TECHNICAL EFFECIENCY OF HOSPITALS: ACASE OF PUBLIC, MISSION AND PRIVATE HOSPITALS IN NORTH RIFT REGION, KENYA

BY

CLEMENT CHERUIYOT TISON

THESIS SUBMITTED TO IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF ARTS IN ECONOMICS

SCHOOL OF BUSINESS AND ECONOMICS
MOI UNIVERSITY

NOVEMBER 2014

DECLARATION

DECLARATION BY CANDIDATE

I do declare that this thesis is my original work and has not been submitted for examination in any other institution for award of any degree.

……………………….. ………………………

Clement Cheruiyot Tison **Date**

Reg No.SBE/PGE 018/2008

DECLARATION BY SUPERVISORS

This Research Thesis has been submitted to the Department as a copy with our approval as University Supervisors.

Prof.Mary.J.Kipsat
Professor- School of Business and Economics
Moi University
Kenya

…………………………….. ……………………..

Date

Dr. Simion.K Ruttoh
Senior Lecturer
Department of Health Policy and Management of Public Health
College of Health Sciences
Moi University
Kenya

………………………………………. ….…………………..

Date

DEDICATION

To my wife Rebecca and our children, Nickline, Delvine and Nicessive for their ever lasting love, affection and dedication.

ACKNOWLEDGEMENTS

My first gratitude goes to the Almighty God for guiding and taking care of me, his name be praised forever, AMEN.

My gratitude is extended to my supervisors: Prof.Mary.J.Kipsat and Dr. Simion.K Ruttoh for their guidance and professionalism that shaped this study and enlighten me intellectually.

God bless you richly.

Special thanks to the Head of Department: Dr.Michael.K.Korir for his support and encouragement, which took me through the academic pressures.

Thank you.

ABSTRACT

The demands for health care resource have been increasing overtime in Kenya. Quantitative methods of measuring the efficiency of management and resources use are increasingly being applied in the health care sector. This study focuses on the measurement of technical efficiency of public, mission and private hospitals in North Rift Region of Kenya. The specific objectives of the study were to compare levels of technical efficiency among selected hospitals in North Rift Region and to identify the major determinants of variations in technical efficiency in the hospitals. Although the number was 43, a purposive sampling technique was used to pick all the 28 hospitals with admission facilities. This represented 65.12 % of the target population. Document analysis was used as the instrument of data collection. Comparative analysis, descriptive statistics and DEAP version 2.1 computer program was applied in analyzing data. The study used data envelopment Analyses methodology to identify and measure individual public and private hospital efficiencies. This involves the standard Constant Return to Scale and Variable Return to Scale models that involve the calculation of Technical and Scale efficiencies. The key results were as follow: The overall average level of TE among hospitals in the North Rift is 83.7%. Under pure technical efficiency, 10 of the 28 hospitals (35.71%) are relatively inefficient compared with the other hospitals in the data set. On the scale efficiency side 15 of the 28 hospitals (53.57%) were scale inefficient. In total the following inputs are wasted and not utilized in the production of outputs among hospitals in the North Rift Region, 16.978 doctors (12.04%), 211.377 nurses (13.58%), 35.943 technicians (15.97%), 14.092 administrative staff (12.25%), 153.748 general staff (16.44%), and 279.68 beds (13.41%).Under pure technical efficiency scores the mission hospitals are the best performers with average score of 1.000, While under scale technical efficiency scores the public hospital have highest average scores of 0.9292 The study gained empirical knowledge about efficiency of public, mission and private hospitals. This knowledge about differences in efficiency will contribute to the public policy debate on how to improve the efficiency with which the government spend their scarce resources for the provision of health care to the population as has been noted by the Health Reform Strategy in Kenya. The findings of the study are intended to benefit health financing, hospital management, academic, policy makers, in government and general public. The study recommends that excess labour force and beds be relocated to the under staffed primary healthcare facilities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATION

AIDS	Acquired Immune Deficiency Syndrome
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DHT	District Health Team
DMU	Decision Making Units
EU	European Union
GDP	Gross Domestic Product
HIV	Human Immune Deficiency Virus
MOH	Ministry Of Health
NACA	National Aids Coordinating Agency
NDP	National Development Plan
SE	Scale Efficiency
SFM	Stochastic Frontier Method
SIDA	Swedish International Development Fund
TE	Technical Efficiency
UNAIDS	United Nations AIDS Program
UNDP	United Nations Development Program
UNICEF	United Nations Children Fund
USAID	United States Agency for International Development
VRS	Variable Returns to Scale
WB	The World Bank
WHO	World Health Organization.

DEFINITION OF TERMS

Efficiency The state, quality operational of producing a desired or satisfactory results,

Technical It is practical use of machinery, method and skill needed for particular job.

Production efficiency (economic efficiency): It is the degree to which the observed use of resources to produce outputs of a given quality matches the optimal use of resources to produce outputs of a given quality. This is composed of two intervening components technical and allocative efficiency.

Technical efficiency: It is expressed as the potential to increase quantities of output from given quantities of inputs or the potential to reduce the quantities of inputs used in producing given quantities of outputs. It is defined independent of prices or costs. It is decomposed to scale efficiency and pure technical efficiency. It is sometimes referred to as efficacy.

Allocative efficiency: Deals with whether, for any level of production, inputs are used in the proportion that minimizes the cost of production, given input prices. Therefore in allocative efficiency price information becomes necessary. It is sometimes referred to as appropriateness.

Sale efficiency: Proportion of technical efficiency where an organization can take advantage of returns to scale by altering its size towards optimal scale.

Non-scale technical efficiency / pure technical efficiency: Proportion of technical efficiency, which cannot be attributed to divergences from optimal scale (scale efficiency) that is, solemnly deals with input –output combinations. It is sometimes known as managerial efficiency.

Decision Making Units (DMU) these are the organization or units being measured in a Data Envelopment Analysis measurement technique.

Best Practice. The set of management and work practices that result in the highest potentials or optimal quality combination of outputs for a given quantity and combination of inputs (productivity) for a group of similar organizations.

Peers: In Data Envelopment Analysis model, these constitute a group of best practice organizations with which a relatively inefficient organization is compared.

Efficiency reference set: Are the group of hospitals against which DEA (Data Envelopment Analysis) locates the inefficiency hospitals and the magnitude of inefficiency.

Slacks. The extra amounts by which an input (output) can be reduced (increased) to attain technical efficiency after all inputs (outputs) have been reduced (increased) in equal proportion to reach the production frontier. This is a feature of piece –wise linear production frontier derived when using Data Envelopment Analysis.

Output; The quantity of goods produced.

Input; It is a substance, idea piece of information provided during the process

CHAPTER ONE

INTRODUCTION

1.1 Introduction

This chapter discuses the background of the study, statement of research problem, research questions, objectives, justification, significance, scope, study limitation, theoretical and conceptual framework of the study.

1.2. Background of the Study

Health care is a basic human need, but in most developing countries the prospect of achieving even minimal adequacy of health and health services is a distant goal. The basic health needs of vast numbers of people remain unmet and the pursuit of improved standards of health has become a primary concern over recent years. The nature and scale of health sector problems have been described and have become better recognized (Maynard and et al Kanavos, 2000). In response to such problems the World Health Organization (WHO) declared in 1977 that the main social target of governments should be the attainment by all citizens of the world by the year 2000 of a level of health that will permit them to lead a socially and economically productive life.

There are many factors that influence the health of an individual or population. Bloom and Lucas (1999) identify some of these factors , which include, access to nutritious food, clean water, adequate clothing and shelter and the means for hygienic disposal of human wastes; positive leisure behaviour, freedom from contamination and skills (education levels). They further state that many if not all factors tend to improve with rising incomes and that the common experiences are that declining levels of poverty are generally associated with increasing health status.

The Kenyan population has grown ignorantly from a modest count of 2.5 million in the first census of 1897 to 39,610, 097 million in the year 2012 and 2013. Fifty years later the number had more than doubled to 5.4 m in 1948 and by the time the country approached independence the population was hardly the current Rift valley population, at merely 8.6 million.[Republic of Kenya 2009 Census]

Seven years later, in 1969 the number hit two digit figure of 10.9 m and since then the census has been done after each decade such that in 1979 the population had reached 15.3 million not accounting deaths which are silently relating to not less than 10 million to date.

By the year 1989, the population status had reached 21.4 million by the turn of this century in a census that was largely blamed on flawed grounds only registering an approximate figure. Table 1.1 shows the population of Kenya from 1897 to 2009.

Table 1.1 shows the population of Kenya from 1897 to 2009

1897	2.5 m
1948	5.4 m
1962	8.6 m
1969	10.9 m
1979	15.3 m
1989	21.48m
1999	28.7 m
2009	38.6m

Source: Republic of Kenya 2009 Census

1.2 Health Sector of Kenya

1.2.1 Background of the Health Sector

At independence in 1963, the Kenya government inherited an economic structure that was underdeveloped even by African standards. The health sector was no exception. In 1955, there were 22 government hospitals (with 1045 beds and cots) and 36 missionary hospitals (with 868 beds and cots). The health program had put more emphasis on secondary health care as opposed to primary health care, the great majority of the people had no access to health services at all (Harvey and et al 2002). Kenya only had one doctor per 32,006 people and only one nurse for every 26,000 people in 1965, though this was considered to be better than the average in Sub-Sahara Africa, which was one doctor per 39,000 people and one nurse for 57,000 people. (Harvey and et al 2002)

Despite all these short falls in the health sector, Kenya's health statistics of 1965 tended to show that Kenya was better off than the average Sub-Sahara Africa, with higher life expectancy and lower infant and child mortality, and this was attributed to :

The hot dry climate in most of part of the county, which is not conducive many of Africa's debilitating diseases. 2. The abundant availability of milk and meat from the large livestock population which provided Kenya with relatively good protein intake and. 3. The missionary provision of health care, as opposed to government provision (though the government spending in health had increased from the year 1964 to 1990) (Harvey and et al 2002)

From 1973, there was a major shift in emphasis from the previous hospital based curative system (secondary health care), towards a decentralized primary health care system, which included promotive, preventive and rehabilitative services. With this came the establishment of a network of basic health facilities throughout the country, which reduced the maximum distance from a health post from 15 to 8 kilometres. It is worth noting that the role of the traditional healers in primary health care is also recognized by the Ministry of Health, there are about 2,000 traditional practitioners in the country who are herbalists, faith healers or diviners (NDP 6 Mid-term Review 2008) in the death rate in this country. There are many challenges in Kenya's fight against HIV / AIDS and these include among others problems inherent in coordination of a multi-sectoral response to disease, Stigmatization of HIV / AIDS patients, Increasing demand for counseling and Community Based Care, Overcrowding in hospitals and Shifting of resources within the health system from other health problem to HIV / AIDS problem (NDP 8 Mid-term Review 2010).

1.3. Statement of the Problem

The expenditure in health care (be it as a result of increase in demand for health or due to technological changes) are financed by government through general revenues and user fees. Therefore sources of finances have raised the question of sustainability of the health care system in the long –run. It is because of this that the Ministry of Health decided to make sustainability (financial sustainability and system sustainability), quality and Appropriateness (efficiency) the main objectives for the remainder of NDP(National Development Plan). These objectives are related to funding and to management of resources as well as to the management of the health system in general.

Therefore one of the major concerns of the Ministry of Health is the improvement of efficiency in the allocation and utilization of the limited public health budget. It is clear that to address this concern as well as other concerns of Kenya's health sector reform, there is need to generate and use relevant evidences which are currently lacking. For the case of improvement of the efficiency in the allocation and utilization of the public health budget, needed evidence include the degree of efficiencies that currently exist in the provision of public services as well as factors responsible for the existing level (s) of (in) efficiencies in the provision of these services. It is for this reason that this research is being undertaken. To evaluate whether indeed there is efficiency in the delivering of health care system.

The study concentrates on hospitals because it has identified (Kwak 1992) that hospitals have the largest expenditures compared to other health facilities, and that the quantities of resources used to care for the same number of patients has been found to vary by more than 50 % in hospitals. Kinga et al. (2000) also state that hospitals absorb a disproportionate share of health sector resources.

1.4 Objectives of the Study

The main objective of this study was to determine whether resources allocated to hospitals are used efficiently in the provision of health care in North Rift Region of Kenya.

1.4.1 The Specific Objectives are to:

i. To compare levels of technical efficiency among selected hospitals in North Rift Region.

ii. To identify the major determinants of the variations in technical efficiency in the hospitals.

1.5 Hypothesis

H_1. There are levels of technical efficiency among selected hospitals in North Rift Regions

H_2: There are major determinants of variation in technical efficiency in the hospitals.

1.7 Significance of the Study

A study estimates that in 1997, public health spending for AIDS alone already exceeded 2% of GDP in 7 of 16 African countries, which is very alarming, considering the fact that these nations have total public spending on health which account for 3%-5% of GDP. Kenya is no exception, where resources that are assumed to be needed (mainly due to HIV / AIDS) are given to the health, sector, which is leading to rapid and persistent rising costs in the health care system. As the growth rate of health care accelerates, efficient allocation of health care resources becomes an issue. UNAIDS/WHO, (2005)

Therefore the contribution of this study to North Rift Region and Kenya as a whole is that it will assist the government in realizing whether in deed more resources are needed in the health sector (specifically hospitals) or that the resources being given in the health sector are sufficient and that the problem is the inefficiency in the usage of these resources.

Furthermore the Kenya government is concerned about the efficiency of many public sector activities, hence this study will help the government identify whether indeed there is inefficiency in the delivery of health services and ways to go about improving the efficiency level, since there is no study done in this area hence it will contribute to the body of knowledge.

The study has assisted Kenya Health Sector Reform process where reforms are primarily concerned with improving efficiency, equity, quality, cost effectiveness and consumer satisfaction. To address this issue, relevant evidence that is needed include the type and extent to which health facilities are efficient or inefficient. Therefore this study will identify the efficient or inefficient health facilities (more precisely hospitals), which need to be identified in the reform process.

The study also helped the private sector (private hospital) in knowing the operational aspects of the already existing hospitals. This is in the case where more investors want to venture into Kenya health sectors.

The study has demonstrated the use of the existing method of measuring overall technical efficiency (DEA) especially in developing countries where according to the literature that was reviewed in this study, little research has been done on measuring efficiency of health facilities in Africa. It will also contribute to the existing body of knowledge in Kenya since information available indicates that this will be among the measure of hospital efficiency in Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The review of literature begins with the identification of concepts that will be used in this study; this will be followed by literature on the theoretical framework of study and different efficiency measurement techniques that can be used. Conceptual framework looked at common concepts used in the efficiency studies; this is to ensure proper interpretations of the results. Though efficiency is applied in many fields, such a s health, finance, agriculture and education, the concepts used are similar from author to author and the definitions below were compiled from Coelli (2005), Folland and Hofler (2007), Hartwich and Kyi (2009) and Forsund and Sarafoglou [2010), and are the definitions that are used in this study.

2.2 Application of Efficiency in the Health Sector

The main reason that has been identified by Britain and block (2009) as to why efficiency is applied in the health sector is that improvement in efficiency provides the possibility of producing more, or higher quality services with the same level of resources and thereby expanding access to health care. This is done due to the realization that large shares of health resources are wasted because of poor managerial practice and use of inappropriate

Technologies or human resources. This is more serious in many sub-Saharan African countries where health services are scarce, not accessible, and are of poor quality making better use of resource and an effective means of cost –containment [Chilingerian 2004] identifies ways in which efficiency in the use of resources can be enhanced; first equitable access to health services improves cost efficiency, second aiming for greater efficiency in the use of human resources, third individual and families can be educated to take greater responsibility for their health, fourth careful choice of technology and lastly the strengthening of management support services which is crucial to sustain efficient delivery of services.

Hsiao (2005) stated that efficiency can be achieved in the presence of a good health policy. The fundamental questions that guide the formation of a good health policy are:; on health care: *health care financing:* Second, a country must decide how and by whom scarce health resources are to be allocated among programmes, diseases and regions: *equity, need and cost benefit analysis.* Third, countries must endeavour to obtain maximum efficiency in the production of health services. Therefore efficiency, equity and quality are termed as the major fundamental objectives of any health care system. Hsiao also identifies the nature of ownership, organization and management of health facilities as factors that affect the productive efficiency of health services.

Mooneys' (1986) view of efficiency in health is summarized that *Without a wider use of economics into health care, inefficiencies will abound and decisions will be made less explicitly and hence less rationally than is desirable. We will go on spending large sums to save lives in one way when similar lives in greater numbers could be saved in another way. The price of inefficiency, inexplicitness and irrationality in health care is paid in death and sickness.*

Unlike other services produced, measuring efficiency in producing health services is a challenging exercise because of a variety of factors; quality of care, case mix, input prices and scale of operation vary among providers. All these factors affect the relation between required resources and health service outcome (Bitran and Block 1992) the following provides a brief discussion of the relationships among these factors:

2.2.1 Efficiency and Quality of Care Heterogeneity

There is a need to incorporate the quality of care when measuring technical and allocative efficiency in the production of health services. This is because different levels of quality often consume different levels of production inputs. Therefore, unless quality differences are considered, efficiency measures of a group of providers may obtain a distorted picture. According to Bitran and Block's (2009) example to illustrate this point. Consider two providers D and E of [figure 2.1], each providing the same level of output Q, (outpatient visits) according to their respective production possibilities frontiers.

While both providers operate at the same output level Q, they produce care of different technical quality; provider D is assumed to provide care of greater technical quality, H_1 while provider e is supposed to produce care of a lower technical quality, H_2 Because of these differences in the quality of output, provider D operates at point 1 while provider E operates at point 2.

If researchers compare technical and economic efficiency without including the differences in technical quality, they would conclude that provider E is technically and allocativelly more efficient than D. This is because provider E is seen to use fewer production inputs than D (X_{a2} and X_{B2} versus X_{a1} and X_{b1} respectively), but produces the same level of output Q as provider D. This conclusion would be wrong, because provider D is seen to use more inputs simply due to the fact that the output is of higher quality.

Hence the quality of care has to be taken into consideration and a correct comparison would be one where at any given level of output, relates technical quality to input use.

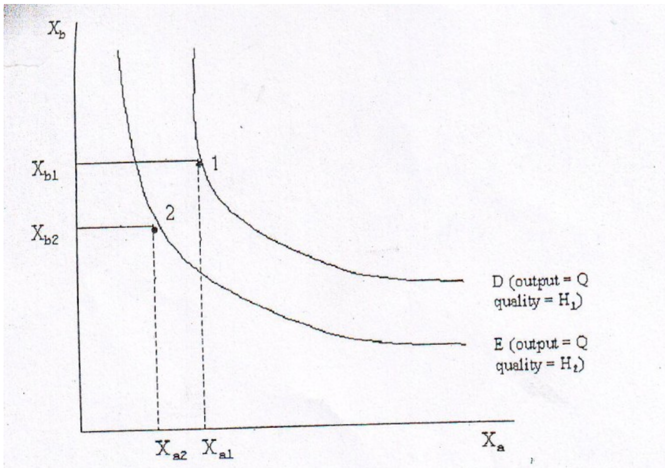


Figure 2.1 Efficiency and Quality of care

Source: Bitran and Block, 1992

Controlling for quality variations in the measurement of efficiency brings about important methodological problems, mainly because of difficulties in measuring quality. Bitran and Block (2009) noted that, even if only a technical or supply-side definition of quality is adopted, data limitations generally preclude an appropriate quality adjustment in studies of efficiency. If technical quality is to be measured on the basis of compliance with technical treatment standards and health outcomes, then a study of efficiency that controls for quality of care would have to obtain information about both the procedures performed and the patients' health status after the care is provided.

2.2.2 Efficiency and Case Mix Differences

Case mix (or the types and complexity of medical conditions of patients, along with related service mix), is a second factor that complicates the measurement of efficiency in the health care sector. This is because case like quality varies among providers. Therefore everything else being constant, one would expect providers with different case mix to use different levels of production inputs.

