EFFICIENCY IN PUBLIC PROVINCIAL AND DISTRICT HOSPITALS IN KENYA. A STOCHASTIC COST FRONTIER APPROACH

BY

JOHN WANJOHI KARANI
C/50/7229/99

NOVEMBER, 2001
DECLARATION

This Research paper is my original work and has not been presented in another university.

20-11-2001
DATE
JOHN W. KARANI

This research paper has been submitted to the University of Nairobi for examination with our approval.

20.11.01
DATE
PROF. B. M. NGANDA

20.11.01
DATE
MR O AKETCH
ACKNOWLEDGMENTS

I want to acknowledge the reserved guidance from the supervisor Mr. Obilo Aketch and particularly to Prof. Nganda. Their critiques to my drafts are highly appreciated.

The financial support from the University of Nairobi made it possible for me to pursue a masters' degree in Economics, and on top of it the university gave me an allowance to survive on. Similarly I want to thank the African Economic Research Consortium (AERC) for giving me a thesis grant to help me in writing my thesis.

I am also grateful to Mr. Charles Muriuki Karani for his encouragement when I was down with swollen thyroid glands. I say thank you very much.

Finally I wish to record my gratitude for the continued support of all my colleagues, not forgetting my family.
DEDICATION

This dissertation is dedicated to my parents Mr. G. Karani and Mrs. Mary Karani.
# TABLE OF CONTENTS

1.0 CHAPTER ONE: INTRODUCTION  
1.1 Background and policy issues  
1.2 The public health system in Kenya  
1.3 Statement of the problem  
1.4 Objective of the study  
1.5 Justification of the study  

2.0 CHAPTER TWO: LITERATURE REVIEW  
2.1 Concepts of efficiency  
2.2 Efficiency evaluation models  
2.3 Hospital cost and production function  
2.4 Hospital output  
2.5 Stochastic cost frontier  
2.6 Critique of the literature review  

3.0 CHAPTER THREE: METHODOLOGY  
3.1 Theoretical framework  
3.2 Model specification  
3.3 Estimation method  
3.4 Variables and their definitions  
3.5 Data sources and types  

4.0 CHAPTER FOUR: RESULTS  
4.1 Correlation coefficient matrix  
4.2 Determinants of hospital costs  

5.0 SUMMARY AND POLICY IMPLICATION  
5.1 Policy recommendations  
5.3 Conclusions  

REFERENCES
LIST OF TABLES

TABLE 4.1 Correlation coefficient matrix 23

TABLE 4.2 Results from stata 25

TABLE 4.3 Maximum likelihood estimates results from frontier program 27
ABSTRACT

This thesis determines efficiency in the public provincial and district hospitals. The resources spent on these hospitals has to be justified. This is done by establishing the efficiency level of each hospital. In this way inefficiency levels are determined.

The main objective of this paper is to determine the level of efficiency for district hospitals and the provincial hospitals. Second objective is to determine the various inputs used in producing the health care and lastly the thesis to develops indicators for measuring the performance of the district and provincial hospitals in Kenya.

A stochastic cost function for hospital was estimated. The mean technical efficiency for provincial hospital is 64.4% while that of the district is about 54.2%. The efficiency level of the whole hospital system is 55%.

The stochastic cost frontier function indicates that there is potential to improve on quality of care without infecting additional resources in the hospitals. This is important given the financial constraint that the Ministry of Health is currently going through. Finally it is desirable to replicate this study to include the health centers so as to access the problems, its causes and possible efficiency savings.
CHAPTER ONE

1.0 INTRODUCTION

1.1 BACKGROUND AND POLICY ISSUES

The issue of efficiency in hospitals and other health facilities has received particular attention due to slowing down of economic growth. This has been caused by the dwindling of the resources available for the Ministry of Health (MoH) and by extension to the hospitals. The Gross Domestic Product (GDP) averaged 6.5% between 1964 and 1970. Since the mid 1980s the economy has deteriorated, due to a turbulent macroeconomic environment and financial resources have been insufficient to keep pace with the population growth. The AIDS pandemic, and the resurgence of other diseases (Collins et al, 1996), have brought about the need to use health resources more efficiently. This then requires the use of technically efficient methods of producing health care.

Efficiency in the health sector requires that the inputs should be effectively used to increase the net change in the desired output, e.g., an increase in the immunization of children. For technical efficiency (see Nganda, 1994 and Christopher, 1995) it is assumed that the methods under consideration are all effective. The decision has to be arrived in deciding the best way to do an activity, which has a net positive change on the desired output. Thus the adopted method should maximize the output with the least cost. This may involve changing the mix of personnel, drugs, physician time, patient's own time, and surgical procedure and so on. These are the technically efficient methods.

The lack of adequate funding and the poor use of resources meant that by the mid 1980s there were signs of deterioration of government health facilities. This included poorly paid and unmotivated staff, buildings in poor state of disrepair, broken equipment and shortage of drugs and other supplies (Collins et al 1996). To address these issues the government has to ensure that the resources in the health facilities are put into their
intended use. This study will establish the efficiency levels of the public health sector, but before then we have to look briefly at the organization of the public health system.

1.2 THE PUBLIC HEALTH SYSTEM IN KENYA

The Kenya health system is organized as a pyramid. At the apex being the referral and teaching hospital that is Kenyatta National Hospital. At the base of this pyramid are the health centres and dispensaries. The system was a creation of the colonialist, who concentrated on curative care. The organization of the pyramid is in four levels, which are central, provincial, district and the rural. The central level acts as the headquarters (that is the Ministry of Health (MoH) headquarters) from which the administrative and the professional policies are coordinated and the decisions made. The permanent secretary in the ministry of health heads the administrative services. The Director of medical services heads the professional services. The administration of hospitals happens to be the work of the central government.

The Provincial Health officer runs the provincial hospitals, while the District medical officer and the hospital assistant are in charge of the District hospitals and the sub district hospitals respectively. A qualified nurse heads the dispensaries.

