992). Sensitivity analysis for DEA models: An empirical example using public vs. NFP hospitals. *Journal of Public Economics*, 48: 185-205.

Van Rensburg, H. C. J., A. Fourie and E. Pretorius (1992). *Health Care in South Africa: Structure and Dynamics*. Pretoria: Academica.

World Bank (1998a). *African Development Indicators 1998/99*. Washington DC: The World Bank.

World Bank (1998b). *World Development Indicators 1998*. Washington DC: The World Bank.

World Bank (1999). *World Development Report 1998/99*. Washington DC: The World Bank.

Wouters, A. (1993). The cost and efficiency of public and private health care facilities in Ogun State, Nigeria. *Health Economics*, 2(1): 31-42.