A facility with a greater proportion of complex cases should be expected to use more resources in production than an otherwise identical facility treating a set of patients with fewer severe cases. Bitra and Block (2009) example used to illustrate this point. Figure 2.2 shows the two providers L and M, with L treating high severity patients (children with severe dehydration from dysentery) and M treating low severity patients (children with mild dehydration from dysentery)

Highly dehydrated children may need to be hospitalized for days and receive intravenous feeding and re-hydration, while children with mild dehydration can be sent home while parents feed them oral ere-hydration salts. Provider L operates at points 3 to treat high severity cases, while M operates at point 4 to teat the milder case. If case severity were not taken into consideration, the researcher would wrongly conclude that provider M, one with lower input use, is more technically and allocatively efficient. Without further analysis one could not make any definitive statements about relative efficiency. This problem is more serious in large hospitals.

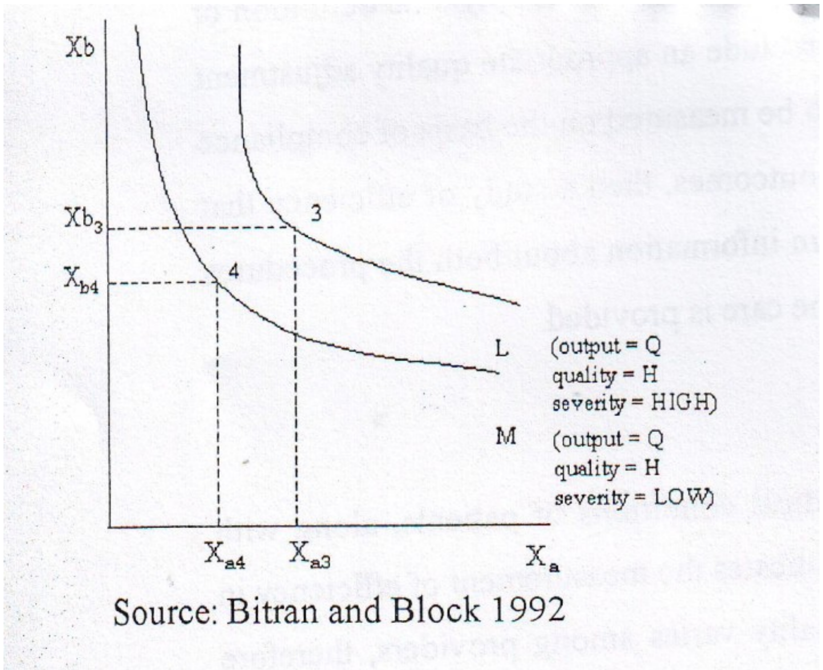


Figure 2.2 Efficiency and Care Mix

Source: Bitran andBlock 2009

Bitra and Block (2009) suggest methods to control for case mix when evaluating heath provider efficiency.Limiting the sample of providers to be compared on the basis of efficiency to those that provide a similar case mix. Here case mix information should be derived on the provider's outputs and not on the inputs, since input usage is an element of productive efficiency.

In econometric studies of efficiency, include explanatory variables such as the types and volumes indicating whether the provider performs other activities that affect resources use, such as the provision of medical education and Limiting the comparison of efficiency to a few, well defined medical services.

2.2.3 Efficiency and Price Information Differences

When there is price information we can measure both technical and allocative efficiency (cost minimization or revenue maximization) Byrnes Valdmanis (1990) note that when you reduce the employment of excess inputs it would increase technical efficiency and select the cost-minimizing mix of inputs, and given relative input prices would lead to allocative efficiency. Hospitals that attain both types of efficiency can lower their costs and thereby compete more effectively with relatively cost-inefficient competitors. Therefore in addition to knowing quality and case mix, the researcher interested in measuring the economic efficiency of producers must also know their input prices. But the prices of producing health resources also vary hence complicating the analysis of efficiency.

2.3 Efficiency Measurement Techniques

The primary purpose of this section is to outline a number of commonly used efficiency measurement techniques. They are divided into frontier (data envelopment analysis and stochastic frontier method) and non-frontier estimations (ratio and regression analysis).These efficiency measures are discussed below:

2.4 Non-Frontier Measures

There are two types of non-frontier estimations, namely ratio and regression analysis.

2.4.1 Ratio Analysis Approach

According to Bitra and Block (2009), measurement of efficiency through Ration Analysis represents a calculation of a ratio relationship between variables. This is done using either output ratio or cost of inputs to output ratio

Input to out put ratios (physical): This approximates technical efficiency.

$$\text{Efficiency} = \frac{\text{Output}}{\text{Input}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad [2.1]$$

This ratio Analysis is conceptually simple to compute; it has low cost of application and can be applied in small samples. The major limitation of Ratio Analysis (1) is that each ratio is limited to only one input and one output and cannot easily accommodate cases of multiple inputs and outputs.

The second: Cost of inputs to outputs ratios: approximates economic efficiency, is measured as:

$$\text{Efficiency} = \frac{\text{Weighted sum outputs}}{\text{Weighted sum of inputs}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad [2.2]$$

Dyson and Frank (2007) identified a difficulty in obtaining a common set of weights in health care, for example, costs (or weights) of each type of patients care (output) are needed, so as to get a weighted output measure, but such weights are not known in most hospital services and if they are known the variations are very high.

Both Ratio Analysis (2.1) and Ratio Analysis (2.2) are not immune to the problem of quality and case mix variations among providers. Unless adequately done ratio analysis may also fail to distinguish between technical and allocative efficiency. Furthermore when hospitals are compared using ratios, there is no objective way of pinpointing inefficient hospitals.

Sear (1991) studied the efficiency and profitability of investor –owned and not-for profit hospitals using 142 hospitals in 1988. He used three measures of efficiency: one, the total number of Full Time Equivalent (F.T.E) personnel per active bed, two, the number of man-hours per adjusted patient days and three, the total wage paid per adjusted patient day. The results showed that investor-owned hospitals used significantly fewer FTE staff per bed, had significantly fewer man-hours per adjusted patient day, and paid significantly less in wages (this last finding- difference in the price of an input-does not say anything about relative economic efficiency).

Lee (1984) carried out a comparative study of efficiency between freestanding ambulatory care and hospital –based care using data from three hospitals and three neighborhoods health centres. Using a simple method for allocating fixed and indirect costs, the authors computed the average cost per visit in hospitals and health centres. The results showed that a wide variety in costs and no clear pattern emerged between the two types of facilities. Using other methods for apportioning various costs to visits, the authors demonstrated how sensitive the results were to the cost allocation rule employed.

2.4.2 Regression Analysis Approach

In the past, regression approaches have been commonly used for measuring efficiency, because it is more comprehensive than ratio analysis and it can accommodate multiple inputs and outputs. The other advantage of regression analysis is that statistical significance can be determined and attached to the regression coefficients, and also there is consideration of random error term in a regression model and finally the use of regression may be advantageous simply because the approach is more familiar and thus better understood or accepted.

The limitations of regression analysis; include the inability to identify sources and estimate the inefficiency amounts associated with these sources. There is no clue on the corrective action is provided even when the dependent variable shows that inefficiencies are present. The second limitation is that an estimate of hospital cost function using the technique results in a mean relationship that does not directly locate inefficient hospitals, hence as with Ratio Analysis, it requires that hospitals with arbitrary distance from the mean be labeled potentially inefficient (Bowlin, 1999).

Furthermore, numerous econometric regression types of hospital studies have been used to identify economies of scale, marginal cost of patient care and rates of substitution among inputs and outputs, but such results say nothing about the efficient rate of substitution, efficient scale size or efficiency rate of transformation simply because they reflect behaviour of both efficient and inefficient hospitals combined.

Shearman (2001) notes that the use of regression analysis would only provide sights into efficient hospital behavior of the hospitals in the study were known to be efficient. Despite these limitations, econometric approaches are among the widely accessible and have consequently been widely used in the hospital industry.

Frank and Taube (1987) studied technical and allocative efficiency of 766 mental clinics. They looked at providers' departures from cost minimizing behaviour and hence compared the ratio between factor prices with the ratio between the factors marginal products. They noted that cost minimization is present when those two ratios are the same for all pairs of variables inputs. Difference in price-marginal product ratios means that certain production inputs are either over or underemployed. They were also interested in the system efficiency and thus studied the extent of economies or diseconomies of scale in production function of mental clinics. The authors use two different empirical specifications for the production of functions of the clinic: a Cob-Douglas and a transcendental production function.

The Cobb-Douglas production function is specified as

$$InQ = InA + \alpha \cdot InX_2 \text{...} [2.3]$$

The transcendental Production function is specified as

$$InQ = A + \alpha \cdot InX_1 + \beta_1 \cdot X_2 + \beta_2 \cdot X^2{}_2 \text{................................} [2.4]$$

Estimation was done using Ordinary Least Squares. They found decreasing returns to scale with both specifications and a greater productivity of private clinics as evidenced by an ownership dummy included in the production functions. Finally they compared input price ratios with marginal ratios and found that they differ for the case of physician and other clinical staff, physicians being over-employed. This signals a departure from cost minimizing behaviour. With regard to their empirical specifications, they find that the simpler Cobb-Douglas model performs better than the alternative specification, as measured by goodness of fit criterion.

Vitaliano (1987) argued that econometric studies of efficiency, which use a multiple output approach, are plagued by the statistical problem of multi-collinearity among outputs, which hinders the interpretation of regression coefficients. He went to further point out that hospital cost behaviour should be studied in the context of a system of equations, where price, cost and output are determine jointly, single equation is a reduced form of a system, a fact that hampers the interpretation of statistical coefficients. In his study he used 166 New York hospitals and argued that because prices are exogenous to hospital and "output is non-storable and supplied on demand" estimation of a single output to conduct a comparative study of economic eficiency[6]. He obtained a quadratic functional from below, and estimation was done during weighted least squires:

$$C=a+b.BEDS+c \cdot BEDS^2+d.MEDSCHOOL+e \cdot FACILITIES+f \cdot URBAN+g \cdot SHARE+e \ \ldots\ldots\ [2.6]$$

With the quadratic cost function, he obtains a U-shaped average cost function. However when he uses a different specification for the cost function (log form) he finds a declining average cost curve and thus economies of scale. These results are not entirely surprising since a quadratic function is likely to produce a U-shaped average cost curve, whereas a logarithmic specification is not.

He attributes the economics of scale to the presence of high fixed costs such as specialized personnel and equipment.

Evans et al (2010) measured efficiency using the production function approach using panel data pf 191 member countries[7] of WHO in 1993-1997. They described the two input case (X_1, X_2) translog model for fixed –effect panel data estimation as follows (all variables in logs)

$$Y_{it} = \alpha_i + \beta_1 X_{1it} \beta_2 X_{2it} + \beta_3 (X_{1it})^2 + \beta_4 (X_{2it})^2 + \beta_5 (X1_{it}) (X_{2it}) + V_{it} \ldots\ldots\ldots\ldots [2.7]$$

And state that both the Cobb-Douglas and the Constant Elasticity of Substitution (CES) production functions can be derived as restricted formulations of the translog functional form. To measure the efficiency, three general types of variables were used. DALE (Disability Adjusted Life Expectancies), expenditure and average educational attainment in the adult population.

They found that many of the poorer performers are those in which there has been significant civil unrest over the period. Many of the others are countries with high prevalence of HIV / AIDS, where the prevalence of AIDS can reduce DALE by almost 15 years in some of the most highly endemic countries such as Botswana (which was ranked 188 out of 191). The authors debated whether they should account for the presence of AIDS in the assessment of efficiency, but decided not to on the grounds that the health system should be held at least partly accountable for the fact that AIDS has not been controlled.

2.5 Frontier Measures

The two principal methods under frontier analysis are Data Envelopment Analysis (DEA) and Stochastic Frontier Method (SFM)

2.5.1 Stochastic Frontier Approach (SFA)

In this approach efficiency measures are calculated relative to an efficient technology, which is generally represented by a frontier function. Hofler and Folland (1995) note that in this method each organization is treated uniquely by assuming it to be affected by a potential shock to its ability to produce, hence since each firm can be potentially shocked, the firms best possible practice, its frontier will be randomly shifted.

The frontier function is therefore partially random, that is the frontier is a stochastic process. Stochastic frontier. The SFA adopts a parametric function (for example Cobb-Douglas form) fitted to the data estimated using econometric approaches and makes assumptions in advance about the statistical distribution of the inefficiencies.

Linna and Hakkinen (1996) evaluated the determinants of cost efficiency in 48 acute care hospitals in Finland using cross-sectional data. The five university hospitals were disaggregated into main specialties operating as managerially independent units. Thus the total number of observations was 95.parametric methods were used to obtain overall cost efficiency; the models are specified by a stochastic frontier cost function. Cost efficiency was estimated with short run multi-product cost functions since major capital investments were excluded. Box-Cox transformed frontier cost function (SFMODEL) was used because it was believed that it would best describe the costs of hospitals.

$$\text{In} \; \frac{Ci}{Wd} = \alpha + \sum_{j=1}^{M} \beta_j \, y_{ij}^{(\lambda)} + \delta 1n \; \frac{Woi}{Wdi} + u_i + V_i \qquad [2.8]$$

Where: C is total costs, wd is input prices per doctor, who is input prices for other staff, The Box-Cox transformation is $Y^{(\lambda)} = (Y^{(\lambda)} - 1) / \lambda$

The results indicated that the hospitals are able to produce both teaching and research output at decreasing marginal costs.

2.5.2 Data Envelopment Analysis (DEA) Approach

DEA is the optimization method of linear programming to generalize the Farrell's 1957 single input / single output multiple –in-put / multiple –output case, by constructing a relative efficiency score as the ratio of a single virtual output to a single virtual input. It was then developed by Charnes, Cooper and Rhodes in 2001 with CRS and was extended by Banker, Charnes and Cooper in 2002 to include Variable Returns to Scale (VRS)

DEA simultaneously analysis's the efficiency with which each DMU (Decision Making Units) uses input to produce its outputs. It identifies the optimal input/ output combination and represents it with "best practice frontier", or data envelop. DMUs that compose this frontier are assigned a score of one and are technical efficient relative to their peers (Coelli 2005)

DEA addresses the limitations associated with the other three methods (RA, SFM and Econometric analysis) of measuring efficiency, as outlined by Shearman (1984), Hartwich and Kyi (2009), Bowlin (2008) and Bitran and Block (2009).

DEA enables simultaneous analysis of multiple inputs and multiple outputs (in their natural physical units) Before in the aim to come to overall performance indicators, practitioners had to apply weights (as are needed for ratio analysis and most types of regression analysis) derived from different rather subjective scoring methods and priority setting exercises. Often the thus derived weights were criticized. The concept of Pareto efficiency used by DEA is an appropriate quantitative tool to avoid the weighting problem.

DEA simultaneously considers the multiple outputs and inputs of an organization without the need to know the efficient relative weights DEA conservatively measures the existing inefficiency and the amount of input reductions that would make inefficient units as efficient as other units in the observation set.

DEA unlike econometric techniques permits the study of production efficiency without the need to make any assumptions about the technology of production.

DEA further avoids the assumption that all units of analysis produce under the same conditions. As services offered by different units are highly individual, units operate under very different conditions. It is likely that only some units may have found a particular best practice way of providing services. This technological know how is not attained by the other units. Therefore the assumption of frontier of best practicing units is more valid than the assumption that all units use the same technology.

DEA locates technical or Pareto inefficiencies in a manner more consistent with economic theory than econometric regression techniques, which is DEA measures efficiency compared with the best practice facilities rather than based on a mean or central tendency relationship that reflects a mixture of efficient and inefficient behaviour.
DEA is also unambiguous in its allocation of inefficient units, that is, hospitals located to be inefficient are strictly inefficient. Furthermore DEA indicates the general magnitude of inefficiencies present in the health facilities.

DEA provides a means of ‘decomposing” economic inefficiency into technical and allocative inefficiency. Furthermore, it allows technical inefficiency to be decomposed into scale effects and pure technical efficiency

Furthermore to calculate technical efficiency, only requires information on output and input quantities (not prices). This makes it particularly suitable for analyzing the efficiency of government services providers, especially those providing human services where it is difficult or impossible to assign prices to many of the outputs.

Finally, contrary to econometric techniques, DEA is a deterministic technique and as such does not include explicitly a statistical error term reflecting measurement.

But as in the other efficiency measurement techniques DEA also has some limitations Bowlin (2009) Hartwich and Kyi (2009) and Shearman (1984)

Being a deterministic rather than statistical technique, DEA produces results that are particularly sensitive to measurement error. If one organization’s inputs are understated or its outputs overstated, then that organization can become an outlier that is significantly distorts the shape of the frontier and reduces the efficiently scores of nearby organizations. In regression-based studies, the presence of error-terms in the estimation tends to discount the impact of outliers, but in DEA they are given equal weight to that of all other organizations. It is therefore, important to screen the outliers when assembling the data. One check is to scrutinize those organizations whose output to input ratios laid more than two and half standard deviations from the sample mean. DEA only measures efficiency relative to best practice within the particular sample. Thus it is not meaningful to compare the scores between two different studies because differences in best practice between the samples are unknown. Similarly, DEA study that only includes observations from within n the state or nation cannot tell us how those observations compare with national or international best practice.

DEA does not locate the actual techniques that give rise to the located inefficiencies or the optimal path to improve efficiency. It directs management’s attention to areas where inefficiencies exists and allows management to identify the preferred path to improve productivity using more problem-focused analytic techniques such as those found in operations research and industrial engineering.

DEA does not identify the efficient production function. Consequently, it does not replace the need to continue efforts to estimate the efficient production function using techniques including external econometric regression methodologies such as those found in operations research and industrial engineering.

DEA does not identify the efficient production function. Consequently, it does not replace the need to continue efforts to estimate the efficient production function using techniques including external econometric regression methodologies such as those being developed.

2.6 Synthesis of the General Literature Review

The theoretical study will outline different measures of efficiency and will prove that DEA is a reliable and effective method of estimating the efficiency (or lack of efficiency) in any institutions, unlike the other three measures.

Bowlin et al (1985) developed a hypothetical data set for hospital units with known efficiencies and inefficiencies. They used this data set to test DEA against ratio and regression analysis. They too found that DEA outperformed both ratio analysis and least squires' regression in identifying sources and amounts of inefficiencies. It is therefore for this reason that this study will adopt

2.7 DEA Literature Review

2.7.1 Graphical Approach to DEA

The discussion begins with an input / input space and hence an input reducing focus, which is equally, termed as input-oriented measures. This will be followed by the output oriented measure.

2.7.2 Input-Oriented Measures

This will be illustrated using the example of Coelli (1996), who borrowed the original works of Farrell. Figure 3.1 shows a firm using two inputs (X_a and X_b) to produce a single output (Q) under the assumption of constant returns to scale and knowledge of the unit isoquant of the fully efficient firm.

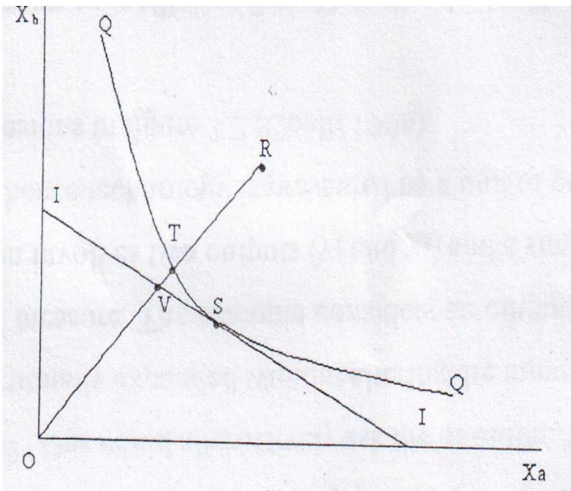


Figure 2:3:Technical and Allocative Efficiencies

Source : Coelli 2005

All points along the isoquant such as S and T are technically efficient, while points such as R are technically inefficient. This technically inefficiency of firm R is represented by the distance TR (which is the amount by which all inputs could be proportionally reduced without a reduction in output), in percentage terms the inefficiency of R is represented by the ratio TR/OR representing the percentage by which all inputs could be reduced.