1.3 STATEMENT OF THE PROBLEM

The hospitals are an important element of the concern about the use of the health resources. This is because they are the largest and the most costly operational units of the health system (Newbrander 1992). During this period of limited public resources, the way in which these resources are used becomes a matter of concern. The hospitals account for a large portion of the health sector financial, human and capital resources. The Ministry of Health uses about 70% of its recurrent expenditure on staff salaries (Collins et al, 1996 & Owino and Korir, 1997); this allocation has to be justified. In Kenya 60% of the physicians and 80% of the nurses are assigned to the hospitals (Bloom et al 1996). Is the Ministry of Health getting the required output with this labor force or could the same
output be produced with the less health personnel? This can be established by carrying out efficiency tests for the hospitals.

In a country like Kenya, the health system is labor intensive and requires qualified and experienced health personnel. There has to be a proper mix between the staff and the equipment. Multilateral and bilateral donors on the other hand, have been known to give different equipment for the same purpose. This may affect the least cost mix between equipment and personnel. In case of breakdown, servicing these equipment becomes a difficult task due to different specification of the same equipment. This results in under utilization of the equipment. This happens to be a source of inefficiency.

The gap between the available and the required resources in the MoH facilities is still increasing due to the demographic (population increase), social (refugee influx) and epidemiological (emergence of other diseases like AIDS) changes. We need to know if the resources or the inputs are used in the right mix that produces technically efficient outcomes with the least cost.

With the scarce resources in the health sector and the low ability to attract previous funding, there is need to determine efficiency levels.

1.4 OBJECTIVES OF THE STUDY

This research paper seeks to state the types of resources at the disposal of the health care providers and their adequacy. The other objectives include the following:

• To determine the level of efficiency for district hospitals and the provincial hospitals.

• Determine the various inputs used in producing health care and their effect on health.

• To develop indicators for measuring the performance (efficiency) of the district and the provincial hospitals.
1.5 JUSTIFICATION OF THE STUDY

In developing countries there is a concern about the cost of health care in light of scarce national and government resources. Hospitals warrant scrutiny due to the resources that can be generated from savings through efficiency gains. The gap between the resources available and the required resources is growing hence the need to establish the level of efficiency. The health needs have not diminished but they have increased. Hence, this paper is timely to determine the levels of inefficiency faced by these facilities and advice on the relevant policies to be adopted.

There has been similar studies done in Kenya. The study by Owino and Korir, (1997) uses a sample of 36 hospitals divided into four (4) provincial hospitals, twenty two (22) district hospitals and ten (10) sub district hospitals. They used data for two fiscal years. Similarly a study by Nzioya, M et al 2001 determined efficiency in 32 health centres in Taita-Taveta, Kajiado and Kiambu Districts. These studies left the categories of the district and the provincial hospitals as a whole.

This study exclusively deals with the public provincial and the district hospitals. Thus this study deviates from the other two studies in that all the provincial and the district hospitals are used in this study. This is because hospitals happen to get the bulk of the MoH recurrent expenditure, for curative care. The way these funds are used needs to be established in the face of dwindling allocations from the government. This then means that the hospitals should use the resources efficiently. Hence this study will establish this.
CHAPTER TWO

2.0 LITERATURE REVIEW

In this section we look at both the theoretical and the empirical literature. The section is organized as follows. Firstly the concepts of efficiency are reviewed, then hospital cost and production functions, hospital output and the stochastic cost frontier functions. Where applicable both the theoretical and the empirical literature are reviewed under each sub heading.

2.1 CONCEPTS OF EFFICIENCY

Efficiency is customarily addressed by the use of the question, whether health expenditures are higher than they need to be. Efficiency can be as maximizing the benefit to a society at large from the resources available or providing such a mix of effective services at least cost (Nzoya, M 2001). Now we look at the various inefficiencies.

2.1.1 Technical inefficiency

Technical inefficiency is concerned with the mix of inputs, which will produce a certain output. Alternatively given the number of services produced by the hospital, is it employing the least possible input that is, the health professionals, supplies and equipment?

Example of technical inefficiency may occur in drug use. These include extravagant prescribing (a less expensive or a generic drug could be used with comparable effects). Overprescribing, the drug is taken in too large doses or for too long a period. There is the incorrect prescribing that is the wrong drug or it is incorrectly prepared, and underprescribing.

Lack of equipment or malfunctioning ones, lack of laboratory reagents may mean that the hospitals just chance on the way of producing health care. According to Newbrander et al,
(1992) in the case of equipment, low levels of use and incorrect use of equipment are also a source of the technical inefficiencies. In the district hospitals the equipment may be under utilized or misused because it is inappropriate. There is also lack of uniformity and consistency in supply. The Ministry of Health may receive donations from bilateral assistance programs. Frequently the equipment received are a variety of the items from all over the world and with considerable different specifications. Acquiring spare parts and managing the maintenance of machines or equipment from different part of the world can be extremely difficult. Thus the MoH must manage its donors.

Technical inefficiencies may occur also in the area of staff. There are may staff inflexibility. This is more apparent in hospitals with low occupancy rates and a high ratio of staff per bed. Broadening the role of lower level staff may include more support for the primary health care activities (PHC), this could add to the productivity of both hospitals and the PHC operations (Barnum and Kutzin 1993). Deficient knowledge and deficient institutional arrangements may have the hospitals undertaking the PHC activities that can be effectively provided by the health centres and the dispensaries (WHO 2000).

2.1.2 Economic efficiency

This deals with the least cost combination of inputs, which produces the desired output given various possible combinations of inputs. Given the hospital budget is output - patient days, admissions, surgeries maximized (Barnum et al 1992).

The combination chosen to produce a certain quantity and quality of output will have lowest cost among the range of alternatives. The failure to use the economically efficient combination of output that relates to achieving the lowest cost combination of inputs is often not achieved.

The reason could be deficiencies in incentives, lack of the necessary training and skills, institutional constraints beyond the control of the hospital managers, shortages in the flow of the real resources and failures in information. Problems of economic inefficiency are
generally outside the control of the individual hospital managers. It is often at the levels of the MoH that the appropriate types and quantities of the facilities, staff and equipment must be answered.