Therefore a measure of Technical Efficiency (TE) is the ratio:

$$TE_1 = OT / OR \text{ or } TE_1 = 1 - TR / OR \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2.9]$$

Allocative efficiency extends the concept of technical efficiency to take into account the relative prices of production inputs. Therefore when input prices are known (represented by line 11) allocative efficiency can be calculated. The Allocative Efficiency[9] (AE) of firm R is defined as the ratio:

AE_1 = OV / OT or AE1 = 1- VT /OT……………………………………….[2.91]

Note that at point T there is technical efficiency but allocative inefficiency, while at point V there is allocative efficiency but technical inefficiency. Only point S (at the intersection of the isoquant and isocost lines) is both allocatively and technically efficient.

While many technical efficient alternatives might occur to produce a given level of output, (any point along the isoquant line QQ) and many allocative efficiency points can occur (any point along isocost line 11), there can only be one economic efficient point which occurs at intersection between the isocost curve and isoquant, that is point s.

Coelli (2005) thus states that a provider's deviation from minimum cost (or from maximum economic efficiency) can be attributed to or decomposed to technical and allocative inefficiency. Therefore overall Economic Efficiency[10] (EE) equals technical efficiency (TE) multiplied by allocative efficiency (AE): EE = TE . AE

Hence from the diagram overall efficiency of production at point R would be:

EE_1 = OV / OR = OT / OR. OV/OT………………………………[2.9.2]

2.7.3 Output-Oriented Measure

The above input-oriented efficiency measures address the question: "by how much can input quantities be proportionally reduced without changing the output quantities produced? One could alternatively ask the question: "by how much can output quantities be proportionally expanded without altering the input quantities used?" This is an output –oriented measure. The example considers an output –oriented measures in a case where production involves two outputs (y_1 and y_2) and a single input (x_1). Under he assumption of CRTS hence technology is presented by a unit of production possibly curve (QQ) in two dimension in figure 3.2 (Coelli 1996)

Point D shows an inefficient firm, Note that the inefficient point D lies below the curve in this case, because QQ' represents the upper bound of production possibilities. Therefore Farrell output-oriented efficiency measures in figure 3.2 is defined as follows: distance DB represents technical inefficiency. That is the amount by which outputs could be increased without requiring extra inputs. Hence a measure of output-oriented technical efficiency is the ratio: TE_o = OD/OB

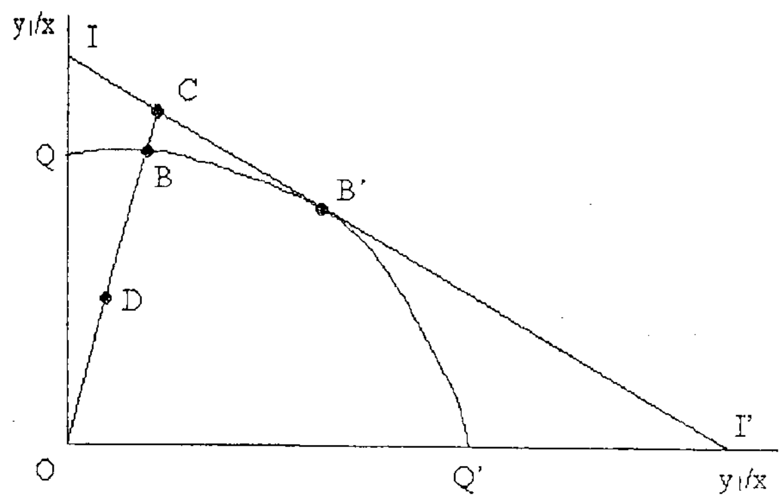


Figure 2.4: Technical and Allocative Efficiencies from an Output-Orientation Source: Coelli, 2005

When we include price information (isorevenue line 11' is included), allocative efficiency is defined as: AE_o = OB / OC. Which has a revenue increasing interpretation (similar to the cost reducing interpretation of allocative inefficiency in the input –oriented case). Hence overall economic efficiency (EE) is the product of TE and EA (just as in the input-orientation) to give: EE_o = (OD / OC) = (OD / OB) * (OB / OC) = TEo * AEo

2.7.4 Difference between Output-and Input –Oriented Measure

The difference between output and input oriented measure is illustrated in figure 3.3 where we have a decreasing returns to scale technology represented by figure 3.3a and a constant returns to scale technology represented by figure 3.3b, still follows Coelli,s work.

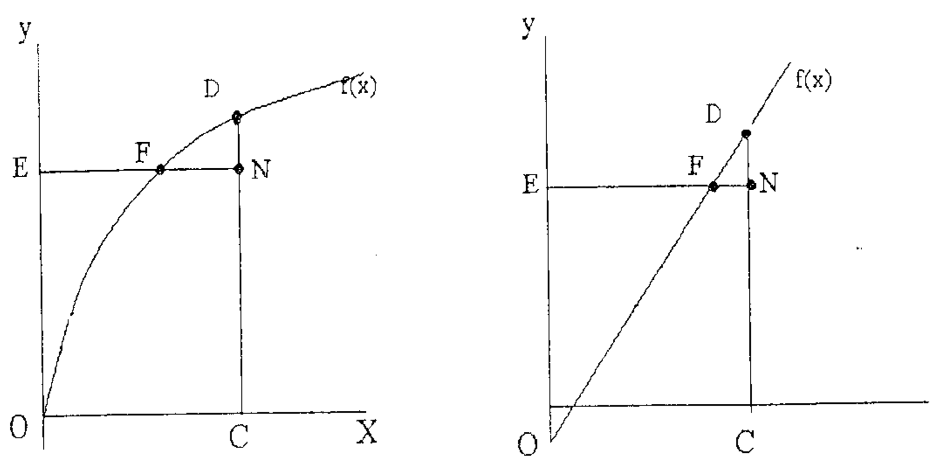


Figure 2.5: Differences in Input-and Output –Oriented TE and Retutns to scale

Source: Coelli 2005

The input-oriented measure of TE is the ratio EF/EN, while the output oriented measure of TE is CN/CD. The output-and input-oriented measures will only provide equivalent measures of technical efficiency when constant returns to scale exist, but will be unequal when increasing or decreasing returns to scale are present. The constant returns to scale case is depicted in Figure (3.3b) where we observed that EF/EN = CN / CD, for an inefficient point N.

2.8 Mathematical Approach to DEA

This section will first look at how the efficiency ratios are convered to the linear programming model DEA (following the original works of Farell), this followed by the original assumptions of DEA, that is CONSTANT Returns to Scale (CRS) DEA and Variable Returns to Scale (VRS) DEA.

2.8. 1 Introduction to DEA

First a formula for relative efficiency incorporating multiple inputs and outputs is introduced and its shortcoming, then the DEA model follows which allows relative efficiency measures, that is DEA measures the relative performance of organizational units where the presence of multiple inputs and outputs makes comparisons difficult. The usual measure for relative efficiency is given by:

$$\text{Efficiency} = \frac{\text{Weighted sum of output}}{\text{Weighted sum of input}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{Equation 2.1}$$

After introducing notations it can be written as:

$$\text{Efficiency of unit j} = \frac{u_1 y_{ij} + u_2 y_{2j} + \ldots\ldots\ldots\ldots}{v_1\, x\text{-}\, x_{1j} + v_2 x_{2j} + \ldots\ldots\ldots\ldots} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{Equation 2.2}$$

Where: u_1: the weight given to output I, y_{1j}: amount of output 1 from unit j, v1: the weight given to input I, x_{1j}: amount of input 1 to unit j

The initial assumption was that this type of efficiency measure requires a common set of weights to be applied across all units, but there are two difficulties in obtaining this common set of weights, (Dyson and Frank 2001)

It may not be simple to value the inputs or outputs, and Different firms may choose to organize their operations differently so that the relative values of the different outputs may legitimately be different.

The measurement of relative efficiency where there are multiple possibilities of inputs and outputs was addressed by Farrell, who constructed a hypothetical efficient unit as a weighted average of efficient units to act as a comparator for an efficient unit. Charnes, Cooper and Rhodes followed Farell's work and proposed that each unit be allowed to adopt a set of weights which shows it in the most favourable light in comparison to the other units. Therefore under these conditions efficiency of target unit jo can be obtained as a solution to the following problem, (Dyson and Frank 2007)

Maximize the efficiency of unit jo,

Subject to the efficiency of all units being $\leq$

The variables of the above problem are the weights, and the solution produces the weights most favourable to unit Jo while measuring the efficiency level. The algebraic model is as follows: (Dyson and Frank 2007)

$$\text{Max } ho = \frac{\sum u_{jo}\, y_{rj}}{\sum u_i\, y_{ij}}$$

$$\text{Subject to } \frac{\sum u_r\, y_{rj}}{\sum u_j\, y_{ij}} \leq 1 \text{ for each unit } j$$

$$u_r, V_i \geq €$$

…………………………… Equation 2.3

DEA model in equation 3.3 is a fractional linear programme. To solve the model it is first necessary to convert it into form so that linear programming can be applied. The linearization of equation 3.3 is shown in equation 3.4 In maximizing a fraction or ratio it is the relative magnitude of the numerator and denominator that are of interest and not their individual values. Therefore it is possible to achieve the same effect by setting the denominator equal to a constant and maximizing the numerator. The result is as follows:

Max $ho = \sum u_r y_{rjo}$

Subject to : $\sum v_i x_{ijo} = 1$

$\sum u_r y_{rj} - \sum v_i x_{ij} \leq 1,$ $j = 1,\ldots\ldots\ldots..n,$

$U_r, v_i \geq €$ (equation 2.4)

Just like in any linear programming it is possible to formulate a partner (dual DEA model) to the original DEA problem (primal DEA model). The dual model is constructed by assigning variables, (dual variables) to each constraint in the primal model and constructing a new model on these variables. This is shown in equation 3.5 and 3.6:

Primal Model

Max $ho = \sum u_r y_{rjo}$ Dual variables

Subject to: $\sum v_i x_{ijo} = 1$ Z

$\sum u_r y_{rj} - \sum v_i x_{ij} \leq 1,$ $j= 1,\ldots\ldots.n,$ y_0

$-U_r \leq - €,$ $r = 1\ldots\ldots\ldots t,$ $s\bar{\imath}$

$-vi < - €,$ $I = 1,\ldots\ldots\ldots m$ $s^{+}{}_{1}$ (Equation 2.5)

Dual model

Min Zo - $€ \sum s^{+}{}_{r} - € \sum s^{-}{}_{i}$

Subject to : $xij\ Zo - s^{-}i - \sum x_{ij}\lambda j = 0$ $I = 1,\ldots\ldots..,m,$

$s^{+}r + + \sum yrj^{\lambda j} = y_{rj}$ $\quad$ r = 1,………,t,

λ_j, $s^{+}r$, $s^{-}I \geq 0$, Zo $\quad$ Unconstrained. $\quad$ (equation 2.6)

It is usually preferable to solve the dual DEA model (equation 2.6) rather than the primal model (equation 2.5). This is because the primal model has more constraints (hence more difficult to solve) than the dual model.

The dual variables Lamba (j) (from theory of linear programming) are shown prices related to the constraints limiting the efficiency of each unit to be greater than 1. When a constraint is binding, a shadow price will be positive normally and when the constraint is non-binding the shadow price will be zero. In the solution to the primal model therefore a binding constraint means that the corresponding unit has an efficiency of 1 and there will be a positive shadow price (dual variable). Hence positive shadow prices in the primal, or positive values for the Lambdas (j)'s in the dual, correspond to and identify the peer group for any inefficient unit (Dyson and Frank 2007)

2.8.2 The Constant Returns to Scale Model (CRS)

Following the 1957 works of Farrell, Charnes, Cooper and Rhodes in 1978 proposed a DEA model, which had an input orientation and assumed CRS. Subsequent papers such as Banker, Charnes and Cooper in 1984 then considered alternative sets of assumptions, and proposed VRS DEA model. The following begins with input-oriented CRS model, because this model was the first to be widely applied.

This will be explained following the literature of Coelli (2005). Assume that there id data on K inputs and M outputs on each of the N firms (DMUs). For the i^{th} DMU these are represented by the vectors x_i and y_i respectively. The K*N input matrix, x- and the M*N output matrix, Y, represent the data of all N DMUs. DEA constructs a non-parametric envelopment frontier over the data points such that all observations lie on or below the production frontier. For each DMU we obtain a measure of the ratio of all outputs over all inputs, such as $u' y_i / v' x_i$ where u is an M*1 vector of output weights and v is a K * 1 vector of inputs weights. To select optimal weights Coelli specifies the mathematical programming problem.

$\text{Max}_{u,v} (u'y_i / v'x_i)$

st $\quad u'y_j / v' x_j \leq 1$, j =1,2,…….N,

$\quad u,v \geq o,$ ………..…………… Equation 2.7

The values for u and v are found, such that the efficiency measure of the i-th DMU is maximized, subject to the constraint that all efficiency measures must be less than or equal to one. A problem with this ratio formulation is that all it has an infinite number of solutions. To avoid this Collie imposes the constraint $v'x_i=1$, which provides:

$\text{Max}_{u,\, v}(u,y_i)$,

st $\quad v'x_i = 1$

$u'y_j – v'x_j \leq, 0$,j = 1,2,….,N,

$u,v \geq 0$,

Where the notation change from u to v ﻪύ reflects the transformation. This is known as the multiplier form of the linear programming problem. Using a daily in linear programming, an equivalent envelopment form of this problem is derived:

$\min_{\theta,\lambda} \theta$

St $-yi + y\,\lambda > 0$,

$\theta_{Xi} - X\lambda \geq 0$,

$\lambda > 0$,

Where θ is a scalar and λ is a N. 1 vector of constants. This envelopment form involves fewer constrains than the multiplier form (K+ M < N+1) and hence is generally the preferred form to solve. The value of θ obtained will be the efficiency score for the i^{th} DMU. It will satisfy θ < 1, with a value of 1 indicating a point on the frontier and hence a technically efficient DMU, according to Farrell's 1957 definition. The linear programming problem is solved N times once for each DMU in the simple and a value of θ is obtained for each DMU.

2.8.3 Slacks

The piecewise – linear form of the-parametric in DEA causes a few difficulties in efficiency measure. These difficulties are caused by the piecewise linear frontier which run parallel to the axes. This problem is shown in figure 3.4, where the DMUs using input combinations C and D are efficient and define the frontier DMUs A and B are inefficient. The measure of TE gives the efficiency of DMUs A and B as OA'/OA and OB'/OB respectively. However, it is questionable as to whether the point A' is the same output. This is known as *input slack*, when we have a case involving more inputs and/or multiple outputs, the possibility of the related concept of output slack also occurs.

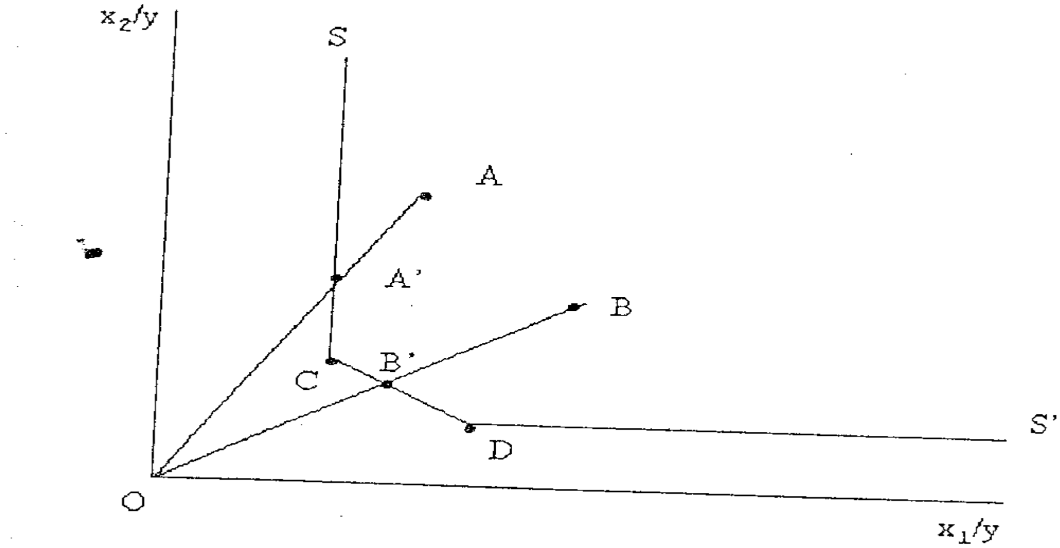


Figure 2.6 Efficiency Measurements and Input Slacks

Source: Coelli 2005

Thus Coelli argues that both the technical efficiency (θ) and any non-zero input or output slacks should be reported so as to provide an accurate indication of technical efficiency of a DMU in a DEA analysis. For the i^{th} DMU the output slacks will be equal to zero only if $Y_{\lambda - yi} = 0$, while the input slacks will be equal to zero only if Θ xi - $X_{\lambda} = 0$ (for the given optimal values of Θ and λ).

In figure 3.4 the input slacks associated with point A' is CA' of input X_2. In cases where

There are more inputs and outputs than considered in this example, the identification of the nearest efficient frontier point and hence the subsequent calculation of slacks becomes a difficult task. The solution suggested for this, is a second-stage linear programming problem, so as to move to an efficient frontier point by *maximizing* the sum of slacks required to move from an inefficient frontier point (A' in fig 3.4) to an efficient frontier (C). This second stage linear problem is defined by 11 (Coelli 2005)

$\text{Min}_{\lambda,\, os,\, 1s^-}$ (MI'OS + K1'IS),

st $\quad -yi + y\,\lambda\text{-OS} = 0,$

$\theta xi - X\lambda - OS = 0$

$\lambda \geq 0, OS \geq 0, IS \geq O$ ………………… Equation 2.10)

However, there are two problems associated with the second –stage LP: first, the sum of slacks is maximized rather than *minimized*. Hence, identifies the furthest and not the nearest efficient point. Second, it is not invariant to units of measurements, which are in the alteration of units of measurement; say for an input from kilograms to tones (while leaving other units of measurement unchanged) could result in the identification of different efficient boundary points and hence different slack and lambda measures.

The two problems are not revealed in figure 2.4 because there is only efficient point on the vertical facet, but if slacks occur in more dimensions then the above problems can occur. Because of these problems many choose the first-stage linear program (equation 2.9) to obtain the technical efficiency measures (θ) for each DMU and ignore the slacks completely. The alternative approach is to report both the technical efficiency score (θ) and the residual slacks, were the slacks are calculated as follows:

$OS = -yi + Y\lambda.$ and $IS = (\theta)xi - X\lambda.$

However this approach also has problems; it is the radical slacks may not always provide all slacks, (for example when a number of observations appear on the vertical section of the frontier in figure 2.4 and therefore may not always identify the nearest efficient point for each DMU. There are three different ways of treating slacks in Coelli's 2005 DEAP software:

1. One-stage DEA, where we conduct the LP in equation 12 and calculate the slacks.
2. Two-stage DEA, where we conduct the LPs in equation 12 and 13.
3. Multiple-stage DEA, where we conduct a sequence of radial LPs to identify the efficient projected point.

Though multi-stage DEA is more computationally demanding it is more recommended than the other two, because it identifies the projected points which have input and output mix and which are as similar as possible to those of the inefficient points, it is also invariant to units of measurement.

2.8.4 The Variables Returns to Scale Model

Certain conditions such as imperfect competition, constraints on finance, may cause a DMU not to operating at optimal scale (the flat portion of the LRA curve). Hence it would not be appropriate to use CRS; this is because usage of CRS when not all DMUs are operating at optimal scale will result in measures of TE, which are confounded by SE. Banker, charnes and Cooper in 1984 suggested an extension of the CRS DEA model to account for VRS situations. Where the use of the VRS specification will permit the calculation of TE devoid of these SE effects.

Still following Coelli's [2005] work, the CRS linear programming problem can be modified to account for VRS by adding the convexity constraint: N1,$_{\lambda = 1}$ to equation 2.9 to provide:

$$\text{Min}_{\theta,\lambda}\theta,$$

$$\text{St} \quad -y_i + Y\lambda \geq 0,$$

$$\theta x_i - X\lambda \geq 0,$$

$$N1'\lambda = 1$$

$$\lambda \geq 0, \quad \text{………………… Equation 2.11}$$

Where N1 is an N*1 vector of ones, this forms a convex hull of intersecting planes

Which envelope the data points more tightly than the CRS conical hull and thus provides

TE scores which are greater than or equal to those obtained using the CRS model.

2.8.5 Scale Efficiencies:

By conducting both a CRS and a VRS on the same data, the TE scores can be decomposed into two components, one due to *scale inefficiency* and one due to “pure” *technical inefficiency.* If there is a difference between the CRS and VRS TE scores for a particular DMU it indicates that there is scale inefficiency. This is illustrated in Figure 3.5, where we have a one –input –oriented technical inefficiency of the point P is the distance PPc, while under VRS the technical inefficiency would only be PPv.

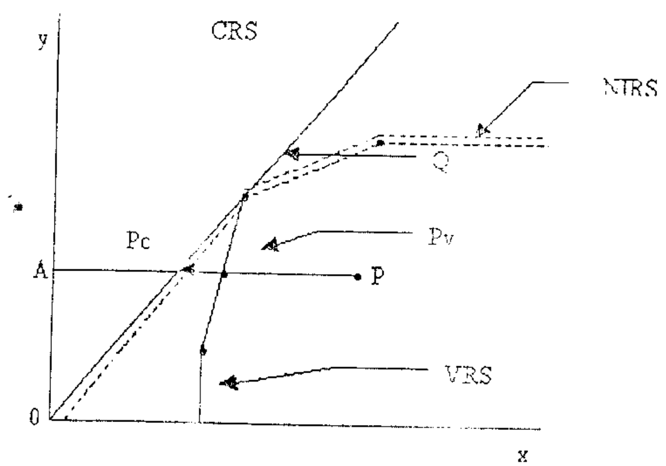


Figure 2.7: Calculation of Scale Economics in DEA

Source: Coelli, 2005

The difference between PcPv, is put down to scale inefficiency. In ratio efficiency measures his is expressed [12] as:

TE1,crs = AP

A limitation of scale efficiency measure is that the value does not indicate whether the DMU is operating in an area of increasing or decreasing returns to scale. But this can be overcome by solving an additional DEA problem with non-increasing returns to scale (NIRS) imposed, by altering the DEA model in equation 2.11 and substituting the N1'λ ≤ 1 to provide:

$\min_{\theta,\lambda}{}^{\theta},$

$$^{St}\ -y_i + Y\lambda \geq 0,$$
$$\theta x_i - X\lambda \geq 0,$$
$$N1'\lambda \leq 1$$
$$\lambda \geq 0, \qquad \text{(equation 2.12)}$$

Coeli Notes that from figure 2.5, the sale inefficiencies (due to increasing or decreasing returns to scale) for a particular DMU is determine by seeing whether the NIRS TE score is equal to the VRS TE score. If they are unequal (point P) increasing returns to scale exist. If they are equal (as point Q) decreasing returns to scale exists.

2.8.6 CRS AND VRS: Input and Output Orientation

The technical inefficiency is identified by a proportional reduction in input usage: input-based measure or it can be measured as a proportional l increase in output production. See section 2.1 figures 2.3 (a) and figure 2.3 (b), the two measures: *output-and input-oriented will estimate the same frontier and therefore by definition identify the same set of DMUs as being efficient. It is only the efficiency measures associated with the inefficient DMUs that may differ between the two methods:* Furthermore given that linear programming cannot suffer from such statistically problems as simultaneous equation bias, the choice of an appropriate orientation is not a s crucial as it is in the econometric estimation case. Many studies opt for an input-oriented model because they have particular orders to fill, and hence the input quantities become the primary decision variables. While other studies opt for an output orientation approaches because some industries are given a fixed quantity of resources and asked to produce as much output as possible.

Therefore Coeli notes that one should select an orientation according to the type of quantities that is most controllable, furthermore the choice orientation will only have minor influences on the scores obtained. An output –oriented DEA with two –output is represented in figures 3.6; Coeli shows that the observations lie below the curve, and that the sections of the curve that at right angles to the axes will cause output slack to be calculated when a production point is projected onto those parts of the curve by a radial expansion in outputs.

2.9 Empirical Framework

The purpose of this section is to provide a computational discussion about issues in the measurement of technical and allocative efficiency using DEA in the production of health services. Therefore reviews of selected studies of health services efficiency that have been conducted in both developed and developing countries are presented here, and measurements issues and measurements technique described.

In the developed countries criteria, Finland (where a study by Linna and Hakkinen is looked at) and United States of America (where two tudies are looked at; one by Brynes and Valdamanis, and the other by Shearman) are presented. While in the Developing countries criteria, Tawan (where a study by Chang is looked at) Korea (where a study by Wong is looked at) Kenya (where a study by Kiriga, Emrouznejad and Sambo is looked at) and South Africa are presented.

2.9.1 Review of Efficiency in Hospitals Studies in Developed Countries

Linna and Hakkinen (1996) in their study of determinants of costs efficiency of Finnish Hospitals used cross- sectional data from 1994 of 48 acute care hospitals, and applied both DEA (data Envelopment Analysis) and SFA (Stochastic Frontier Approach).

Cost efficiency was estimated with short-run multi-product cost functions: Box –Cox transformed frontier cost function (SFMODEL)

$$\ln \frac{C_i}{W_d} = \alpha + \sum_{j=1}^{M} \beta_j \, y_{ij}^{(\lambda)} + \delta 1n \frac{W_{oi}}{W_{di}} + u_i + V_i \qquad \text{………………….. Equation 2.13}$$

Where: C is total costs; W_d is input prices per doctor; W_o is input prices for other staff: The Box-Cox transformation is $Y^{(\lambda)} = (Y^{(\lambda)} - 1) / \lambda$.

While the cost efficiency model (DEACEI) was specified as:

$$\begin{aligned} \text{Min zx} \quad & \sum w_{jo} x_j \\ \text{s.t} \quad & z.Y \geq y_o \\ & zi \geq 0 \\ & \sum z_{i=1} \qquad \text{………………….. Equation 2.14} \end{aligned}$$

Where: Y is an n-m matrix of observed outputs for n hospitals; X- is an $n_* k$ matrix of inputs for each hospital: Z is a 1_{*n} vector of intensity variables; w = (w1, ………………wk)€R^k+ denotes input prices;

The constraint of equation 2.14 defines the input requirement set given by

$$L(y) = \{x\text{-}:z \cdot Y \geq y_0, z \cdot X, \leq x, z_i \geq 0, \sum^{n} z_i = 1\}$$

The findings indicated that the choice of modeling approach did not substantially affect the results, though using DEA it was possible to decompose overall cost efficiency into allocative and technical components. Level of cost inefficiency was estimated to lie between 8-15 %, which suggested that improving the overall efficiency of hospitals could reduce t he hospital costs by FIM 1.1-1.6 billion ($300-400 million). Approximately half of the observed inefficiency was due to technical and half to allocative inefficiency. Scale inefficiency was found to be a minor factor in overall inefficiency.

Byrnes and Valdmanis (1983) conducted a research in the United States of Analyzing Technical and Allocative efficiency for a sample of 123 communities (non –teaching) not-for profit hospitals of the year 1983 using the DEA approach. They used three outputs:medical –surgical acute discharges, medical-surgical acute discharges, medical –surgical intensive care discharges and Maternity discharges. Six inputs are specified: five labour inputs and one capital input (beds), the input prices were also used. The model specification was as follow s, for hospital R the minimum cost: MC_R (Y, X, P) for the technology is the solution to the following linear programming problem.

$$MC_R(Y,X,P) = \text{minimize } x\ _R{}^{\lambda}j\ P_{R.}\ X_R$$

$$\text{s.t} \quad Y\lambda \geq YR$$

$$X\lambda \leq X_R$$

$$\lambda_j \geq 0 \qquad \text{……………………. Equation 2.15}$$

The measure of technical efficiency for hospital R is denoted TE_R (Y,X) and was obtained by solving the following linear programming problem:

$$TE_R\ (Y,X\text{-}) = \text{minimize } \lambda_{j\theta}\theta$$

$$\text{s.t} \quad Y.\ \lambda \geq Y_R$$

$$X.\ \lambda \leq X_R$$

$$\lambda\ j\theta \geq 0 \qquad \text{……………………. Equation 2.16}$$

The scale efficiency component was computed by solving the following linear programming problem:

$$W_R(Y,X\text{-}) = \text{minimize } \lambda_{j\theta}\theta$$

$$\text{s.t} \quad Y.\lambda \geq Y_R$$

$$X.\lambda \leq X_R.\theta$$

$$\sum \lambda_j \geq 1$$

$$\lambda_j\theta \geq 0 \qquad \text{…………………} \text{ Equation 2.17}$$

The pure technical and congestion component for observation R, were solved by the following linear programming problem:

$$PTER(Y,X\text{-}) = \text{minimize } \lambda_{j\theta}\theta$$

$$\text{s.t} \quad Y.\lambda \geq Y_R$$

$$X.\lambda = \delta.\theta.X_R.$$

$$\delta \leq 1$$

$$\lambda_j\theta\,\delta \geq 0 \qquad \text{………………….} \text{ Equation 2.18}$$

Using the specification described above they computed cost minimizing efficiency, allocative efficiency and technical efficiency, including scale congestion and pure technical efficiency. The results showed that the primary source of inefficiency is allocative inefficiency that is most hospitals employed the wrong input mix given input prices so that their costs were 27% higher than the cost minimizing level. Almost 40% of the sample hospitals were technically efficient; the technical inefficient hospitals could reduce inputs in-puts by 16% on average and still have reduced the same level inpatient services. They also realized that scale inefficiency is the primary cause of technical inefficiency (60% of the sample).

Shearman (2001) used DEA analysis in the estimation of hospital efficiency in Massachusetts, in the United States of America in the year 1976. The researcher had no a priori knowledge about which of these hospitals was relatively inefficient, nor was any accepted benchmark available that independently indicated which of these hospitals. The outputs used were patient days with *** years of age, patient days with *** 65 year of age, number of nurse students, number of interns and residents in training. The inputs used were, full time equivalent non-physician, bed days available, supply dollars, DEA measured the efficiency of hospital O compared with n Hospitals in the set as follows:

$$\text{Max } E_o = \frac{\sum Ur\, y_{ro}}{\sum vi\, x_{jo}} \qquad \text{……………………. Equation 3.19}$$

Where O is the hospital being evaluated in the set of j = 1,………….n

Less than-unity constraints

$$1 > \frac{\sum Ur\, y_{rj}}{\sum vi\, x_{jo}}$$

Positivity constraints:

0 – Ur ; r = 1,……………s

0<vi;j =1,………………m

Where: ouputs: y_{rj} is the observed amount of r^{th} output for the j^{th} hospital; and inputs: x_{ij} is the observed amount of i^{th} input for the j^{th} hospital.

Two out of the six hospitals were found to have relative technical efficiency ratings less than one. Sherman compares these results with the ranking used by the Massachuttes Rate Setting Commission to measure hospital efficiency, where ratio analysis was used. He demonstrates that the commission's ratios (average cost per patient and per patient day) fail to distinguish technical inefficiencies. But because his analysis was limited to the estimation of technical efficiency, Sherman was unable to judge the validity of the Commission's ratios from an economic efficiency standpoint.

2.9.2 Review of Efficiency in Hospitals Studies in Developing Countries

Chang (1998) combined DEA (stage one) with Regression analysis (stage two) to evaluate the efficiency of 6 central government- owned hospitals in Taiwan over the fiscal years 1900 to 1994. Efficiency was first estimated using DEA with a choice of 3 inputs: full time equivalent employment of general and administration personnel. The two outputs were number of clinic visits and the number of weighted patient days. A multiple regression model was then employed in which the efficiency score obtained from the DEA computation is used as the dependent variables and a number of hospitals operating characteristics were chosen as the independent variables which were: scope of services reflecting the supply side of services to represent operating complexity (SCOPE), OCCUP is the occupancy rate for a hospital, the proportion of patients who are veteran is denoted by PCRET and YEAR is the proxy for the time of study, denoting the fiscal year for each unit.

The DEA model employed in the paper was specified as the following linear programme:

$$e(Xo; Yo) = \text{Min } e$$

$$\text{s.t } \sum_{j=1} \lambda_j X_j \leq e X_o$$

$$\sum_{j=1} \lambda_j Y_j > Y_o$$

$$\sum_{j=1} \lambda_j = 1$$

$$e \text{ and } \lambda_j \geq 0 \qquad \text{……………….. Equation 2.20}$$

The dual to the linear programme wa s formulated as follows:

$$H(X_o; Y_o) = \text{Max } UY_o + u_o$$

$$\text{s.t } U Y_j - VX_j + U_o \leq 0 \text{ for } j = 1,\ldots\ldots.n$$

$$VX_o = 1$$

$U \geq e$, $V \geq e$ and u_o is unconstrained in sign …………………. Equation 2.12

The Regression model was specified as:

$EM = \beta_o + \beta_1\, SCOPE + \beta_2\, OCCUP + \beta_3\, PCRET + \beta_4\, YEAR + e$ (equation 2.22)

The results of stage one reveal that from 1991 to 1994 the central government-owned hospitals, on average, experienced improvement in efficiency over time. The results of stage two indicate that the scope (SCOPE) of services and the proportion of retired veterans patients (PCRET) are negatively and significantly associated with efficiency, where as occupancy rates (OCCUP) is positively and significantly associated with efficiency. The coefficient for YEAR also has a positive and significant impact on efficiency.

Wong (1996) measured the technical efficiency of 560 hospitals in Korea. The first analysis began with measuring technical efficiency based on DEA, he used 16 outputs and 8 inputs, the second analysis used the DEA results as dependent variables and some independent variables (number of beds, beds squired, labour intensity, high tech indicator, length of stay, proportion of insured patients, ownership, location and teaching) in the two part-model (Logit model and Truncated regression model).

The DEA CRS model to measure overall technical efficiency was:

$$\text{Max } Ek = \sum u_r Y_{rk}$$

$$\text{s.t} \quad \sum v_i X_{ik} = 1$$

$$\sum {}_{ur} Y_{rj} - \sum v_i X_{ij} \leq 0$$

$$V_i, \ldots\ldots v_s > 0$$

$$U_i, \ldots\ldots u_m > 0$$

Where E_k is the efficiency score for hospital k. The input and output variables were as follows; Y_{rk} is the actual amount of output r produced by hospital k; X_{jk} is the actual amount of input 1 used by hospital k; U_r is the weight to input r, computed in the solution by DEA; V_i is the weight to input I, computed in the solution by DEA.

The DEA VRS model to measure pure technical efficiency and returns to scale was;

The two-part model in his study was

$$\text{Max } Ek = \sum u_r Y_{rk} + w_k$$

$$\text{s.t} \quad \sum v_i X_{ik} = 1$$

$$\sum {}_{ur} Y_{rj} - \sum v_i X_{ij} + W_k \leq 0$$

$$V_i, \ldots\ldots v_s > 0 \qquad r=1, \ldots\ldots\ldots\ldots, s$$

$$U_i, \ldots\ldots u_m > 0 \qquad i = 1, \ldots\ldots\ldots\ldots, m$$

$$j = 1, \ldots\ldots\ldots\ldots, n$$

The two-part model in his study was:

Logit model:

Pr (Y = 1) = 1 / [1 + exp – (β1 + β2x2 + β3 X3+ ……….)1 …………… Equation 2.25

Where Y = 1 if efficiency score (E) = 1,

Y = 0 if efficiency score (E) < 1

Trunscated regression model:

Ln (E) = $\beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\ldots\ldots + \varepsilon$ for the data E < 1

Where εi is the usual disturbance term …………....... Equation 2.26

The results were as follows: with respect to overall efficiency, the CRS model showed that 16 hospitals (25.8 %) are efficient. Regarding pure technical efficiency, the VRS model found 23 hospitals (37.1%) efficient. Asd for scale efficiency 19 hospitals (30.6 %) were efficient, and it is one of the main f actors that reduce overall efficiency in the hospital industry. The results from the two-part model showed that labour intensity, location and the portion of insured patients are significant factors in determining hospital efficiency, while high-tech indicators, ownership and proportion of insured patients are more significant variables in explaining the degree of inefficiency among the selected hospitals with a level of efficiency of less than one.

Kiriga, Emrouznejad a nd Sambo (2000) conducted a research on measurement of technical efficiency of 54 district level hospitals in Kenya using the DEA approach. They computed the efficiency by solving two fractional programming models: CRS (equation 2.27) and VRS (equation 2.28):

Max ho = $\sum_{ur} y_{rjo}$

Subject to : $\sum v_i X_{ijo} = 1$

$\sum u_r y_{rj} x_{ij} < 0$, j=1,……..n,

$-u_r \leq -$ €, r=1,……..t,

$-v_i < -$ €, r=1,………..,m, …………………… Equation 2.27

Max ho = $\sum_{ur} y_{rjo} + u_o$

s.t $\sum v_i X_{ijo} = 1$

$\sum u_r y_{rj} - S v_i x_{ij} + uo \leq 0$, j = 1,………….. N

Ur Vi ≥ 0 …………………… Equation2.28)

Where: y_{rj} is the amount of output r produced by hospital j: x_{ij} is the amount of input I used by hospital j; u_r is the weight given to output r; V_i is the weight given to input 1; n is the number of hospitals; t is the number of outputs; m is the number of inputs; € is a small positive number.

The findings are that, out of 54 district hospitals included in the analysis, 39 (72%) were technically efficient, whils the remaining 15 (28 %) were technically inefficient. On the other hand out of the 54 hospitals analyzed, the average scale efficiency score was 89.9 % implying that there is room to increase total outputs by 10.1%.

Kiriga, Lambo and Sambo (2000) conducted a technical efficient study on the public hospitals in Kwazulu-Natal province of South Africa. They employed a DEA analysis to identify and measure individual hospitals inefficiencies in 56 provincial hospitals

The model used was as follows:

$$\text{Max } E = \sum_{t=1}^{m} V_i \; X_{ijk}$$

s.t { ∑ur Yrj) / [Vi X- Xij] < 1, j = 1,2,……………n hospitals (equation 2.29)

Their findings were: forty percent of the hospitals had some degree of technical efficiency. Fifty eight percent were scale inefficient. In total the following inputs were being wasted and not utilized in the production of hospital outputs in Kwazulu-Natal public hospitals: 117.4 doctors (8%), 2709 nurses (11.9 %), 61 paramedics (11.5 %), 58 technicians (13.1 %), 295 administrators staff (11.1 %), 835 general staff (11.3 %), 1193 labour provision staff (14.3 %), 38 other staff (!0.7 %), and 1752 beds (7.1 %).

2.9.3 Synthesis of the DEA Literature Review

Since time in memorial, policy-makers and developing countries acknowledge that hospitals absorbed a misappropriate share of resources of the health sector. This is why most efficiency studies that are carried in the health sector concentrate on hospitals.

The theoretical study outlined the usage of DEA in measuring efficiency and has proven that DEA is a reliable and effective method of estimating the efficiency levels in health institutions. Furthermore the studies in Africa (like in the case of Kenya and South Africa) in Section 3.4 that used DEA only measured technical efficiency aspect (and decomposed it into: pure technical efficiency and scale efficiency). This was mainly due to data limitation, which is strongly evident in AFRICA. Therefore since this study will also be constrained by data and time limitation, it will adopt the models by Kiriga (2000) used in Kenya and kwazulu Natal (South Africa.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter outlines the method use to carry out this study. This includes description of the study area, method of data collection and model specification. First we look at theory behind the DEA model used.