2.1.3 Scale efficiency

Newbrander et al (1992) scale efficiency deals with, whether a health system as a whole is producing service (health care) at the least cost. This can lead to understanding the economies of scale. They are important not only for planning of the number of hospitals from a macro point of view but also for the individual hospital size. This has an impact on the operational cost and efficiency. In the short run the beds are fixed, and this limits the range of input combination for example staff, supplies, equipment, and buildings that can be used to produce the desired output. This limits the efficiency that the hospital can achieve in the short term. In the long run at the time when the factors can be varied. Thus as the hospital size increases there happens to be a greater combination of inputs to produce a given output. As the size increases, the fixed costs are spread over a greater number of patients, there is also a greater degree of specialization.

2.2 THE EFFICIENCY EVALUATION MODELS

RATIO ANALYSIS

Here we examine the three interrelated hospital services indicators. These are bed occupancy rate, average length of stay and the bed turn over. This methodology is used to assess the hospital performance based on the simultaneous analysis of the three statistics. This methodology was devised by Pabon Lasso 1986 quoted in Owino and Korir 1997. Each of the indicators may be defined on an annual basis and it can also refer to a particular ward, inpatient department, an entire hospital or group of hospitals. The variables are interrelated because, knowledge of any two defines the third variable.

A graphical technique using the indicators simultaneously is presented on how to examine the relative performance of the hospitals. The x-axis on the graph is the average
bed occupancy rate while the y-axis represents the annual bed turn over rate. This mathematical relationship among these three indicators of the hospital performance allows for a ray from the origin that passes through any point on the graph to represent a constant average length of stay. This increases monotonically from left to right and across the top and the right hand side of the graph.

The mean values of the bed occupancy and turn over rates (and by extension the mean value of average length of stay) divides the graph into four regions. Region 1 the first quadrant on the bottom left hand side shows hospitals with excess bed availability, low demand for hospital in relation to the installed capacity. Patients can be diverted to other institution. Hospitals in region two on the right hand side bottom quadrant have excess capacity and a high turn over rate. Hospitals in region three have high occupancy and a high turn over rates, while hospitals in region four have low turn over rate. One of the shortcomings of this method is that it is only used for descriptive purposes rather than for policy purposes. The technique doe not suggest specific policies to be addressed.

DATA ENVELOPMENT ANALYSIS

This is a linear programming technique for measuring the relative performance of institutional units with multiple inputs and outputs. These inputs and outputs make comparison difficult. Nzioya, M 2001 uses a DEA approach to estimate the efficiency of health centres in some districts in Kenya. This approach converts multiple inputs and outputs into a simple summary measure of productive efficiency, thereby determining the optimal input output combination.

In their study where, they were determining efficiency in some health centres, they found out that there are technical inefficiency as well as allocative inefficiency. Although wastages are in all health systems, their prevalence and magnitudes are more profound in the developing countries. This leads to inflation of the costs of the service delivery. The presence of inefficiency in the health centres means that these facilities have excess
inputs or insufficient output, in their study they found out that six of the health centre (out of the 32 health) are using more resources than the other 26 health centres. Their DEA results indicate that there is a potential to improve the quality of care without injecting additional resources into the health centres. This is important given the financial constraint of the Ministry of Health.

2.3 HOSPITAL COST FUNCTIONS AND PRODUCTION FUNCTIONS

A production function describes the technology of the firm (Varian, H 1992). It is usually satisfactory to think of inputs and outputs as a flow. That is a certain amount of inputs per time period are used to produce a certain amount of output per unit time period. In this case there is the short time period and the long run period. In the short time some of the factors of production are fixed while in the long term all factors of production vary.

A production plan is efficient if there is no way to produce more output with the same inputs or to produce the same output with the less inputs. There are Leontief, CES function and Cobb Douglas production technologies.

Batesse and Coelli (1992) proposed a stochastic frontier production function for panel, data which has firm level effects which are assumed to be distributed as truncated normal variables, which are also permitted to vary systematically with time. The model may be expressed as

$$Y_{it} = x_{it} \beta + (V_{it} - U_{it}) \quad \text{eqn (1)}$$

$$I = 1, \ldots, N \quad t = 1, \ldots, T$$

Where $Y_{it}$ is (the logarithm of) the output of the i-th firm in the t-th time period;

$x_{it}$ is a $k \times 1$ vector of transformation of the input quantities of the i-th firm in the i-th time period;
\[ V_{i} \] are the random variables assumed to be iid N(0, \( \sigma_v^2 \)), and independent of the

\[ U_{i} \] are the non-negative random variables which are assumed to be IID as truncated at zero of the N(\( \mu, \sigma_u^2 \)) distribution.

The above specification is in the form of a production function. According to Coelli (1994) the \( U_{i} \) is interpreted as the technical inefficiency effects, which causes the firm to operate below the stochastic production frontier. If we wish to specify a stochastic frontier cost function we simply alter the error term from \( (V_{i} - U_{i}) \) to \( (V_{i} + U_{i}) \). This substitution would transform the production function to

\[
Y_{i} = x_{i} \beta + (V_{i} + U_{i}) \quad \text{eqn}(2)
\]

\[ i=1, \ldots N, \]

Where \( Y_{i} \) is the (logarithm of the) cost of the product of the \( i \)-th firm;

\( x_{i} \) is a kx1 vector of (transformation of the) input prices and the output of the \( i \)-th firm;

\( \beta \) are the parameters to be estimated;

the \( V_{i} \) are the random variables which are assumed to be iid N(0, \( \sigma_v^2 \)), and

\( U_{i} \) which are non-negative random variables, which account for the cost efficiency in production, which are often assumed to be iid N(0, \( \sigma_u^2 \)). In this cost function the \( U_{i} \) defines how far a firm operates above the cost frontier. The above cost frontier is identical to the one proposed by Schmidt and Lovell (1979).

The measures of technical efficiency relative to the production frontier as specified in equation (1) and of the cost efficiency relative to the cost frontier specified in equation (2) are both defined as:
EFF_i = E(Y^*_i \mid U_i, X_i) / E(Y^*_i \mid U_i=0, X_i) \quad \text{eqn (3)}

Where the Y^*_i is the product of the I-th firm, which will be equal to Y_i when the dependent variable is in the original units and will be equal to exp (Y_i) when the dependent variable is in logs. In the case of the production frontier, EFF_i will take the a value between zero and one, while it will take a value between one and infinity in the cost function case (Battese and Coelli (1992).