3.2 Theoretical Framework of the study

The concept of DEA was introduced in the journal literature by the highly influential 1978 papers of Charnes, Cooper and Rhodes. The ideas were borrowed from Farell's who laid the foundation in his seminal 1957 paper on concepts of efficiency and productivity, and how to calculate the benchmark technology and the efficiency measures. The fundamental assumptions was the possibility of efficient operations, immediately pointing to *frontier production function* concept as the benchmark, as opposed to notion of *average performance* underlying most econometric literature on production function up to the time of seminal contribution.(Forsund and Sarafoglou 2010)

The of Farrel was path breaking as to three aspects:

Efficiency measures were based on uniform contractions or expansions from inefficient observations to frontier,The production frontier was specified as the specified as the most pessimistic piecewise linear envelopment of the data and The frontier was calculated through solving systems of linear equations, obeying the two conditions on the unit isoquant:

(a)That its slope is not positive.(b)That no observed point lies between it and the origin.

(Forsund and Sarafoglou 2010: Farell 1957)

Efficiency and productivity are core concepts of economics; the new aspect of farell was to offer decomposition into technical efficiency, price (allocative) efficiency and overall efficiency (See section 2.1 for Farell’s graphical illustrations) at the micro level of a firm (or production unit).

In the choice of a production frontier benchmark Farell adopts a most practical approach, starting with engineering considerations and ending up with recommending observed best practice. Inspired by the activity of analysis of Koopmans in 1951 his contribution was to introduce a *piecewise linear* envelopment of the data as the most pessimistic specification of the frontier, in the sense of the production being as close to the observation as possible and to show how the frontier could be established by solving linear questions.

Farell's original ideas above were developed in the 1978 paper "Measuring the efficiency of decision making units" by Abraham Charnes, William Cooper and Edwardo Rhodes (CRR), who covered the same ground ad regard the efficiency measure concept as Farell. That is both the proposed efficiency measures and the framework of a piecewise linear production technology was identical. But the linear programming model formulated was a genetic one, and was quite superior to Farell's unit isoquant approach in the case of a singe output, and the change of origin approach proposed in the multiple output case in Farell and Field house in 1962 (for constant returns to scale) one unique contribution of CCR is the explicit connection made between a productivity index in the form of weighted sum of outputs on a weighted sum of inputs and thee Farell technical efficiency measure (in the case of constant returns to scale). This was the starting point in CCR: Finding weights by maximization of such a productivity ratio subject to the best practice and normalization constraints, the so-called ratio form of CCR, corresponds to the natural science –engineering definition of efficiency. A bridge was offered between the engineering concepts of micro productivity ratios and the economists' concept of efficiency.

Economics is *the social sciences concerned with the problem of using or administrating scarce resources (the means of producing) to attain the greatest or maximum fulfillment of society's unlimited wants (the goal of s producing). Hence economics is a science of efficiency; efficiency in the use of scarce resources; society wants to use its resources* *efficiently so as to get maximum amount of useful goods services from its scarce resources (McConnell and Bruel 1996).*

Therefore economics is concerned with doing the best with what we have. But if our resources are scarce, we cannot satisfy all our unlimited material wants. The next best thing is to achieve the greatest possible satisfaction of these wants, in doing so we need to apply the concept of efficiency. Mooney (1995) defines efficiency as maximizing the benefit (however defined to society at large) from the resources available (however constrained). Bitran (2009) sees efficiency as a point where inputs are combined to produce a given level of output at minimum cost.

There is a long tradition of measuring efficiency in the applied economics literature, especially in the fields of agricultural and industrial economics. Today, efficiency measures are being applied too an increasing number of fields and used to evaluate and compare educational departments (schools, colleges and universities), health care facilities (hospitals, clinics), prisons, agricultural production, banking, sports and many and many others (Evans etal., 2000).

3.3 Conceptual Framework of Study

*Hospitals are assumed to use **7*** **inputs*****:*** (1) number of beds, (2) number of full time doctors (3) number of full time nurses, (4) number of full time technicians, (5) number of full time administrative staff, (6) number of full time general staff. (7) Number of short term training

*Hospitals are assumed to produce **4 outputs**:* The number of admission measures these, and is consistent with practice elsewhere in the health economics literature, where output is equal to services provided or the outcomes of the services give n. For this study they are (1) maternity admissions, (2) general admissions, and (3) tuberculosis admissions. (4) out-patients

Where:

General admissions: The total number of male, female and paediatric admissions in each DMU in the year in question.

Maternity admissions: The total number of TB admissions in each DMU in the year in question.

Short term training: The total numbers of personnel who have undergo training in a period of less than one year in various field of their specialization in the year in question.

Beds: The total number of beds found in all the wards in each DMU.

Doctors: This refers to the number of full time Medical officers found in each DMU.

N*urses*: The total number registered and enrolled nurses in each DMU.

Technicians: The total number of full mechanical, pharmacy, medical, laboratory, audiology and radiology technicians in each DMU.

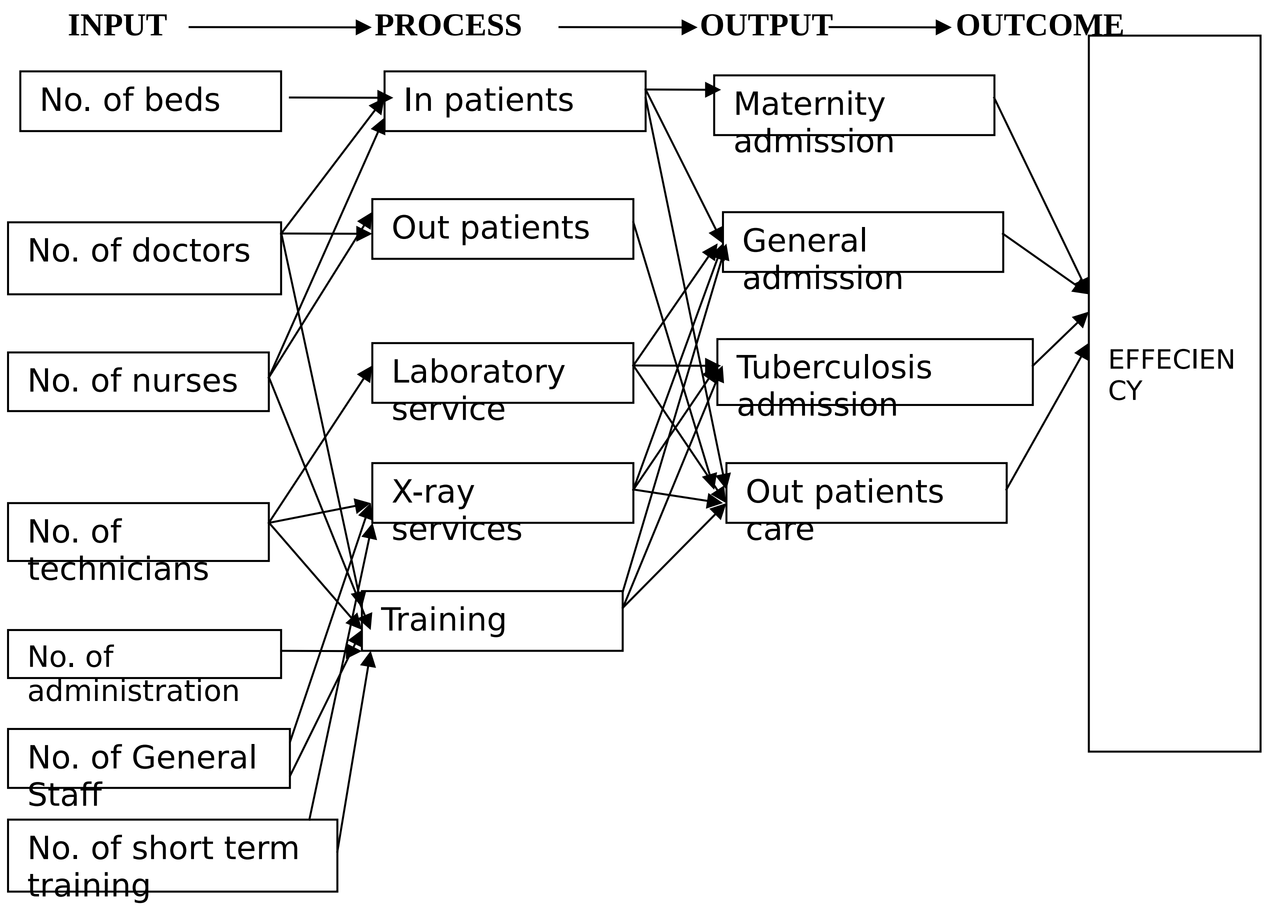
Administrative staff: The total number of individuals responsible for the administrative work in each DMU.

General staff: The total number of individuals who support the running of the hospitals, cooks, gardeners, cleaners, drivers

TB admission: The total number of TB admission in each DMU in the year in question.

Out patients: the total number of patients treated and let to go home in the year in question.

Figure 3.1 Production Process for Health Institutions



Source: Author

2013 3.4 Model Specification

The study adopts a DEA model based on those used in studies by Kiriga et al (2000) in Kenya and Kwazulu Natal.

Under the restriction that each units efficiency is judged against its individual criteria (individual weighting system), efficiency of a target unit Di is obtained as a solution to the following problem: maximize the efficiency of unit i, under the restriction that the efficiency of all units is≤ 1. The algebraic model ifs (following the definition of Dyson and the Frank 2001 in Section 2.2)

$$\text{Max } Di = \frac{\sum m_{r1}\, y_{r1}}{\sum n_{j1}\, X_{j1}}$$

s.t $\frac{\sum m_{r1}\, y_{r1}}{\sum n_{j1}\, X_{j1}} \leq 1$

$Mr,nj \geq 0$ …………… Equation3.1

Where: D_1 is technical of unit 1 to be estimated; M_r and n_j are variables to be estimated: Y_i are the outputs of the ith unit; Xi are the inputs of the ith; i indicates the k different units; r indicates the t different outputs; j indicates q different inputs.

The m's and n's are variables of problem and are constrained to be greater or equal to 0. The solution of the above model in relation to unit 1 gives the value D_1, the efficiency of unit 1 and the weights, m and n, leading to that efficiency.

DEA problem of equation (3.1) is a fractional linear program where the numerator is maximized and the denominator minimized simultaneously, which will lead to infinite number of solutions. Hence the easier way is to convert it into a linear form and apply the maximizing the numerator. Applying the transformation that was developed by Charnes, cooper and Rhodes in 1978, the model becomes:

Max m,n $D1 = \sum m_{r1}\, y_{r1}$

s.t $\sum m_{r1}\, y_{r1} - \sum n_{j1}\, X_{j1} \leq 0$ for each unit i

$\sum n_{j1}\, X_{j1} = 1$

$m_r, n_j \geq \varepsilon$ ……………………. Equation 3.2

Since this model includes input and output slacks, the constraint: m_r, n_j . 0 in model 1, is replaced by the constraint m_r and n_j are > to some small positive quantity e. This is to avoid any input or output (that lay parallel to the axes in the piecewise linear frontier) being totally ignored in determining the efficiency scores.

This linear programming problem will be solved k times (where k is the total number of DMUs in the sample), once for each unit in the sample. A value of D_i (efficiency score) is then obtained for each units. Since with linear programming the more the constraints the more difficult a problem is to solve, the dual model (which has less constraints) is instead used. The dual model is be constructed by assigning a variable (dual variables) to each constraint in the model and constructing a new model on these variables.

$$\text{Min}_{\Theta\lambda}\ \Theta 1-[\sum \varepsilon P^{+}_{JI} + \sum E p^{-}_{ri}]$$

$$\text{s.t} \quad -yrI + \sum \lambda i1yri\text{-}P\text{-}r1 = 0$$

$$\Theta 1 x_{ji} - \sum \lambda_{i1}\, x_{ji\text{-}} P^{+}_{j1} = 0$$

$$\lambda_{i1} \geq 0 \quad \text{……………………. Equation 3.3}$$

Where: Θ_1 is the technical efficiency score for unit 1 to be estimated; λ_i is a k-dimensional constant to be estimated; I indicates the k different unit as; e is some marginally small, but positive quantity; S^{+}_{r} are the slack variables for t outputs; S^{-}_{r} are the slack variables for q inputs.Since CRS assumes that all units are at optimal scale (which is not always the case), VRS is also used (equation 3.4), where technical efficiencies of variables that are confounded with scale efficiencies are measured. This will be done by adding $\lambda_{i1} = 1$to equation (3.3) meaning that under VRS the λ add to one.

$$\text{Min}_{\Theta\lambda}\ \Theta 1-[\sum \varepsilon\, P^{+}_{JI} + \sum E P^{-}_{ri}]$$

$$\text{s.t} \quad -yrI + \sum \lambda i1yri\text{-}P^{-}r1 = 0$$

$$\Theta 1 x_{ji} - \sum \lambda_{i1}\, x_{ji\text{-}} P^{+}_{j1} = 0$$

$$\sum \lambda_{i1} \geq 1 \quad \text{…………….. Equation 3.4}$$

The reason for modeling both CRS is that it enables the decomposing of technical efficiency into scale efficiency and to pure technical efficiency.

3.5 Justification for the Model Used

3.5.1 CRS and VRS DEA and TE

Most researchers when studying hospital performance focus on assessing hospital efficiency in technical terms only, and decompose this technical efficiency into pure technical efficiency and scale efficiency, by using CRS and VRS. For example, Shearman (2006), Kiriga et al (2000), and the main reason is that calculation of allocative component and hence, overall cost-minimizing efficiency, requires information on the relative prices of inputs and outputs. But such information on relative prices of inputs and outputs is difficult to obtain especially in health care, and this is because, there are usually large dispersion of prices in both inputs and output of health care caused by differences in quality and case mix.

This study also measured technical efficiency and decomposed it to pure technical efficiency and scale efficiency, by using CRS and VRS) on the grounds that information on input and output prices will be difficult to determine especially on the output side, and this is because: In general, health services are heterogeneous in nature (due to quality and case mix) and therefore heterogeneous in price.In Kenya, the user-charge for service in government hospital gives waivers and exemption on those who are not able to pay. This decision is made at the local level by local authority or social workers, therefore services are offered at different prices (Moalosi 2008).

3.5.2 Input-Oriented Measures

Coelli (20005 noted that studies tend to select input-oriented models rather than the output based ones, and this is because most DMUs have particular orders to fill and hence the input quantities appear to be the primary decision variables, but this is not the case in all DMU, that is some DMUs may be given a fixed quantity of resources and asked to produce as much output as possible. In this case an output orientation would be more appropriate. Therefore states that one should select an orientation according to the quantities of decision making variables and that the choice of orientation will have only minor influences upon the scores obtained.

Therefore this study uses the input-oriented model, because in the case of Kenya the inputs are the primary decision variables, and furthermore the health sector has had an increase in most of its inputs (finances from donors, human capital) due to HIV / AIDS epidemic.

3.4. Sample Selection, Target population and Variable Measurement

3.4.1 Sample Selection

Given the size of Kenya, undertaking a study that covers all the health care facilities in the Country is not feasible. What is feasible is to focus on a particular portion of the health facilities. Hence this study specifically looks at hospitals. Furthermore to avoid the problems brought about by case mix and quality of care when measuring efficiency in health facilities (see Chapter 2 section 2.2.1), the sample only consist of hospitals.

However since the total number of hospitals in North Rift is small, this study had a small population in which to choose the sample from. That is, out of 43 hospitals, only 28 were sampled. This is because the rest of the hospitals are health centers without admission facilities. There were three types of hospitals namely:

Government / public hospitals; these are hospitals that are funded and run by the government, and serve the whole public.

Mission hospitals; these are hospitals that are funded by the government and churches and subsidized by government, although churches run them, there is some presences of government control, and they serve the public and lastly.

Private hospitals: these are funded and run by private organization. Though they are open to the public, are not subsidized by the government.

3.5.2 Target Population

The study focused at Mission, Private and Government hospitals in four counties namely Nandi,Uasin Gishu,Trans-Nzoia and Elgeiyo/Marakwet.

Table 3.2 Public, Mission and Private Hospital in the North Rift Region

County	Hospitals	Classifications	Code
Nandi County	Nandi Hill Hospitals	Public	1
Nandi county	Kapsabet District Hospital	Public	2
Nandi County	Maraba District Hospital	Public	3
Nandi County	Alwara kapsabet Hospital	Private	4
Nandi County	Baraton University Hospital	Private	5
Nandi County	Kobujoi Mission Hospital	Mission	6
Nandi County	Mosoriot District Hospital	Public	7
Uasin Gishu County	Elgon View Hospital	Private	8
Uasin Gishu County	Uasin Gishu District Hospital	Public	9
Uasin Gishu County	MediHeal Hospital	Private	10
Uasin Gishu County	Eldoret Hospital	Private	11
Uasin Gishu County	Memorial Hospital	Private	12
Uasin Gishu County	Moi-Teaching and Referral Hospital	Public	13
Uasin Gishu County	Moi‘s Bridge District Hospital	Public	14
Uasin Gishu County	Ziwa District Hospital	Public	15
Uasin Gishu County	Soi District Hospital	Public	16
Uasin Gishu County	Burnt Forest Mission Hospital	Mission	17
Trans-Nzoia County	Kitale District Hospital	Public	18
Trans-Nzoia County	Mt .Elgon Hospital	Private	19
Trans-Nzoia County	Cherangany Nursing Home	Private	20
Trans-Nzoia County	Cherangany District Hospital	Public	21
Trans-Nzoia County	Kapsakwony District Hospital	Public	22
Trans-Nzoia County	Kitale Nursing Home	Private	23
Elgeiyo/Marakwet County	Iten District Hospital	Public	24
Elgeiyo / Marakwet County	Kaptarakwa Hospital	Public	25
Elgeiyo/Marakwet County	Plateau Mission Hospital	Mission	26
Elgeiyo/Marakwet County	Kaptagat Hospital	Private	27
Elgeiyo/Marakwet County	Tambach District Hospital	Public	28

Source;Ministry of Health and sanitation.

3.5.2 Variable Measurement

Hospitals are assumed to use ***7*** **inputs*:*** (1) number of beds, (2) number of full time doctors (3) number of full time nurses, (4) number of full time technicians, (5) number of full time administrative staff, (6) number of full time general staff.(7) number of short term training.

Hospitals are assumed to produce ***4 outputs****:* The number of admission measures these, and is consistent with practice elsewhere in the health economics literature, where output is equal to services provided or the outcomes of the services give n. For this study they are (1) maternity admissions, (2) general admissions, and (3) tuberculosis admissions. (4) out –patients.

Where:

General admissions: The total number of male, female and pediatrics admissions in each DMU in the year in question.

Maternity admissions: The total number of maternity admissions in each DMU in the year in question.

Short term training: The total numbers of personnel who have undergone training in a period of less than one year in various field of their specialization in the year in question.

Beds: The total number of beds found in all the wards in each DMU.

Doctors: This refers to the total number of full time Medical officers found in each DMU.

N*urses*: The total number registered and enrolled nurses in each DMU.

Technicians: The total number of full mechanical, pharmacy, medical, laboratory, audiology and radiology technicians in each DMU.

Administrative staff: The total number of individuals responsible for the administrative work in each DMU.

General staff: The total number of individuals who support the running of the hospitals, cooks, gardeners, cleaners, drivers.

TB admission: The total number of TB admission in each DMU in the year in question.

Out patients: the total number of patients treated and let to go home in the year in question.

Table 4.2 is reproduced in appendix 1 for easier reference when analyzing the results.

Table 3.3 Codes Used in the Measurement of Variables.

Variables	**Measurements**
Outputs	
Y_1	No. of maternity admissions
Y_2	No. of general admissions
Y_3	No. of tuberculosis admissions
Y_4	No. of out patients care
Inputs	
X_1	No. of beds.
X_2	No. of full time doctors.
X_3	No. of full time nurses.
X_4	No. of full time technicians.
X_5	No. of full time administrative staff.
X_6	No. of full time general staff.
X_7	No. of short term training

The selection of the variables was influenced by;

The availability and uniformity of data in the hospitals, and

The DEA degrees of freedom rule (Dyson and Frank 2001) which states that if there are t outputs and m inputs, then the number in the set should be substantially greater or equal to tm in order for there to be suitable discrimination between the units.

Therefore in this study DMU (28) = tm (4 x 7 = 28)

3.6 Data Sources, Types and Estimation Method

The study used cross sectional data for the year 2010. Information was obtained from published records obtained from Statistics Unit of Ministry of Health, Health Research Units and in the individual hospital where necessary. Information was also obtained from annual reports and statistical bulletins of Kenya. The study used the mult-stage “DEAP Version 2.1 Computer Programme” by Coelli (2005) to estimate equation 3 and 4 respectively with each hospital to obtain individual efficiency rates.

A sensitivity analysis was done to test the sensitivity of the efficiency results to changes in the input-output specification. This was done by running a series of DEA, with gradual increase in the number of outputs and inputs used. The rationale for this approach, which was also used by Mickillop (1999), was to determine whether there was any consistency between the different results.

CHAPTER FOUR

EMPIRICAL RESULTS AND DISCUSSION

4.1 Introduction

This chapter is mainly concerned with the presentation and discussion of results obtained from DEAP version 2.1 computer programme. The computer programme was used to compute the standard CRS and VRS DEA models that involve the calculation of technical and scale efficiencies. A summary of the descriptive statistics of inputs and output is given, this is followed by the DEAP version 2.1 computation of equations (3) and (4) from chapter 3 repetitively with each hospitals in the sample to derive individual overall technical efficiency, which is decomposed into pure technical and scale efficiency respectively. These efficiency scores are presented in the first section of this chapter.

Next a calculation of excessive inputs based on DEA evaluation for each inefficient hospital is done; this is achieved by using Efficiency Reference Set (ERS) for each hospital constructed by applying weights (the dual variables from the DEA linear programme) to the actual and inputs in each of the hospitals. An alternative method of calculating excessive inputs in each inefficient is the use of Input Slacks, which is also presented.

Comparative analysis is then done on the hospitals based on ownership / management and location differences. The sensitivity analysis revealed a considerable degree of stability in the efficiency scores when using different input-output specifications around the 4 outputs and 7 inputs specification that was decided upon.

4.2 Descriptive Statistics

The results obtained showed that in overall input, the Out-patient contributed the highest number 17354.60 compared to Maternity 1026.39,General 1742.43 and Tuberculoses 114.25

Table 4.1 Descriptive Statistics (Inputs and Outputs: n=28)

Variables	Mean	Median	Standard Deviation	Minimum Value	Maximum Value
MAT	1026.39	695.5	743.11	171	2588
GEN	1742.43	1300	1145.79	346	4352
TB	114.25	93.5	74.35	10	279
Out- patients	17354.60	12948	11412.07	3446	43346
Beds	74.5	50	53.02	23	177
Doctors	5.035	3	4.07	2	22
Nurses	54.5	31	38.94	17	147
Technician	8.035	6	5.03	2	19
Administrative	4.107	3.5	2.35	1	9
General	33.32	31	21.59	8	93
Short training	34.99	24.83	23.99	10	97

Source: Authors Data Analysis Results [2013]

4.3 Constant Returns to Scale DEA Results and Discussions

Recall that DEA identifies the optimal input / output combination and presents it with "best practice frontier" or data envelope. The DMUs that make up the frontier are assigned a score of one, are technically efficient and become peers to the inefficient ones. While all the others not on the frontier are assigned a score between zero and one, and are technically inefficient. The results obtained from solving the CRS DEA are presented in tables 4.2; the table also contains the peers group for each hospital and the peer count.

The results show that the overall level of TE (Technical Efficiency) among the North Rift Region private, public and mission hospitals are 83.7 %, which means that in average the hospitals should be able to reduce the consumption of all inputs by 16.3% without reducing output levels.The results further show that 13 of the 28 (46.43 %) hospitals are technically efficient, that is had scores of 100% (1,000), when we look at the peer count column however, it is far more likely that hospitals 5,6,10,11,25 and 28 are truly efficient, because they are peers for five or more other hospitals in the sample. Hospitals 5,8,12 and 18 each appear in only two or three peer groups, meaning that there is scope for them to improve their efficiency further even though they receive efficiency scores of 100%.

Table 4.2: CRS DEA efficiency rating of hospitals in North Rift Region

Hospital	CRS TE (Overall TE scores)	Efficiency reference set / peer group	Peer count
1	0.576	13,25,6,12	0
2	0.855	13,6	0
3	0.309	13,28,10,5	0
4	1.000	3	0
5	1.000	5	2
6	1.000	6	7
7	0.752	13,6,28	0
8	1.000	8	1
9	0.908	13,25,11,10,18	0
10	1.000	10	3
11	1.000	11	5
12	1.000	12	2
13	1.000	13	14
14	0.773	13,25,6,11	0
15	0.303	13,25	0
16	0.661	25,6,12	0
17	0.830	13,25,28	0
18	1.000	18	2
19	1.000	8	0
20	0.616	13,11	0
21	0.571	13,28,118	0
22	0.970	13,28,11	0
23	0.786	13,25,6	0
24	0.995	13,25,6	0
25	1.000	25	9
26	1.000	26	1
27	0.535	13,6,10,18,27,28	0
28	1.000	28	6
29			

Source: Authors Data Analysis Results [2013

The results also show that 15 of the 28 (53.57 %) hospitals are relatively inefficient compared with the other hospitals in the data set, this means that they have a TE score of less than one. The technically inefficient hospitals should be able to produce the same level of outputs with fewer inputs, and therefore at lower cost. Among the inefficient hospitals the ranges of the overall TE scores were:

7.14 % are below 0.50, 10.71% between 0.51 and 0.607.14 % between 0.61 and 0.70, 10.71 % between 0.71 and 0.80,7.14% between 0.81 and 0.90,10.71% between 0.90 and 0.99

Tables 4.3 and 4.4 show the descriptive statistics of efficient (4.4) and inefficient (4.3) hospitals. On average the efficient group and inefficient group have the same hospital size (beds range 25-175 inefficient group and beds range 23-177 in efficient groups), but the hospital group in Table 4.4 is using on average less inputs, (than the group in Table 4.3) but is producing more outputs, that is why the DEA ranked this group as being efficient relative to hospital group in Table 4.3)

Table 4.3 descriptive Statistics of Inefficient Hospitals (TE <1.000 n= 15)

Variables	Mean	Median	Standard Deviation	Minimum Value	Maximum Value
Maternity	834.06	465	689.46	255	2432
General	1550.26	968	1070.85	346	3711
Tuberculoses	105.86	74	84.33	10	279
Out –patients	15440.59	9641	10665.67	3446	36962
Beds	74.53	50	49.42	25	175
Doctors	5.86	3	5.15	2	22
Nurses	59	28	44.14	23	147
Technicians	8.66	6	5.35	3	19
Administrative	4.33	4	2.22	2	9
General	39.8	31	22.52	17	93
Short term training	39.22	24	26.46	16	97

Source: Authors Data Analysis Results [2013]

Table 4.4N Descriptive Statistics of Efficient Hospitals (TE = 1.000 n = 13)

Variables	Mean	Median	Standard Deviation	Minimum Value	Maximum Value
Maternity	1248.30	1045	766.99	171	2588
General	1964.15	1435	1231.50	464	4352
Tuberculoses	123.92	114	62.84	26	220
Out- patient	19562.93	14293	12265.74	4621	43346
Beds	74.46	41	58.95	23	177
Doctors	4	3	2.08	2	8
Nurses	49.30	41	32.91	17	114
Technicians	7.30	6	4.73	2	19
Administrative	3.84	3	2.54	1	8
General Staff	29.46	29	19.89	8	83
Short term training	3130	27	2072	10	28

Source: Authors Data Analysis Results [2013]

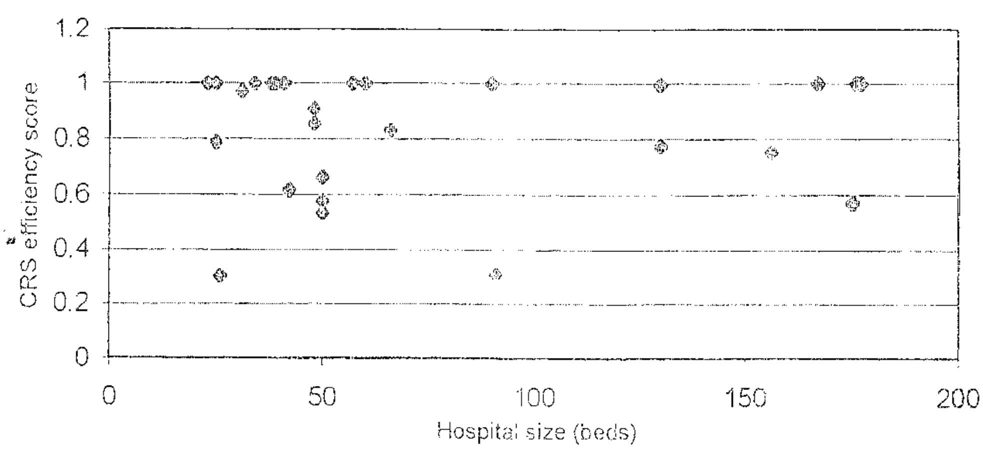


Figure 4.1: Hospital Size and CRS Efficiency Scores.

Source: Author’s Data Analysis Results [2013]

Figure 4.1 is a chart of the CRS efficiency scores and bed size, it indicates that inefficiency scores in small hospital groups are very high and volatile but they are getting higher and narrower as hospital size increases up to around 50 beds, then decreases again.

It has usually been expected that the larger the hospital, the less efficient it will be. However, the results in this study are quite different from the conventional view. It seems that the minimum optimal scale in hospitals is larger than expected from the literature. Although the results in figure 4.1 do not exactly imply the existence of minimum optimal scale, it may be inferred that hospitals are likely to be efficient at a range of approximately 60 to 130 beds.

4.5 Variable Returns to Scale DEA Results and Discussions:

The DEA approach enables the change to Variable Returns to Scale (VRS), this change relaxes the simplistic assumption that inputs normally will move in exact proportions to the scale of operations: and allows for the existence of economies and diseconomies of scale. The results for the DEA run with VRS for the 28 hospitals are presented in Table 4.5 Column 2 of Table 4.2 is equal to column 2 of Table 4.5, which is both look at the overall technical efficiency score

Table 4.5 VRS DEA efficiency rating of hospitals in North Rift Region.

Hospital	CRS TE (Overall TE)	VRS TE (Pure TE)	CRS / VRS (Scale efficiency)	Return to scale	Efficiency reference set /peer group	Peer council
1	0.576	0.926	0.622	IRS	13,22,25	0
2	0.855	0.877	0.975	DRS	13,6,12,25	0
3	0.309	0.432	0.717	IRS	10,25	0
4	1.000	1.000	1.000			0
5	1.000	1.000	1.000			
6	1.000	1.000	1.000		6	4
7	0.752	1.000	0.752	DRS		
8	1.000	1.00	1.000		12	
9	0.908	0.935	0.971	IRS	10,11,2,2,13,25,12	
10	1.000	1.000	1.000		10	2
11	1.000	1.000	1.000		11	3
12	1.000	1.000	1.000		12	3
13	1.000	1.000	1.000		13	6
14	0.773	1.00	0.773	DRS	14	1
15	0.303	0.906	0.335	IRS	22,25	
16	0.661	0.880	0.751	IRS	6,22	
17	0.803	0.982	0.845	DRS	13,14,26,25	
18	1.000	1.000	1.000		18	1
19	1.000	1.000	1.000			
20	0.661	0.837	0.737	IRS	25,13,12,6,11,22	
21	0.571	0.785	0.725	DRS	12	
22	0.970	1.000	0.970	IRS	22	5
23	0.786	1.000	0.786	IRS		
24	0.995	1.000	0.995	DRS		
25	1.000	1.000	1.000			9
26	1.000	1.000	1.000		26	1
27	0.535	0734	0.730	IRS	6,22,25,12	
28	1.000	1.000	1.000			3

Source: Authors Data Analysis Results [2013]

But the VRS DEA further decompose this overall technical efficiency score into one due to pure technical (column 3) and one due to scale efficiencies attributed to either the pure TE score is less than one or the SE score being less than one. The peer counts and peers groups are different between CRS DEA (table 4.2) and VRS DEA (table 4.5). This is because the VRS DEA efficiency scores also include scale efficiency in the assigning of peer groups and peer counts.

Table 4.6: Differences between CRS and CRS DEA peer count

Hospital	CRS Peer Count	VRS Peer Count
1	0	0
2	0	0
3	0	0
4	0	0
5	2	0
6	7	4
7	0	0
8	1	0
9	0	0
10	3	2
11	5	3
12	2	3
13	14	6
14	0	1
15	0	0
16	0	0
17	0	0
18	2	1
19	0	0
20	0	0
21	0	0
22	0	0
23	0	5
24	0	0
25	9	9
26	1	1
27	0	0
28	6	3

Source: Authors Data Analysis Results [2013]

The differences in peer count between n CRS and VRS is shown in Table 4.6 while in some hospitals the peer counts have decreased especially in hospital 13 (where the peer count has decreased from 14 to 6), in others the peer counts have increased, especially hospital 23 (where the peer count increased from 0 to 5), while in others like hospital 25 the peer count has remained the same.

As before in CRS DEA, hospital 6,13 and 25 are still peers for more hospitals in the sample and have now been joined by hospital 23 Under VRS DEA, which means that they are truly efficient.

Hospitals 14,18 and 26 are peers to themselves , while 4, 8 and 9 do not appear in any peer groups, this indicates that they were found apparently efficient by default because there were no other hospitals of comparable organization.

Table. 4.7 Summary of Efficiency Results

	Overall Technical Efficiency (CRS)	**Pure Technical Efficiency (VRS)**	**Scale Efficiency (CRS / VRS)**
Mean	0.837	0.939	0.882
Median			
SD	0.2137	0.1217	0.1660
Minimum Value	0.303	0.432	0.335
Maximum Value	1.000	1.000	1.000
Hospitals on frontier	5,6,8,10,11,12,13,18,19,25,26,28	5,6,7,8,10,11,12,13,14,18,19,22,23,24,25,26,27,28	5,6,8,10,11,12,13,18,19,25,26,28
Total number of Hospitals on frontier	13	18	13

Source: Authors Data Analysis Results [2013]

Table 4.7 summaries efficiency result, which shows that 13 hospitals (46.43 %) in the sample have overall efficiency scores of 1.000 and those are hospitals making up the frontier. Under pure technical efficiency 18 hospitals (64.29 %) had efficiency scores of 1.000, which are also the hospitals making up the frontier. Under scale efficiency 13 hospitals (46.43 %) had efficiency scores of 1.000 and are the hospitals making up the frontier.

The average pure technical efficiency score of the hospitals in the sample is 0.939, while the average scale efficiency score of the hospitals is 0.882 , this shows that on average the hospitals in the sample performed better under pure TE than under Scale efficiency.

In technical efficiency, results showed that the major contributory factor of inefficiency in the North Rift Region Hospitals is the scale size.

Table 4.8: Ranges of Inefficiency Scores of Hospitals.

Ranges	Overall TE		Pure TE		Scale efficiency	
< 0.50	7.14 %	2	3.57 %	1	3.57 %	1
0.51-0.60	10.71 %	3	0.0 %	0	0.0 %	0
0.61-0.70	7.14 %	2	0.0 %	0	3.57 %	1
0.71-0.80	10.71 %	3	7.14 %	2	28.57 %	8
0.81-0.90	7.14 %	2	10.71 %	3	3.57 %	1
0.91-0.99	10.71 %	3	14.28 %	4	14.29 %	4
Total	**53.57 %**	**15**	**35 . 71 %**	**10**	**53. 57**	**15**

Source: Authors Data Analysis Results [2013]

Table 4.8 shows the ranges of efficiency scores among the inefficient hospital in the sample, with relation to overall, pure and scale inefficiency where under the overall TE scores we had 15 hospitals that are inefficient, 10 hospitals had inefficiency scores under pure technical efficiency and 15 hospitals had inefficiency scores under scale efficiency. While the overall inefficiency scores of the hospitals in the sample are almost evenly distributed throughout the six ranges, under the pure TE scores, most of the hospitals had inefficiency scores ranging from 0 .81 to 0.91 to 0.99. This range of 0.81 and 0.999 represents 24.28 % of the total hospitals in the sample. And under scale efficiency, most of the hospitals had inefficiency scores ranging from 0.71 to 0.80 and 0.90 to 0.99, which represents 42.86 % of the sampled hospitals.

4.5.1 Findings Related to Pure TE

Recall that pure technical efficiency is the portion of overall technical efficiency, which is solely attributed to the combination of inputs that produce the maximum amount of outputs.

Under pure technical efficiency (column 3 of table 4.5), the results show that 18 of the 28 (64 .29 %) hospitals are technically efficient, while 10 of the 28 (35.71 %) hospitals are relatively inefficient compared with the other hospitals in the data set, this means that they have a TE score of less than one. Among the inefficient hospitals (35.71 %) the ranges of the pure TE scores were:3.57 % had less than 0.50,0.0 % between 0.51 and 0.60,0.0 % between n 0.61 and 0.70, 7.14 % between 0.71 and 0.80,10 % between 0.81 and 0.90,14.28 % between 0.90 and 0.99

These inefficient hospitals need to reduce inputs so as to become efficient. There are two methods of calculating input reductions needed in individual hospitals. Using Efficiency Reference Set (ERS) or using the Input Slacks.

Method 1: How to obtain the input reductions needed in each hospital (using ERS): This is the method that was used by Kerugma (2000) in calculating the input reductions that were needed in Kwazulu-Natal hospitals.

The ERS hospitals are the basic vectors of the linear programme solution. That is, a convex combination of the actual outputs and inputs of these ERS hospitals results in a composite hospital that produces as much or more outputs as the inefficient hospital while using fewer inputs than the inefficient hospitals. This information about ERS is usually a direct output of DEA (see in Table 4.5) By identifying the ERS, DEA allows one to focus on a subset of these hospitals to understand better the inefficiencies present.

Recall that the Efficiency Reference Set (ERS) is the group of hospitals against which DEA locates the inefficiency hospitals and the magnitude of inefficiency. The meaning of the inefficient rating derived from ERS, can be better understood by looking at the results of, for example, hospitals 27. DEA shows that hospital 27 should be able to produce its actual outputs levels using 46.5 % = (1.000 – 0.535) x 100 less of each input. More specifically, DEA indicates that the inefficiency was located and measured by comparing hospital 2 with its ERS; hospitals 12, 13, 19 and 25, which is shown

4.9: Summary of Hospital peers and peer Weight / Lambda Weights

Hospital	**Hospital Peers and Peer Weight / Lambda (:) Weights**					
1	13(0.053)	22(0.223)	25(0.723)			
2	13 (0.065)	6(0.446)	12 (0.428)	25 (0.060)		
3	13(0.265)	28(0.603)	10(0.132)			
4	4 (1.000)					
5	5 (1.000)					
6	6 (1.000)					
7	7 (1.000)					
8	8 (1.000)					
9	10 (0.074)	22(0.084)	25(0.131)	12(0.397)	11(0.046)	13(0.269)
10	10(1.000)					
11	11 (1.000)					
12	12 (1.000)					
13	13 (1.000)					
14	14 (1.000)					
15	22 (0.719)	25(0.281)				
16	25 (0.509)	6(0.051)	120.131)	22(0.309)		
17	13 (0.092)	25(0.409)	28(0.002)	14(0.497)		
18	18(1.000)					
19	19(1.000)					
20	13 (0.086)	6(0.483)	11(0.016)	22(0.016)	25(0.011)	12(0.044)
21	13 (0.857)	28(0.143)				
23	22 (1.000)					
23	23 (1.000)					
24	24(1.000)					
25	25(1.000)					
26	26(1.000)					
27	13(0.113)	6[a](0.201)b	10(0.261)	18(0.426)		
28	28(1.000)					

Source: Author's Data Analysis Results [2013]

The weighed composite of the ERS hospitals yields a hypothetical hospital 25 that produces as much or more outputs as the inefficient hospital 27 but uses fewer inputs than 2. This composite is constructed by applying the lambda weights (the dual variables from the DEA linear programme which are shown in Table 4.9) 0.113,0.201, 0.216 and 0.426 of hospital 27 respectively to actual inputs of hospital 12,13,19 and 25.