On the other hand the cost function measures the minimum cost of producing a given level of output for some fixed factor prices. Just as the production function is the primary means of describing the technological possibility of production, the cost function describes the economic possibility of a firm or a hospital. Under the principle of duality it is possible to define a cost function from a production function.

According to Newbrander et al 1992, the production function describes the relationship between the output and the quantities and the combinations of inputs required to produce them. It indicates how resources can be combined to produce various levels of output in a technically efficient manner. In effect it describes the productivity of the inputs. The hospital is the unit of analysis and the production function specifies the minimum quantity of inputs required to achieve the given output or conversely, determine how output can be maximized with the given inputs.

Now we look at the various empirical literature, according to Feldstein, (1967) the current state of medical art or technology determines how these inputs are used to produce the case weighted output. Feldstein used a production function for the National Health Services hospitals in Britain. He specified hospital output as function of inputs consisting of the hospital beds medical staff, supplies, nursing and housekeeping, using a weighted case mix.
In a study of hospital function in Colombia, Barnum and Kutzin, (1993) estimated a non-linear cost function, logit regression. The independent variables included output that is the number of inpatient bed days and outpatient visits. The number of beds was used as proxy for capital stock. Secondary data was used for eight hospitals from 1975 – 1978. The results showed that the inpatient cost at the sample average gave almost equal marginal, average and constant short run returns to the variable factors.

In a study in Ethiopia by Bitran-Dicowsky and Dunlop 1989 (quoted in Owino and Korir), a flexible short run variable cost function was employed. The explanatory variables were inpatient days and outpatient visit, number of deliveries, laboratory tests, surgical operations and beds. Using pooled cross sectional and time series data from 15 hospitals, it was that marginal costs were slightly above average costs for inpatient bed-days and for deliveries and laboratory services, hospitals exhibited nearly constant short run returns to the variables factor inputs, and mild economies of scope existed between inpatient and outpatient services.

A study of Nigerian hospitals modeled health care provision by employing frontier production and cost function (Barnum and Kutzin, 1993). They used 24 health institutions in Ogun State. An efficiency index was obtained by firstly computing the marginal productivity of both health and non-health marginal productivity, and then comparing this with the ratio of their wages. The multi product log linear function estimated included, outpatient visit wages of both health and non-health worker, an index of drug availability as measure of quality, and the number of beds. The results indicated that outpatient visit, drug availability and wages are important. The ratio of marginal productivity of non-health workers to that of health workers was about two thirds of the ratio of their wages. The results may suggest that the facilities used most of the non-health workers than was economically efficient.
Anderson (1980) carried out an estimation of the scale efficiency of the health facilities in Kenya. A cost function was used for public hospital to measure scale efficiency. The study suffered from various limitations. The formulation adopted a variable to capture inefficiency, which was difficult to ascertain in case of hospitals.

In a study of 30 Chinese hospitals Howard and Kutzin (1993) found the extensive diseconomies of scale and mild economies of scope, short run inefficiency with respect to bed days and outpatient, and that low and middle level hospitals operated slightly below efficient levels of output, while larger hospitals operated above the efficient volume of output. The conclusion in this case is that hospital unit cost could be reduced by both using a large number of small institutions and reducing the average occupancy rates found to long stays, in which case, an increase in efficiency could be attained by reducing the average occupancy slightly. The study attributed long occupancy rates to long stay, in which case, an increase in efficiency could be attained by reducing the length of stay at the facilities. It should however be noted that although reductions in the length of stay is likely to increase the total cost per day, it could also reflect overall efficiency gains in terms of more patients being accommodated and, subsequently, reductions in cost per admission.

The study by Owino and Korir (1997) uses a hospital cost function. They estimate a short run cost function using a Cobb Douglas functional form. They adopted a cost function because it is more powerful in capturing the economic behaviour of the health facilities as opposed to a production function, which focuses on production technology. They used outpatient, admissions and operations in hospitals as the hospital output. This specification was situated for their needs. Their dependent variable was the short run costs. The explanatory variables were beds, average wage of the health professionals and output (admissions, operations and outpatient visit). They found beds, average wage and
their measure of output except admissions to be important. These variables played a role in reduction of technical inefficiencies.

From above the variability of the results indicate that caution should be used in estimating and interpreting cost functions. The country specific context in which the hospitals operate should be considered. The example from China shows great inefficiencies since high occupancy rates were maintained despite marginal cost exceeding average costs. This was due to extremely long length of stay. Thus a reduction in the length of stay was recommended to expand capacity and also lower unit cost.

2.4 HOSPITAL OUTPUT

Basically there has been no encompassing definition of hospital output. There is a controversy as to what should constitute hospital output. Hospitals lack homogeneity in their output. This is because of the different disease (case mix) treated in the hospitals.

For homogeneity in defining output the following is used.

The Diagnosis related Groups (DRG), which is basically a construction based on case types, and can be used as a measure of output. One of the difficulties is estimating a production function with all the cases.

The decision as to what constitutes hospital output can be classified into four categories. These are the inpatient treatment, outpatient treatment, teaching and research. The following can be used in measuring the hospital output. These are the effectiveness of a treatment of an episode of illness, severity of treated cases and inpatient days (Feldstein 1987).
Hospital output has been classified as internal medicine, surgery, gynecology, pediatrics, intensive care and other. However this classification does not exhaust classification of hospital output. This then happens to be a limitation to this approach (Wagstaff 1989).

Barnum and Kutzin (1993) consider inpatient admissions and outpatient visits to estimate a two-product function in a study. This study was about the cost and efficiency of public and private health care facilities in Ogun State in Nigeria. In this case the output was found to be inversely related to the cost of seeking health care. Assuming a production possibility frontier increase in production of inpatient days reduces the number of outpatient visits.

Barnum and Kutzin (1993) estimated a non-linear cost function. The output they specified the number of the inpatient days and the outpatient visits to a health facilities. Anderson (1980) in a study in Kenya to Determine efficiency of the hospitals the output was similar to that used by Owino and Korir.