This construction is done in Table 4.10, where Column z indicates that a combination of the actual operations of these four hospitals would result in a hypothetical hospital 27 that would use 0.797 fewer doctors, 7.173 fewer nurse, 2.128 fewer technicians, 1.086 fewer administrative staff, 8.244 fewer general staff, 5.875 fewer short term training and 17.367 fewer beds to produce the same amount of patient care / outputs as were inefficient hospitals 27.

Table 4.10: Comparison of Hospital 27 with its ERS Hospitals 12,13,19 and 25

	Hospital 12 (t)	**Hospital 13 (u)**	**Hospital 19(v)**	**Hospital 25 (w)**	**Composite (x) = t+u+v+w**	**Hospital 2(y)**	**Z = y-x**
Inputs							
Doctors	2^a x 0.113^b	3^a x 0.201^b	2^a x 0.436^b	2^a x $0{,}426^b$	2.203	3	0.797
Nurses	27^a x 0.113^b	19^a x 0.201^b	20^a x 0.426^b	20^a x 0.426^b	19.827	27	7.173
Technicians	5^a x 0.113^b	2^a x 0.201^b	6^a x 0.426^b	6^a x 0.426^b	5.872	8	2.128
Administrative staff	3^a x 0113^b	1^a x 0.201^b	2^a x 0.426^b	2^a x 0.426^b	1.914	3	1.086
General Staff	20^a x 0.113^b	9^a x 0.201^b	31^a x 0.426^b	31^a x 0.426^b	31	39.244	8.244
Beds	25^a x 0.113_b	23^a x 0.201^b	34^a x 0.426^b	34^a x 0.426^b	32.633	50	17.367
Short term training	20^a x 0.113^b	19^a x 0.201^b	31^a x 0.426^b	31^a x 0.426^b	23.125	29	5.875

Source: Author's Data Analysis Results, [2013]

The above is done to all inefficient hospitals (that is hospitals with pure TE scores of less than one), and Table 4.13 is formed which shows the amount of input reduction needed to make inefficient hospitals efficient.

Method 2: how to obtain the input reductions needed in each hospital (using input slacks)

An alternative method of obtaining the input reductions needed each hospital is to use pure TE scores and the input slacks from table 4.11. This method is derived from the works of Coelli (2005).

Recall the efficiency scores estimate the extent to which inputs would need to be reduced in equal proportions to reach the production frontier. In addition, for some hospitals, after inputs have been reduced proportionally, one input could still be reduced further without reducing output, these are what are referred to as slacks in the DEA literature.

Table 4.11 Summary of Input slacks

	TE (θ)	S^-_{X1}	S^-_{X2}	S^-_{X3}	S^-_{X4}	S^-_{X5}	S^-_{X6}
1	0.926	21.025	-	-	-	-	-
2	0.877	8.147	-	1.009	-	0.753	6.848
3	0.432	10.199	0.453	5.593	-	0.763	-
4	1.000	-	-	-	-	-	-
5	1.000	-	-	-	-	-	-
6	0.877	-	-	1.009	-	0.753	6.848
7	1.000	-	-	-	-	-	-
8	1.000	-	-	-	-	-	-
9	0.935	7.227	-	-	-	-	-
10	1.000	-	-	-	-	-	-
11	1.000	7.227	-	-	0.161	-	-
12	1.000	-	-	-	-	-	-
13	1.000	-	-	-	-	-	-
14	1.000	-	-	-	-	-	-
15	0.906	-	-	3.219	2.594	2.063	25.969
16	0.880	17.236	-	-	-	0.968	7.522
17	0.982	-	2.301	5.2989	10.935	-	32.852
18	1.000	-	-	-	-		
19	0.837	-					-
20	0.837	4.284	-	-	-	-	-
21	0.785	7.57					
22	1.000	-					
23	1.000	-					
24	1.000	-					
25	1.000	-					
26	1.000	-					
27	0.734	4.069	-	-	-	0.266	-
28	1.000	-	-	1.496	-		

Source: Author's Data Analysis Results [2013]

Table 4.11 above shows that all the hospital on the frontier, that is having pure TE scores of 1.000,do not have slacks in may inputs. This is because these efficient hospitals do not need any input reduction.

This will be illustrated using hospital 27 which has a TE score of 0.734 (73.4%), this means that DMU 27 is radially inefficient *in* input usage by a factor 73.4 % plus it has (non-radial) input slacks (S-) of 4.069 units of X1 and 0.266 units of X5 (see

4.11 Summary of Input Slacks

The targets of DMU 2(27) would therefore be to reduce usage of all its inputs by 0.266

(26.6 %) =1.000 -0.734, and also to further reduce the usage of x1 by 4.069 (S-X1: slack of input x2) and x3 by 0.266 (S-X4: slack of input X5). Table 4.12 gives us the result of the input reduction needed to make the inefficient hospital 27 efficient, which are the same results obtained in Table 4.10, where ERS was used instead.

Table 4.12 : Input Reductions using IS and TE scores

Hospital 2	Inputs (A)	(B) = (A) X 0.266	©= (B) + S-	
Beds	50	13.3	13.3 + 4.069	17.369
Doctors	3	0.798	0.798 + 0	0.793
Nurses	27	7.182	7.182 + 0	7.182
Technicians	8	2.128	2.128 + 0	2.128
Admin staff	3	0.798	0.798 + 0.266	1.086
General staff	31	8.246	8.246 + 0	8.246
Short term training	24	6.384	6.384+0	6.382

Source: Authors Data Analysis Results [2013]

Therefore by using either method 1 from Table 4.10 or method 2 from table 4.12, the input reductions needed in each of the inefficient hospitals, so as to make them efficient is summarized in Table 4.13

Table 4.13: Input Reductions Needed to Make Inefficient Hospitals Efficient.

Hospital	Excessive inputs based on DEA evaluation						
	Beds	**Doctors**	**Nurses**	**Technicians**	**Administrative Staff**	**General Staff**	**Short term training**
1	24.775	0.226	3.532	0.227	2.237	13.542	6.588
2	14.085	0.371	4.596	0.62	1.123	10.923	5.878
3	61.913	5	46.51	4.545	3.604	13.072	24.243
9	10.304	0.192	1.799	0.544	0.128	1.095	1.253
15	2.438	0.281	5.752	3.157	2.438	29.909	13.845
16	23.237	0.36	2.759	0.48	1.327	11.235	5.387
17	1.125	2.434	55.037	11.262	1.106	34.115	34.465
20	11.155	0.49	3.93	0.981	0.328	3.114	2.948
21	113.281	6.857	80.289	11.999	1.715	26.574	42.478
27	17.367	0.767	7.173	2.128	1.086	8.246	6.467
Total	279.68	16.978	211.377	35.943	14.092	153.748	144.046

Source: Author's Data Analysis Results [2013]

4.3.2 Findings related to SE

Recall that scale efficiency is the portion of overall technical efficiency, which is solely attributed to the divergences from optimal scale.

On the efficiency side, out of the 28 hospitals 13 (46.43 %) were scale efficient (that is are operating at the most productive scale) while 15 (53.57%) were scale inefficient, which means that the hospitals are operating at an inefficient size (bigger or smaller than optimal). Among the inefficient hospitals (53.57 %) the ranges of the SE scores were: 3.57% below 0.50,0 % between 0.51 and 0.60,3.57 % between 0.61 and 0.70, 28.57 % between 0.71 and 0.80,3.57 % between 0.81 and 0.90,14.29 % between 0.90 and 0.99

After examining Table 4.5 it became apparent that the major cause of overall technical inefficiency is scale inefficiency in North Rift Region hospitals. Therefore the next stage in the analysis is to explore the source of the scale inefficiency for each of the hospitals, this information is presented in Table 4.14, which shows 32.14 % of the 28 hospitals are operating in increasing returns to Scale (IRS), 46.43 % in constant returns to scale (CRS), and 21.43 % in decreasing return s to scale (DRS). Hence the 53.57 %=32.14 % + 21.43 % of scale inefficiency imply that hospitals need to adjust their capacity in order to enhance efficiency.

Table 4.14: Returns to Scale

	IRS	CRS	DRS
No. of hospitals	9	13	6
% no. of hospitals	32.14	46.43	21.43
Average beds	45.89	74.46	117.5

Source: Author's Data Analysis Results [2013]

As has been noted when efficiency scale is less than one, the implication is that the hospital is not operating at the most productive scale size for its observed input mix. In order to operate at the most productive scale size, a hospital exhibiting DRS should scale down both outputs and inputs. Similarly, if a hospital is showing IRS, it should expand both outputs and inputs.

All of the 9 hospitals that depicted IRS are public hospitals. Hence some consideration should be made as to whether these 9 public could be expanded. This would of course be influenced by other factors as well, like distance from other provincial or referral and population size.

All the hospitals depicting DRS are public hospitals, which implies that they need to be scaled down, hence some decision here too has to be made about their scale size.

4.6 Comparative Analysis of Hospitals:

4.6.1 Comparisions of Urban and Rural Hospitals:

The hospitals were classified according to their locations: urban area or rural area, as per classification of the population project 1990-2020, where urban areas in Kenya are defined as all settlements on state and those on tribal land with a population of 5000 or more persons with at least 75 % of labour force in non-agricultural occupations (subsistence farming). Therefore those classified as urban are; Nandi Hills District, Kapsabet District Hospital, Alwara Kapsabet Hospital, Elgon View Hospital, Uasin Gishu District Hospital, Medi Heal Hospital, Eldoret Hospital, Memorial Hospital, Moi Teaching and Referral Hospital, Moi's Bridge District Hospital, Kitale District Hospital, Mt. Elgon Hospital, Cherangany Nursing Home, Kitale Nursing Home, Iten Hospital, Mosoriot District Hospital, Kapsokwony District Hospital, Tambach Hospital, Kobujoi Mission Hospital. The rest not mentioned above are classified as rural hospitals.

Therefore applying this classification, 19 hospitals were found to be in the urban areas, while 9 hospitals were found to be in the rural areas. Table 4.15 summaries the efficiency sores of hospitals found in urban and rural areas. Urban hospitals on average have higher efficiency sores (both pure TE scores and scale TE scores) than the rural hospitals, this means that the urban are better performers in relation to both scale size and in usage of inputs, than the rural hospitals

Table 4.15: Efficiency Scores of Urban and Rural Hospitals

	Overall TE	**Pure TE**	**Scale TE**
Rural Hospitals (n=9)			
Mean	0.703556	0.90666	0.7706
S.D	0.22915	0.08857	0.21279
Min	0.303	0.734	0.335
Max	1.000	1.000	1.000
Hospitals on Frontier	2	3	2
Urban Hospitals (n = 19)			
Mean	0.91872	0.95441	0.93421
S.D	0.17318	0.13602	0.10832
Min	0.309	0.4324	0.717
Max	1.000	1.000	1.000
Hospitals on Frontier	11	15	11

Source: Author’s Data Analysis Results [2013

When compared, Table 4.16 and 4.17, which summarizes the descriptive statistics of urban and rural hospitals respectively. The descriptive statistics show that the urban hospitals have on average 90.52 beds (which range between 23-177 beds), while the rural hospitals have on average 40.66 beds. Since the number of beds usually represents the size of the hospital, the urban hospitals seem to be larger than the rural hospitals (which range between 25-50 bed

Table 4.16 Descriptive Statistics of Urban Hospitals (n = 19)

Variables	Mean	Standard deviation	Minimum value	Maximum Value
Maternity	1305.21	751.261	171	2588
General	2187.68	1140.5	464	4352
Tuberculoses	132.05	79.053	10	279
Out patients	21789.29	11359.380	4621	43346
Beds	90.52	57.687	23	177
Doctors	6.10.	4.556	2	22
Nurses	68.68	40.088	19	147
Technicians	9.10	5.714	2	19
Administrative	4.68	2.625	1	9
General Staff	38.15	25.347	8	93
Short term training	72.41	45.339	10	96.667

The urban hospital are on average more efficient than the rural hospitals, as was shown in table 4.15 were there average overall efficiency scores were 0.9872 and 0.7035 respectively. These findings are consistent to the findings in section 4.2 in graph 4.1 which also showed that large hospitals seem to be more efficient relative to t he smaller ones.

Table 4.17 Descriptive Statistics of Rural Hospitals (n =9)

Variables	Mean	Standard deviation	Minimum value	Maximum Value
Maternity	437.77	139.299	255	668
General	802.44	198.253	346	1007
Tuberculoses	76.66	47.457	19	157
Out-patient	7992.30	1974.59	3446	10029
Beds	40.66	10.136	25	50
Doctors	2.66	0.5	2	3
Nurses	24.55	3.940	17	29
Technicians	5.77	1.855	3	9
Administrative	2.88	0.781	2	4
General Staff	28.33	7.433	19	42
Short term training	21.397	4.836	14	29

Source: Author's Data Analysis Results [2013]

Note that this study is simply pointing out the differences in efficiency scores between urban and rural, before any conclusions can be drawn or decisions made, more analysis therefore has to be done to measure the relationships between efficiency scores and locations of hospitals.

4.6.2 Comparisons of Government and Non- Government Hospitals

It is often argued that the private sector is more efficient than the public sector in the production of health services, and that government reliance on private provision would help improve the efficiency and quality of public spending in health. The analysis in this study also found the same evidence, although in order to make this point a fact in the case of North Rift Region further analysis has been done on Relationship between the efficiency scores and ownership/management of the hospitals. Comparing the three categories from table 4.18, we see that the mission hospitals seem to be the best performers, with an average overall TE score of 0.904, which is followed by private and finally public, though the difference in average overall TE score between the public and private hospitals is very minimal.

Table 4.18: Efficiency Scores of Government and Non-Government Hospitals

	Overall TE	Pure TE	Scale TE
Private hospitals (n=9)			
Mean	0.8355	0.9401	0.8657
S.D	0.2117	0.0829	0.1815
Min	0.303	0.734	0.335
Max	1.000	1.000	1.000
Hospital on frontier	8	9	8
Mission hospitals (n=3)			
Mean	0.904	1.000	0.904
S.D	0.1294	0	0.1294
Min	0.752	1.000	0.752
Max	1.000	1.000	1.000
Hospital on frontier	2	5	2
Public hospitals (n=16)			
Mean	0.8272	0.8580	0.9292
S.D	0.3455	0.2840	0.1415
Min	0.309	0.432	0.717
Max	1.000	1.000	1.000
Hospital on frontier	3	3	3

Source: Author's Data Analysis Results [2013]

Under pure TE scores, the mission hospitals are the best performers with an average TE score of 1.000, meaning that they are producing the maximum output with their available inputs (are on the frontier), hence the private and government hospitals could learn a lot from the mission hospitals in terms of input use and output production. On the other hand under scale TE scores the public hospitals have the highest average score, 0.9292, hence when it comes to learning about right sizing the hospitals, the private and mission hospitals could learn from the public hospitals. Table 4.19, 4.20 and 4.21 show the descriptive statistics of public, mission and private hospitals respectively.

Table 4.19: Descriptive Statistics of Government Hospitals (n=16)

Variables	Mean	Standard deviation	Minimum value	Maximum Value
MAT	934.21	712.62	255	2588
GEN	1598.57	1224.84	346	4352
TB	92	52.02	19	194
Out –patient	15921.76	12199.41	3446	43346
Beds	151.8	43.02	25	177
Doctors	3.78	2.29	2	10
Nurses	44.31	35.80	17	147
Technicians	8	4.91	3	19
Administrative	3.73	2.13	1	8
General	31.73	16.93	8	72
Short term training	30.52	20.69	10.33	85.33

Source: Author's Data Analysis Results [2013]

The mission hospitals have on average 57, 68 beds, the government hospitals have on average 151.8 beds, and private hospitals have on average 57.25 beds, which depict the average sizes of the hospitals. Hence the government hospital seems to be larger on average than the private and mission hospitals.

Table 4.20 Descriptive Statistic of Mission Hospitals (n=3)

Variables	Mean	Standard deviation	Minimum value	Maximum Value
MAT	1 831.2	479.92	1105	2432
GEN	2734	317.03	2289	3148
TB	206	76.26	97	279
Out-patient	27230.6	3157.62	22798	31354
Beds	57.78	21.12	130	176
Doctors	10	6.74	6	22
Nurses	104.6	13.57	86	121
Technicians	9.8	6.37	2	18
Administrative	6.6	2.07	4	9
General	56	32.31	14	93
Short term training	62.3	20.02	37	87

Source: Author's Data Analysis Results [2013]

Table 4.21 Descriptive Statistics of Private Hospitals (n=9)

Variables	Mean	Standard deviation	Minimum value	Maximum Value
MAT	458.25	228.13	171	723
GEN	1186.25	705.60	464	2087
TB	105.25	92.07	10	197
Out patient	11815.05	7027.78	4621	20787
Beds	57.25	38.40	23	91
Doctors	4.5	2.38	3	8
Nurses	40.25	24.51	19	72
Technicians	6	4.241	2	11
Administrative	2.75	1.707	1	5
General	24	11.13	9	35
Short term training	25.8	14.656	11	42

Source: Author's Data Analysis Results [2013]

The mission hospitals were on average the best performers with an overall TE score of 0.904 which was shown in table 4.18, when one looks at the descriptive statistics of the government hospitals in table 4.20, it shows that the average beds range from 130-176, meaning that the government hospitals are mainly large hospitals, which is consistent with our earlier results that large hospitals in North Rift Region seem to be relatively more efficient than the smaller ones.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND POLICY IMPLICATION

5.1 Introduction

The study focused on conducting empirical analysis on the technical efficiency performance of 28 public, private and mission hospitals in North Rift Region, and of those inefficient hospitals what inputs and outputs contribute most to inefficiency. The empirical analysis shows the following observations.

5.2 Summary

The overall average level of TE among North Rift region is 83.7 % while Under pure technical efficiency 18 of the 28 hospitals (64.29 %) are technically efficient while 10 of the 28 hospitals (35.71 %) are relatively inefficient compared with the other hospitals in the data set. On the scale efficiency side, out of the 28 hospitals 13 (46.43 %) were scale inefficient (that is are operating at the most productive scale) while 15 (53.57 %) were scale inefficient, which means that hospitals are operating at an inefficient size (bigger or smaller than optimal)

The major contributory factor of inefficiency in North Rift Region hospitals is the scale size (scale inefficiency as opposed to pure technical inefficiency) This was shown in table 4.7 where the average pure technical efficiency score of the hospitals in the sample was 0.939, while the average scale efficiency score of the hospitals was 0.882.

The results also revealed that urban hospitals on average have higher efficiency scores (both pure TE scores and scale TE scores) than the rural hospitals, this means that the urban hospitals are better performers in relation to both scale size and in usage of inputs, than the rural hospitals. The mission hospitals are better performers than the private and public hospitals, with an average TE scores of 1.000, meaning that they are producing the maximum output with their available inputs (are on the frontier), hence the private and government hospitals could learn a lot from the mission hospitals in terms of input use and output production.