Vita, 1990 used various units to measure hospital output. These have included the effectiveness of treatment of an episode, its severity and the inpatient days.

2.4 STOCHASTIC COST FRONTIER FUNCTIONS

A frontier can be defined as giving maximal output, which can be attained, given the set of input quantities. This maximum can be firms in the sample under investigation or with respect to all firms that can be conceived to exist (Forsund et al 1980).

Cost frontier functions were first used by, Farrel (1975). He provided definitions and a computational framework for analyzing both the technical and allocative efficiency of firms. Aigner and Chu followed in 1977 with a Cobb Douglas specification. The observations are required to be on or below the production frontier, while in the case of
the stochastic cost function this shows the cost lying above the cost frontier hence they are cost inefficient. The amounts by which a firm lies below its production function is and by the amount by which it lies above its cost frontier, can be regarded as a measure of inefficiency. Under duality conditions, a non-linear functional form of production function is assumed to be sufficiently tractable to permit derivation of cost and input demand frontier.

2.5 CRITIQUE OF THE LITERATURE REVIEW

The literature on hospital production function and cost function brings out clearly the relationship in the transformation of the process of producing health care. However there are still controversies as to what is hospital output. Each of study defines different variables as output. For example the study by Barnum and Kutzin (1993) defined the output as inpatient days and outpatient days, while Owino and Korir on Kenya defined output as admissions, inpatient days and outpatient days. There happens to be some similarity between admissions and inpatient days. When one is admitted in the hospitals they become an inpatient, and the days one spends there are the inpatient days. Hence I think Owino and Korir erred and should have used one of the two variables. In this study output will be measured using outpatient visit.

A stochastic cost frontier approach will be used to estimate the efficiency levels of the hospitals in Kenya. However it is possible to investigate frontier models without the use of explicit one-sided error terms. But in this case a cost frontier function will be used due to the assumption of the error term, which is composed of two terms. This is the specification of the frontier model used in this study.
CHAPTER THREE

3.0 METHODOLOGY

INTRODUCTION

The examination of cost functions has previously used unit cost functions, which were not theory driven. The average hospital cost functions were considered as a function of interrelated variables such as bed occupancy rates, patients' flow and length of stay (Pabon Lasso, 1986). These cost functions formulations were not based on theory but defined for convenience (Barnum and Kutzin 1993).

Recently there has been a distinction between the short run, the long run cost functions, the production and the cost functions. The functional forms have been specified unlike in the past, and the variables are consistent with the theory have been used.

This distinction is important for measuring the economies of scale, which is a long run concept. The distinction occurs when these functions are related to the specification of time period and the inclusion of a scale or capital proxy as an independent variable.

3.1 THEORETICAL FRAMEWORK

The concepts of the production and the cost function provide a framework for consideration of allocative decisions in the hospitals. In this case we are going to use the cost function. It expresses the relationship between the hospital's costs and output levels. There exists an intimate link between the measurement of firm level efficiency and the estimation of the production function and cost frontier. Here we consider a public health facility using inputs \( x = (x_1, \ldots, x_n) \) to produce a single output \( y \) (health services) that can be sold at a fixed price (user charges) where the price is greater than zero \( (P > 0) \). Efficient transformation of inputs into output is characterized by the production function \( f(x) \), which shows the maximum output that is obtainable from the various input vectors.
Under the duality condition, an equivalent representation of an efficient production technology is provided by a cost function:

\[ C(y, w) = \min \{ w, x : f(x) \geq y, x \geq 0 \}, \]

Which represents the minimum expenditures required in producing output “y” at price “w” which is the expenditure.

3.2 MODEL SPECIFICATION

We have estimated the cost function captured in a Cobb Douglas functional form. The model is specified as follows:

\[ \ln \text{Cost} = \beta_0 + \sum \alpha_i \ln x_i + v-u \] eqn (4)

The above equation is the general equation as put forward by Aigner et al (1977)

where \( \beta \) are parameters to be estimated

Cost = the total operational expenditure in a hospital.

The \( x_i \) is a vector of inputs prices and may be denoted as;

\( P_{Lab} = \text{labor}, \) is the proxy of the price of labour for health the professionals.

\( P_{Util} = \text{measures the utility cost like telephone, water, electricity.} \)

\( \text{Outpat} = \text{number of visits made to a hospital by an outpatient.} \)

\( \text{Cap} = \text{these are capital items bought in a certain year, for example equipment} \)

\( \text{Dss} = \text{drugs and supplies} \)

\( \text{Qs} = \text{quality related variables for example purchase of paints, cleaning materials.} \)
Expanding equation (4) we get

$$\ln \text{cost} = \alpha_0 + \alpha_1 \ln P_{\text{lab}} + \alpha_2 \ln P_{\text{util}} + \alpha_3 \ln \text{outpat} + \alpha_4 D_{\text{ss}} + \alpha_5 \ln \text{cap} + \alpha_6 \ln q_{s} + \varepsilon \ldots$$

eqn (5)

The $\varepsilon$ the error term, $\varepsilon$ has two components: $v_1$ and $u_1$;

where $v_1$ is a symmetric component representing random factors and captures variations in output of the sector due to factors outside the control of the health facilities; and $u_i<0$ is a one sided error component representing technical inefficiencies.

The economic logic behind this specification, where the error-term is composed of 2 term is that that the production process is subject to two economically distinguishable random disturbances with different characteristics. The non-positive disturbance $u_i$ reflects the fact that each hospital's output must lie on or below its production frontier. Any such deviation is the result of factors under the hospital's control such as technical and economic inefficiency. The error term $u_i$ is derived from $N(0,\sigma^2_u)$ distribution truncated above zero. $V_i$ represents the factors, which are not under control of the health facilities e.g. an outbreak of diseases in a certain area, amount allocated for the recurrent expenditure in a certain year.

Using the properties of the cost function then, we expect that the cost should be non-decreasing in the factor prices (Varian, 1992). As the price of labor increases, the hospital costs are expected to rise. Thus the Mol I operational expenditure will rise; this is due to the fact that the factor prices are determined exogenously. As hospital costs of production rise then hospital output is expected to decrease for example on the length of stay in hospitals, is expected to fall.
This shows the relationship between capital cost and the short run cost is positively related. This means that with the purchase of capital goods the costs of producing outpatient days will be high and output will fall. Increase in the utility cost may help in producing the required hospital output. Thus a positive relationship is expected between the output and the utility costs.

3.3 ESTIMATION METHOD

The estimation of the above equations requires the econometrics package known as Stata and Frontier 4.1. We assume that $v_i$ is i.i.d. $N(0, s^2_v)$, $u_i$ is i.i.d. $N(0, s^2_u)$ where $u_i$ and $v_i$ are independent of each other as well as of the independent variables. Based on the above distributional assumptions, the model can be estimated using the maximum likelihood estimation technique.

3.4 VARIABLES AND THEIR DEFINITIONS

(i) Cost of producing the health output (Lcost)

This is the dependent variable, which is captured by the total operational cost for the respective hospitals.

(ii) Labor costs (Iplabs). This is the staff emolument and other allowances, used in producing hospital output, that is outpatient visit.

(iii) Outpatient visit (Ioutpat). In this case it is also classified as hospital output. These are the number of days an individual visits a health facility.

(iv) Capital cost (Icaps). This includes the items that are used in producing health care or the output. This includes drug, hospital linen, new equipment and so on.

(v) Utility costs (Iutil). This includes electricity, water, telephone and so on.

(vi) Drugs and supplies (Idss). These include the drugs given to the patients and other items used in treating them for example bandages and so on.
(vii) Quality costs (lqs). These include the purchase of cleaning detergents, paints and the labor used in the cleaning of the health facilities.

3.5: DATA SOURCE AND TYPES

This study used secondary data, which consisted of yearly observations on price of labor for health professionals, utility cost, capital cost, drugs and supplies, output and quality costs. The data was from the Government Appropriation Accounts for the year 1996/97. Statistical Abstracts and Economic Surveys were also used as sources.
4.0 RESULTS

4.1 CORRELATION COEFFICIENT MATRIX

This section discusses the descriptive results. This study evaluated the efficiency of the district hospitals and the provincial hospitals in Kenya. Previous studies used a sample of some of the mentioned hospitals. This evaluation was by estimating a stochastic cost frontier, using the Stata and the Frontier 4.1 econometric software. The estimation was carried out in two phases by first finding the OLS estimates (using Stata) and then finding the maximum likelihood estimates and the efficiency level (using the Frontier 4.1) of each health facility.

A correlation matrix was estimated using Stata. The following table has the results of the correlation matrix from.

Table 4.1: CORRELATION MATRIX

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.3417</td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.4791</td>
<td>0.3056</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.1033</td>
<td>0.2121</td>
<td>0.1137</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>0.6763</td>
<td>0.2050</td>
<td>0.3023</td>
<td>0.2915</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>0.5549</td>
<td>0.4678</td>
<td>0.2407</td>
<td>0.0293</td>
<td>0.4030</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>0.7369</td>
<td>0.2696</td>
<td>0.3279</td>
<td>0.1157</td>
<td>0.4550</td>
<td>0.3998</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

(1) Cost of hospital output
(2) Capital items
(3) Cost of utilities
(4) Hospital output
(5) Cost of labor
(6) Quality inputs
(7) Drugs and supplies

The greatest correlation coefficient is between drugs and supplies and the recurrent expenditure, which is 0.7369. The lowest correlation coefficient is .1137. This is between the outpatient visits and the utilities used by the hospitals.

4.2 HOSPITAL EFFICIENCY LEVELS

The following are the estimates from Frontier 4.1. Program. The average technical efficiency for the provincial hospitals was found out to be 64.45% with Nakuru provincial hospitals catering to the Rift Valley Province having a technical efficiency of 89%. The provincial hospitals have a mean efficiency of 64.4% with an inefficiency level of around 35%. These results show that there is room for producing the optimal output more efficiently.

The technical efficiency mean for the district hospitals is 54.9%. There are six hospitals within an efficiency level of less than 20%. These hospitals waste 80% of their resources. There can be enormous efficiency gain in these health facility if wastage is reduced.

4.3 DETERMINANTS OF HOSPITAL COSTS

The results from the Stata output are presented below. This is done to find out if the results conform to our a priori expectations. The hospital operational expenditure is used as the dependent variable.
Table 4.2 Results from *STATA*

| Source  | SS      | df | MS        |
|---------+---------|----|-----------|
| Model   | 37.980559 | 6  | 6.33009317 |
| Residual| 10.7330486 | 45 | 238512192  |
| Total   | 48.7136077 | 51 | 955168778  |

Number of obs = 52
F( 6,  45) = 26.54
Prob > F   = 0.0000
R-squared = 0.7797
Adj R-squared = 0.7503
Root MSE   = 48838

| lcost | Coef. | Std. Err. | t    | P>|t|   | [95% Conf. Interval] |
|-------|-------|-----------|------|-------|----------------------|
| llab     | .4308816 | .0554354 | 7.773| 0.000 | .319229 .5425343     |
| tcp      | .0108778 | .0740746 | 0.147| 0.884 | -0.1383161 .1600718  |
| lutil    | .1587936 | .1128803 | 1.407| 0.166 | -.068559 .3861462    |
| lqs      | .0740727 | .1064195 | 0.696| 0.490 | -.1402671 .2884125   |
| ldss     | .3867835 | .1274986 | 3.034| 0.004 | .1299882 .6435789    |
| lout     | -.0434446 | .0208267 | -2.086| 0.283| -.1240114 .0371222   |
| _cons    | 1.05971  | 1.650659  | 0.642| 0.524 | -2.264888 4.384308   |

From the above results we see that the hospital cost decreases as hospital output increases. The outpatient visit to a particular hospital may be occasioned by diarrhea, tuberculosis, coughs and so on. The coefficient of the outpatient visit is -0.04344, this shows that an increase in 1% in inpatient visit will lead to a decrease in cost of production by -0.04%. This is because cost of producing health care (operational cost) will be spread on more outpatient visits. This is suggestive of existence of economies of scale in the hospital sub sector.
Using the properties of the cost function, the costs are non-decreasing in input prices, then we expect a positive correlation between the inputs used in producing health output, which is the case.