The total amounts of specific inputs that are currently wasted (the total amount of inputs which need to be reduced to make inefficient hospitals technical efficient) and not utilized in the production of hospitals in North Rift Region are:
16.97 doctors (12.04 % of the total number of doctors in the sample)211.37 nurses (13.58% of the total number of nurses in the sample)35.94 technicians (15.97 % of the total number of technicians in the s ample)(14.09 administrative staff (12.25 % of the total number of administrative staff in the sample)144.046 short term training (14.06% of the total number of the train personnel in the sample.153.74 general staff (16.44 % of the total number of general staff in the sample)279.68 BEDS (13.41 % of the total number of beds in the sample)

5.3 Conclusions

The study has showed that there is significant difference between technical efficiency levels and variations among selected hospitals.

As noted earlier, inefficient hospitals utilized larger of inputs, but even with their excess inputs inefficient hospitals produced less outputs than their relatively efficient counterparts. Efficiency analysis, like the one in these findings is only a guide to decision making, and ultimately the property rights over decision –making lies with the policy-makers at the national departments of health. However in deciding whether or not to take necessary measures to reduce the inefficiencies revealed by efficiency analysis, policy makers should take note that:

Inefficiency is a kin to a torn rice sack. If the holes are not identified and sealed / mended, it would be impossible to fill the sack. Like-wisely, unless inefficiencies are identified and eliminated, resources will keep on leaking out of the health care system. And thus reducing the extent to which health systems are able to achieve goals of health status improvement…inefficiency signifies the denial of additional citizens of opportunities to realize health improvements at zero cost. This what makes inefficiency both immoral and unethical?

5.4 Recommendations and Policy Implications

The study has therefore identified the efficient and inefficient of hospitals in North Rift Region and of those inefficient hospitals has managed to identify the peers. The study has also identified the specific input reductions required to make inefficient hospitals technically efficient. These inputs that need to be reduced can be termed as excess resources in the hospitals output production.

Therefore in connection with these excess resources that are currently wasted in the production of hospitals outputs, a number of policy implication are suggested to the health system managers as follows:

Options related to excess beds and space; Sell/rent the excess beds and space to the private health practitioners where there is demand and Re-allocate the excess beds to other health facilities or programs.

Private health practitioners, here refers to those operating private clinics in North Rift Region. The best option would be to re-allocate the excess beds to other health facilities /programs, and since HIV / AIDS is the most serious epidemic and hence highly prioritized, these excess beds could be re –allocated to HIV / AIDS programmes, such as the AIDS / STD / TB laboratories and STD training and research centres. The assumption here is that these laboratories / research centres are an independent entity, that is do not operate in or as part of any hospital. This will be in line with NDP mid-term review, were HIV / AIDS is snow being seen as the most critical issue in the country and therefore is highly prioritized.

Options related to excess doctors, nurses and technicians: Re-allocate the excess labour force to under staffed primary health care facilities and Re-allocate the excess labour to other health programmes (for example reproductive health programs, HIV / AIDS programmes, disease surveillance programmes, pulmonary tuberculosis programmes) / facilities.

The best option would be to re-allocate the labour to primary health care facilities (such as clinics), mainly because primary health care deals with health promotions, care and disease preventions, hence it would be more cost –effective to concentrate more on preventive care (primary health care: clinics, health posts) rather than curative care (secondary health care: hospitals). This would strengthen the essential promotive and preventive services, and as with the common phrase, “prevention is better than cure.” This will also be in line with NDP health strategy were one of government decision is to prioritize primary health care as opposed to secondary health care.

Options related to excess general staff; Re-allocate the excess labor to other health programmes and Re-allocate the excess labour force to other under staffed different departments of the Ministry of Health.

Since this is the general staff, they can be re-allocated to any other programme or department, which is experiencing shortage of supportive staff.

Options related to the scale size of hospitals: Conversion of hospitals into community health centres (contraction) and Conversion of primary hospitals into general hospitals (expansion)

These are very delicate options, for it entails the conversion of hospitals into community health centres / clinics where the hospitals were found to have decreasing returns to scale. And conversion of primary hospitals into general hospitals, were hospitals were found to have increasing returns to scale. If these options were to be pursued there would be a need for an intensive study in working out the details of the conversion process (which would have to take into consideration exogenous factors such as population size, distribution of the burden of disease and location.

Options related to ownership / management of hospitals: Introduce hospital autonomy and Performance evaluations and incentives that encourage directors and staff.

Firstly there will be a need here for further research on how ownership / management affect efficiency scores, by for example applying the models used by Wong (1996) where he used a two stage procedure (DEA and Econometric analysis) to assess the relationship between efficiency and ownership and location, refer to Chapter 2 section 2.41.

All in all hospitals that are less efficient should be re-organized to correct widespread inefficiency. We could expect that hospital productivity could be improved through performance evaluations and incentives that encourage directors and staff. Incentives include a variety of performance based personal benefits such as promotions or extended fringe benefits.Granting hospital autonomy is an option of making changes in management as given by NDP mid-term –review, where it involves granting hospitals a budgetary allocation and allowing them to decide on the most cost effective ways of using the funds.

5.7 Limitations of the Study

This study is limited by certain factors, firstly it was not ease to established whether the ERS facilities had exactly the same standards in terms of type of services provided, severity of cases treated, working schedules (number of work days, length of working days) and health care technology.

Secondly the analysis assumed that the case mix of specific DMU and its ERS DMUs are similar, it was not verified whether the assumption is plausible or not. However given that all the DMUs in the data set were all hospitals the above assumption is most likely hold.

Thirdly, there may be variations in quality of care from one health facility to another (for example those facilities offering higher quality of care may require more personnel time and other inputs than those offering low quality of care). Given that this study was based on secondary data, it was not be possible to determine whether there was any variance in quality of care across the facilities. Despite the limitation, sufficient required data was collected to address and meet data requirement for the study.

5.4 Areas for Future Research

There is a need for further research into the measurement of hospital efficiency in Kenya; this section will highlight some of these areas:

Analyzing of economic efficiency (both technical and a llocative efficiency) of the hospitals in Kenya.

Analyzing of economic efficiency (both technical and a llocative efficiency) of clinics in Kenya)

An analysis to determine the relationship between efficiency scores and hospital operating characteristics (for example differences in case mix, differences in quality of care)

An analysis to determine the relationship between efficiency scores and exogenous factors (for example regional variation s in HIV / AIDS epidemic, geographical locations, environmental conditions and ownership)

REFERENCES

Bloom G. and Lucas H. (1999) 'Health and Poverty in Sub-Sahara Africa ." IDS Working Paper 103.

Bitran R. and Blocks S. (2009) ' Provider Incentives and Productive Efficiency.' Health Financing and Sustainability Project, Abt Associates Inc. Bethesda, MD.

Britran R. (2009) 'Efficiency and Quality in the Public and Private Sectors in Senegal.' Health Policy Planning 10 (3):271-283.

Byrners P. and Valdmanis V. (1983) ' Analyzing Technical and Allocative Efficiency of Hospitals.' Duke University.

Bowlin W. (1999) 'Measuring Performance: An Introduction to Data Envelopment Analysis (DEA).' Department of Accounting, University of Northern Lowa Cedar Falls, 50614-0127.

Chang H. (1998)' Efficiency and Quality in the Public and Private Sector in Taiwan.' National Chengchi University, Taiwan.

Chillingerian J. (1994) ' Exploring Why Some Physicians' Hospital Practices are More Efficient: Taking DEA Inside the Hospital.' University of Texas, Austin.

Coelli T. (2005) a Guide to DEAP version 2.1: A Data Envelopment Analysis (Computer) Program.' Working Paper 8/96, Centre for Efficiency and Productivity Analysis (CEPA)

Dyson P. and Frank R. (2007) 'A Data Envelopment Analysis Tutorial.' Warwick Business School, UK.

Evans D.Tandon A. Murray C.and Lauer J. (2000) 'Measuring Overall Health System Performance for 191 Countries.' Discussion Paper no. 30, World Health Organization.

Frank G. and Taube C. (1987)' Technical and Allocative Efficiency in Production of Outpatient mental Health Clinic Services.' Journal of Social Science and medicine, Volume 24, No 10.

Hsiao W. (1993)'A Health Sector Study.' Paper Prepared for the World Bank, China Institute of Health Service Management.

Kiriga M.J. And Lambo E.and Sambo L. (2000) ' Are Public Hospitals in Kwazulu-Natal Province of South Africa Technically Efficient?' World Health Organization Regional Office for Africa.

Kiriga J. Emrouznejad A. and Sambo L. (2000) 'Measurement of Technical efficiency of Public Hospitals in Kenya: Using DEA Analysis Approach.' Warwick Business School, UK.

Maynard A. and Kanavos P. (2000) ' Health Economics: An Evolving Paradigm.' Department of Social Policy and LSE Health, London School of Economics.

Mooney J. (1995) 'Data Envelop Analysis: An Assessment of Performance.' Working Paper 4/95, Centre for Efficiency and productivity Analysis (CEPA)

Mugenda and Mugenda (1999), Research Methods: Quantitative and Qualitative Approaches, African Centre for Technology Studies (ACTS) Nairobi, Kenya.

Mukras, M.S (1993), Elementary Economics, Theory, Application and Policy, East African Educational Publishers Nairobi: Kenya.

Sear R. (1991) 'Comparison of Efficiency and Profitability of Investor-Owned Mult-Hospitals Systems with not-for Profit Hospitals.' Health Care Management review, Volume 16, No. 2.

Shearman H. (2006) ' Hospital Efficiency Measurement and Evaluation: Empirical Test of a new Technique.' Medical Care. Volume 22 No 10

Vitalliano D. (1987) 'On the Estimation of Hospital Cost Functions.' Journal of Health Economics, Volume 6 No 4

Wong S. (1996) ' Technical Efficiency in Korean Hospitals.' Institute of Economics and Sciences, Tropics University.

APPENDIX 1

Codes used in the measurement of Variables

Variables	**Measurements**
Outputs **Y_1** **Y_2** **Y3** **Y4**	 No. of maternity admissions No. of general admissions No. of tuberculosis admissions. No. of out – patients
Inputs X_1 X_2 X_3 X_4 X_5 X_6 x^7	No. of beds No. of full time doctors No. of full time nurses No. of full time technicians No. of full time administrative staff No. of fulltime general staff No. of short term training

APPENDIX II

Codes Used in the Measurement of Variables.

Variables	**Measurements**
Outputs	
Y_1	No. of maternity admissions
Y_2	No. of general admissions
Y_3	No. of tuberculosis admissions
Y^4	No. of out- patients
Inputs	
X_1	No. of beds.
X_2	No. of full time doctors.
X_3	No. of full time nurses.
X_4	No. of full time technicians.
X_5	No. of full time administrative staff.
X_6	No. of full time general staff.
X^7	No. of short term training

APPENDIX III: INTRODUCTION

Good morning /afternoon. My name is Clement Cheruiyot Tison I am a post graduate student of Moi University carrying out research on **TECHNICAL EFFICIENCY OF HOSPITALS: CASE OF PUBLIC, MISSION AND PRIVATE HOSPITAL IN NORTH RIFT REGION –KENYA**

The information request in the questioners is meant for academic purpose only and shall be treated in confidence. Kindly assist in filling the questionnaire. Please note that information given on the questionnaire will be held in strict confidence and will be used only for purpose of the study .In case any of the questions may not be appropriate to your circumstance, you are under no obligation to answer

Thank you.

APPENDIX V—RECOMMENDATION LETTER

MOI UNIVERSITY

SCHOOL OF BUSINESS AND ECONOMICS

Tel: (053) 43620
Fax No: (053) 43360
Telex No. 35047 MOIVARSITY

Box 3900
Eldoret
KENYA

Ref:MU/SBE/PGE/018/08

10TH February, 2010

TO WHOM IT MAY CONCERN:

Dear Sir/Madam,

RE: CLEMENT CHERUIYOT TISON: PGE/SBE/018/08
DATA COLLECTION FOR THE M.PHIL DEGREE DISSERTATION

The above named person is one of our Master of Philosophy (Economics) degree students currently in the field collecting data required for their dissertations The dissertation is an essential part of the said degree program and the student must pass this component of the program before he can be awarded the degree.

The purpose of writing to you this letter is to request you to assist the student as much as you can in his effort to collect the data that he needs. Below we have given the name of the student, his registration number and the title of his dissertation.

Please do accept our sincere thanks for your kind assistance.

Name of student: CLEMENT CHERUIYOT TISON
Registration number.
Title of dissertation: "EVALUATING TECHNICAL EFFICIENCY OF PUBLIC, MISSION AND PRIVATE HOSPITALS. THE CASE STUDY OF NORTH RIFT REGION-KENYA

Yours most sincerely,

M.S. Mukras

PROFESSOR M.S. MUKRAS

HEAD OF DEPARTMENT - ECONOMICS

APPENDIX VI—A UTHORITY TO COLLECT DATA

MINISTRY OF PUBLIC HEALTH & SANITATION

TELEGRAM "MEDICAL" NAKURU TEL.
(052)2062711/2031421 FAX (053)
2062711 *Email:*
mohnak@yahoo.com
When replying please Quote the

NK/POH/VOL.4/11

PROVINCIAL MEDICAL OFFICER OF HEALTH
P.O. BOX 2566
NAKURU

22nd February, 2010

CLEMENT CHERUIYOT TISON

Reg. No.SBE/PGE/018/08

RE: AUTHORITY TO COLLECT DATA

Your request to collect data for your M.PHIL degree dissertation from our health facilities and Medical records office is granted.

If further assistance at any points is necessary feel free to conduct my office

Dr Benedict Osore (pdms)

Provincial Medical Officer of Health

For.(PROVINCIAL MEDICAL OFFICER)
P.O BOX 2566,
NAKURU-20100

APPENDIX VII - FORMAL APPROVAL

INSTITUTIONAL RESEARCH AND ETHICS COMMITTEE (IREC)

MOI TEACHING AND REFERRAL HOSPITAL
P.O. BOX 3
ELDORET
Tel: 33471//2/3

MOI UNIVERSITY
SCHOOL OF MEDICINE
P.O. BOX 4606
ELDORET Tel:
33471/2/3

28* May, 2010

Reference: IREC/2010/45
Approval Number: 000546

Clement Cheruiyot Tison
P.O. Box 7074- 30100,
ELDORET, KENYA.

Dear Mr. Tison

RE: FORMAL APPROVAL

The Institutional Research and Ethics Committee has reviewed your research proposal titled:

"Evaluating technical efficiency of Public, Mission and Private Hospitals. The case study of North Rift Region-Kenya

Your proposal has been granted a Formal Approval Number: **FAN: IREC 000546** on 27th May, 2010. You are therefore permitted to begin your investigations.

Note that this approval is for 1 year; it will thus expire on 26th May, 2011. If it is necessary to continue with this research beyond the expiry date, a request for continuation should be made in writing to IREC Secretariat two months prior to the expiry date.

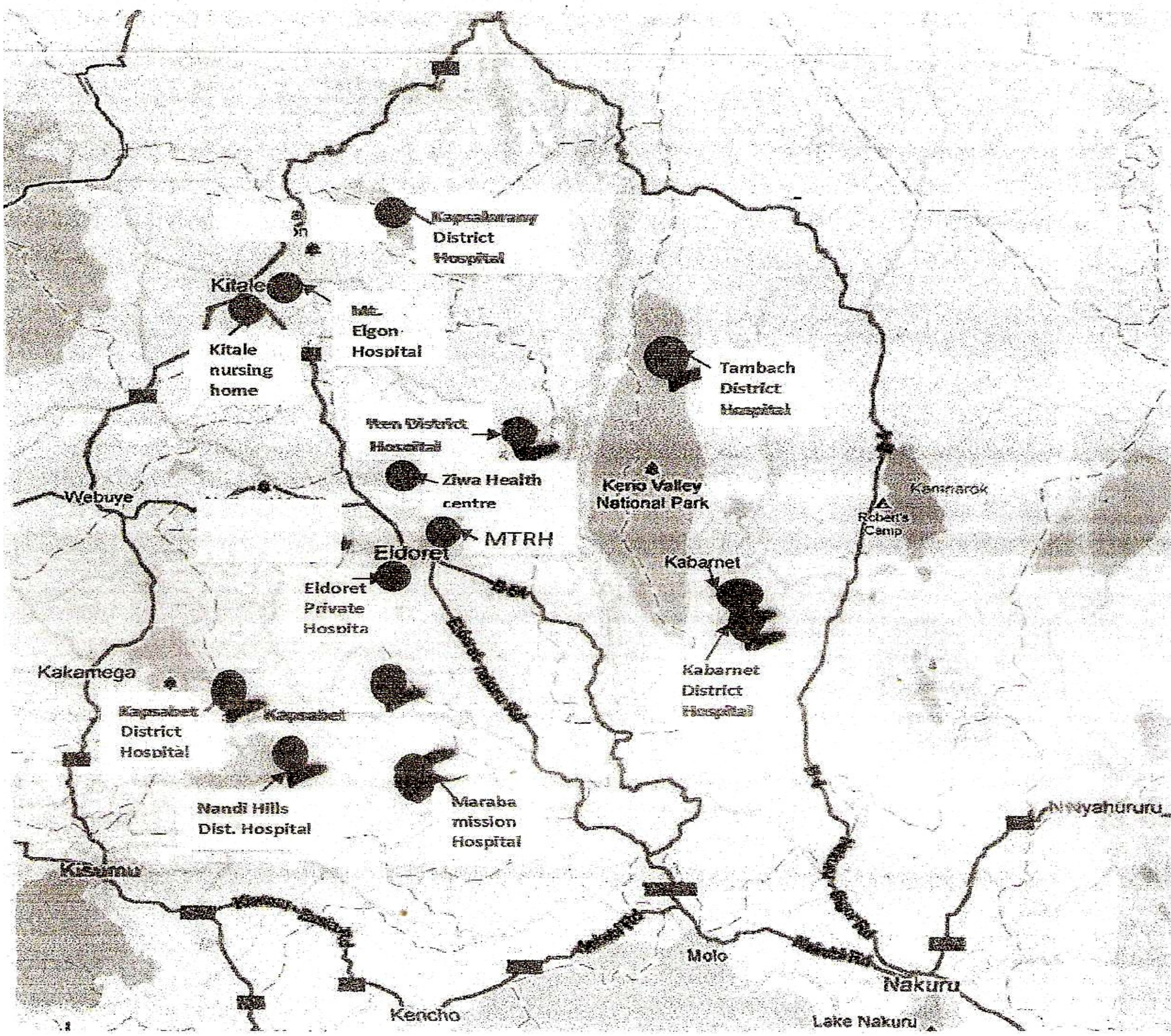
You are required to submit progress report(s) regularly as dictated by your proposal. Furthermore, you must notify the Committee of any proposal change (s) or amendment (s), serious or unexpected outcomes related to the conduct of the study, or study termination for any reason. The Committee expects to receive a final report at the end of the study.

Yours Sincerely,

DR. OMAR ALY
CHAIRMAN
INSTITUTIONAL RESEARCH AND ETHICS COMMITTEE

cc: Directer - MTRH
Dean - SOM
Dean - SPH
Dean - SOD

APPENDIX X –SAMPLE OF HOSPITALS IN NORTH RIFT REGION



SOURCE: ministry of health and sanitation (2010)

APPENDIX VIII - APPROVAL TO CONDUCT RESEARCH MTRH

MOI TEACHING AND REFERRAL HOSPITAL

Telephone:2033471/2/3/4
Fax: 61749
Email: director@mtrh.or.ke
Ref:
ELD/MTRH/R.6/VOL.II/2007

P.o Box 3
ELDORET

28th May, 2010

Clement C. Tison
P.O Box 7074-30100
ELDORET, KENYA

RE: APPROVAL TO CONDUCT RESEARCH AT MTRH

Upon obtaining approval from the Institutional Research and Ethics Committee

(IREC) to conduct your research proposal titled:

"Evaluating technical efficiency of Public, Mission and Private Hospitals. The case study of North Rift Region-Kenya

You are hereby permitted to commence your investigation at Moi Teaching and Referral

Hospital.

PROF. H.N.K. Arap MENGECH
DIRECTOR
MOI TEACHING AND REFERRAL HOSPITAL

CC - Deputy Director (CS)
- Chief Nurse
- HOD. HRISM