The labor costs for the health professionals are important inputs in producing health output in this case it is the outpatient visit. The elasticity of cost with respect to labor costs is 0.4308. This implies that an increase in the labour costs by 1% would lead to an increase in cost of producing health output (operational expenditure) by 0.43%. Hence the labor costs and the operational expenditure of a particular health facility move in the same direction, because there is a positive relationship. For example, with an increase in salaries of the health professional, the input prices for labour will increase. For a particular level of output the patient will thus have to pay more for the hospital service. Thus, the productivity of the health professionals needs to be ascertained.

The health professionals equipment should complement their work, that is, the work of producing health output (outpatient visits). Without equipment, producing health output would be a daunting task. These are some of the capital outlays a hospital engages in. The elasticity of the cost with respect to equipment is 0.01087; this shows that an increase in 1% in capital expenditure would lead to an increase in the cost of producing health output by 0.0187%. Capital expenditure in hospital also includes the contracted doctors.

Drugs and supplies are also very important inputs in producing health output. This is evident from the coefficient of the variable, which is the 0.3867. This means that an increase in 1% in quantity of drugs and supplies will lead to a 0.38% increase in the costs of producing health output. This variable is also highly significant with a t-ratio of 3.034. Availability of drugs in government hospitals would lead to higher utilization rates in the sense that patients would not be turned away for lack of drugs. Reduction in pilferage and spoilage of drugs would lead to increase in technical efficiency. This would affect the way the inputs are combined.
Quality costs are the other inputs used in producing health output and building repairs. This includes such items such as equipment maintenance, cleaning items, and cleaning detergents. These add to the perceived quality of care by the patients (who are the consumers of health care). They need a clean environment from which to seek their treatment. The coefficient of the of the quality costs is 0.0740 which means that an increase in 1% of the quality costs will lead to 0.0747% in the operational expenditure of the hospital.

Utility costs arise from things such as telephone charges, water, and so on. The coefficient of the utility costs is 0.1588. This then shows that an increase in 1% in the utility costs would lead to 0.1588% increase in the cost of producing health output.

The labor and the drugs and supplies are the major determinants of hospitals costs. These are the most visible inputs to the patient. Lack of these inputs in the required proportions would increase the waiting time in the form of long queues. This then would be a sign of technical inefficiency.

Table 4.3 Maximum likelihood estimates (MLE) Results from the Frontier program

<table>
<thead>
<tr>
<th></th>
<th>coefficient</th>
<th>standard-errors</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta 0</td>
<td>0.030003690</td>
<td>0.026472431</td>
<td>0.11356685</td>
</tr>
<tr>
<td>beta 1</td>
<td>0.049879476</td>
<td>0.023725482</td>
<td>2.10235897</td>
</tr>
<tr>
<td>beta 2</td>
<td>0.010852064</td>
<td>0.00350366</td>
<td>3.097345365</td>
</tr>
<tr>
<td>beta 3</td>
<td>0.050576291</td>
<td>0.16526423</td>
<td>0.30603290</td>
</tr>
<tr>
<td>beta 4</td>
<td>0.76498732</td>
<td>0.250981162</td>
<td>3.047987001</td>
</tr>
<tr>
<td>beta 5</td>
<td>0.014912429</td>
<td>0.00373120</td>
<td>3.99035289</td>
</tr>
<tr>
<td>beta 6</td>
<td>-0.052732341</td>
<td>-0.191367302</td>
<td>2.75555647</td>
</tr>
<tr>
<td>sigma-squared</td>
<td>0.5698658</td>
<td>0.10787405</td>
<td>0.052824248</td>
</tr>
</tbody>
</table>
The above results are from Frontier 4.1 software. The betas represent the parameters that were estimated. In this case,

Beta 0 represents the coefficient of the constant term, beta 1 coefficient of the labor costs, beta 2 represents the coefficient of capital costs of the various hospitals, beta 3 represents the coefficient utility costs, beta 4 represents the coefficient quality costs, beta 5 represents the drugs and supplies and beta 6 represents the coefficient outpatient visits to the hospitals.

The maximum likelihood estimates are conforming to our apriori expectations. In that all other variables are positively related to the cost of producing health care apart from the outpatient visits which has a negative relationship. This implies that with a high cost of producing health care some costs may be passed on the patients and this may reduce their use of these facilities.

The mean efficiency of the provincial hospitals is about 64.4%, while that of the district hospitals is about 54.21%. The scale efficiency for the system is 54.91% almost 55%. This shows that there can be improvements in terms of the amount of output produced. This shows that there are technical inefficiencies in these hospitals. The presence of inefficiencies means that the hospitals are using more inputs than they need to produce the current level of output or less than optimal output with the same resources. The district hospitals would have to reduce their inputs by 30%.

There are many factors that can influence the efficiency level of the health facilities. Mostly, these factors are exogenous and thus out of control of the health managers of the health facilities.
A possible cause of inefficiencies is mismanagement of resources. For example poor storage of drugs, pilferage, low staff productivity (due to low staff morale), shortages in health personnel, poor storage facilities, and variation in resources available.

4.4 LIMITATION OF THE STUDY

This study used secondary data and was therefore subject to measurement errors often present in such data. In addition to this, the other limitation arises from the assumption was to assume that all hospitals treat the same case mix, i.e. the type of patient treated, was the same.
CHAPTER FIVE

5.1 SUMMARY AND POLICY RECOMMENDATIONS

This research paper has examined the types of resources at the disposal of the health care providers and their adequacy and the level of efficiency at which they are used. The other objectives include the following:

This study also evaluated the efficiency of the public hospitals, specifically, the districts and the provincial hospitals. The explanatory variable represents the cost of producing health output (in this case the outpatient visits).

Labor inputs are a major determinant of hospital costs. The hospitals (both the district and the provincial) are labour intensive and thus labor is a major cost to the facilities.

Considering the capital costs which includes the cost of contracting additional doctors, the purchase of equipment, we find that total hospitals cost is also positively correlated with this variable.

Drugs and supplies have a positive relationship with cost of producing health care. A 1% increase in the purchase of drugs will lead to a 0.38% in the cost of producing the health output or health care. Thus if the drugs are not always available, this can lead to a reduction of the number of outpatient visit to a hospital. This would be a sign of allocative inefficiency.

5.2 POLICY IMPLICATION AND RECOMMENDATION

The finding from this study can be used to provide guidance, in improving policy-making decisions and their implementation in the area of improving efficiency in the public district and provincial hospitals. From the regression results, the main determinants of the
costs of producing the health output (operational expenditure of the hospitals) are the labour cost of the health professionals, the capital costs, the utility costs, quality costs, utility costs and the drugs and supplies costs.

The first policy recommendation is that, improved allocation of resources require that there be analysis of economic data concerning hospitals and an understanding of the underlying economic concepts, processes and interaction of activities within hospitals. Thus, the government should employ hospital managers who are familiar with the national budget process and the political context. These hospital managers should be in a position of producing services to input to outputs. Hence they will be in a better position to reduce wastage of resources in the hospitals. Inefficiencies in hospitals are expensive because they reduce the resources available, and hinder improvements in health status.

As far as labour costs are concerned, the doctors and nurses are hired are centrally by the Public Service Commission and posted to the various hospitals by the Ministry of Health. The health professionals are quick to leave the hospitals to their private clinics. This then reduces the hospital managers to discipline the health professionals. Some of the health professionals advice the patients in the public hospitals to seek treatment from their clinics. This has the effect of reducing productivity of the doctors or any other medical personnel concerned and hence increasing technical efficiency. The other observation is that the recommendations of the health facilities manager have to go up the hierarchical system up to the MoH headquarters. The hospital managers should be given some degree of autonomy in hiring and punishing the errant health professionals. For equitable distribution of doctors, those working in the rural area should be given some allowances, which their counterparts in the urban areas don’t receive.

The Ministry of Health needs to improve the environment in which the health care providers work, since this can influence their behavior and consequently the technical
efficiency. Payment and the working environment may affect the work morale and by extension the productivity and efficiency.

As far as capital costs are concerned, there may be lack of trained technicians in hospitals. They may be lacking the skills to use certain equipment or repairing them in case of breakdowns. Donors contribute equipment, which may be inappropriate, and do not include proper staff training, the equipment cannot be maintained due to lack of availability of spares, incompatibility with the existing equipment and high cost of maintenance due to high sourcing costs of spare parts. The recommendation here is that the Moll should manage its donors. This means accepting only those equipment that are compatible with the existing ones.

Drugs and supplies are very important inputs in the production of health services. The government should make sure that all the hospitals are supplied with the essential drugs first, that is hose drugs that are in the essential drug list. This will ensure that patients are not turned away due to lack of drugs and other medical supplies. The government should put into place internal control to avoid pilferage, spoilage of drugs and so forth. The Moll should control the inappropriate distributions, over supplies and under supplies at the facility level, check the issue of expired drugs still in the stores. There should also be proper co-ordination between the Moll and the donors' supplies to ensure constant supplies of the medicines and other supplies. These drugs and other supplies should be supplied on time, for example in time of disease out break.

Quality cost may include stopping or reducing the wearing down of the hospital physical infrastructure. The physical capital (buildings, equipment and so on) deteriorates in a more visible way. This includes hospital-building repairs. The state of the building could be a source of inefficiency. This may occur in the sense of reducing space due to dilapidated building, which may leak at times. This then makes the patient to turn away
from the public hospitals to the private hospitals. The government should allocate more money for repairing the hospital building and cleaning and so on. In this case quality includes cleanliness of the facility and so on.

5.2 CONCLUSIONS

This study shows that there is room for improvement in producing health care in the country. The scale efficiency is 54.91%. This shows that the health system is operating at a low level efficiency. Thus there should be proper monitoring of staff, good incentives for competitive service provision.

This study used secondary data for both the district and the provincial hospitals. Regressions were run to find out the efficiency levels for each hospital. The mean efficiency for the district is 54.9% while that of the provincial hospitals is 64.4%.

Allocation decisions must be made within the health sector and more narrowly within the hospitals and alternatively the health centres. For example within the health sector the Ministry of health may be faced with a choice of allocating a 5% increase in the budget between hospitals and primary health care. The Ministry may decide to use that the best use of the additional resources would be to allocate entire increase for the PHC services since this is more likely to reduce the incidence of morbidity than if the resources were provided to the hospitals.

Alternatively, the Ministry may be faced with a choice of scale in construction in the hospital sub sector or a choice of skill mix within the hospitals. If a number of nurses will soon be available, the Ministry may need to decide the number of nurses to be placed in the health centres compared to the hospitals. This decision can be implemented on the basis of the soundness of its allocative efficiency. Allocative efficiency is achieved if after deciding the number of staff to be assigned to the hospitals and the health centres the Ministry cannot increase output any further by reallocating the nurses any differently.
A production function is closely related to the cost function. Cost functions can establish the relationship between hospital cost and size or economies of scale. If the objective is to minimize the operating unit costs the concept of economies of scale can be used in for planning the facilities in the health system. The reasons for the returns to scale are division of labor as well as technical factors associated with large hospitals.

Diseconomies of scale are due to limitations in efficient management caused by the size of the facility, which makes control and coordination much more difficult.

As for the capital expenditure, the hospitals are found to receive a much smaller share of total capital expenditure. Their share is so much less than the recurrent expenditure because a much higher percentage was noted in the 1969-70s (Newbrander and Kutzin, 1992) when many projects were initiated through foreign donors. Donors currently have been funding primary and preventive health care (P&PHC) capital projects thus tipping the resource scale in favor PHC.
REFERENCES


Creese, A.L (1991), Health Policy and Planning; *user charges for health care: a review of recent experiences*


Ngugi, R (1999), Health Seeking Behaviour in the Reform process for rural household: The case of Mwea division, Kirinyaga District, AERC research paper no. 95


World Bank (1987); *Better Health in Africa; Experiences and Lessons Learned*; Washington D.C; World Bank.

World Bank (1987); *Financing Health Services in Developing Countries; An agenda for reforms*, Washington, D.C; World Bank.