The Pakistan Development Review
36 : 4 Part II (Winter 1997) pp. 669—693

Can Cost-effective Reallocation of Inputs Increase the Efficiency of the Public Health System in Pakistan?

M. AYNUL HASAN, HAFIZ A. PASHA, and AJAZ M. RASHEED

Heavy investment in many developing countries in the social sector including health is based on the premise that human capital is vital to the growth and development of a nation. However, Pakistan's spending on this sector has been one of the lowest in the region. In the present environment of high budget deficits, one does not expect substantial public funds to be forthcoming and diverted towards the social sector in the intermediate- or medium-term future. The critical issue facing the public sector should then be to design health policies which must be cost-effective and efficient. This study examines these health policy issues within the context of an optimisation framework for public health system, forecasts future upto (2002-03) and discusses an efficient optimal mix of health inputs, outputs, expenditures, and wage policies under alternative scenarios. The study recommends that, first, growth of health infrastructure building in the urban areas be slowed down in the short-term (two to three years), and some of the resources reallocated towards the rural sector either in terms of building new Basic Health Units or upgrading the existing Rural Health Centres. Second, not only attractive wage policies be formulated for health personnel, but the status of nurses in the public health system be also elevated by giving them higher grades. Third, for every rupee of development expenditure incurred, Public Health Department must plan or keep provisions for recurring outlays. All this reallocation of resources is feasible within the projected actual budget and it will lead to efficiency gains in the order of 8 to 10 percent for the entire public health system.

1. INTRODUCTION

In recent years, many developing countries have invested heavily on the social sector including basic health. This is based on the premise that human capital is vital to the growth and development of a nation. Therefore, keeping the mass healthy is as important as providing them with basic education. Pakistan has had an impressive GDP growth rate of about 8 percent per annum in 1991-92, out of which only a meagre 0.7 percent was spent on the health sector. When this figure is translated in

M. Aynul Hasan is Professor of Economics, Acadia University, Wolfville, Nova Scotia, Canada. Hafiz A. Pasha was Deputy Chairman of the Planning Commission, Government of Pakistan, Islamabad. Ajaz M. Rasheed is Senior Systems Analyst, Social Policy and Development Centre, Karachi.

Authors' Note: This study was initiated Aynul Hasan was visiting the Social Policy and Development Centre, Karachi as Senior Technical Adviser on behalf of the Canadian International Development Agency (CIDA). He wishes to thank the Agency for providing financial support for the study. Research assistance and primary data collection were conducted by Zahid Hasnain, Naveed Hanif, Nooreen Mujahid, and Nasrul Eman. Any errors are the responsibilities of the authors.

monetary value it amounts to only 7 cents per thousand dollar GNP spent on the health sector. This amount is very little by any standard and, in fact, the picture is even more dismal when this figure is compared with those of other developing countries.

Table 1 shows the percentage of total government expenditure on the health sector in relation to government expenditure and GNP for selected developing countries. Of the eight countries selected for comparison, Bangladesh and Sri Lanka appeared to have spent 4.8 percent of their government expenditure on health as opposed to only 1 percent by Pakistan. Even a small, poor country like Nepal spends more money (4.7 percent) than Pakistan on the health sector. In the list of countries considered, Pakistan's standing in terms of spending on health (either as a proportion of government expenditure or GNP) is the lowest which is very discouraging and disappointing. Although, Pakistan's spending for this sector is one of the lowest in the region, in the present environment of high budget deficits faced by the country, the critical issue facing the public sector should then pertain to designing *health policies* which must be *cost-effective* and *efficient*. A recent study by the World Bank on Pakistan's Health sector 1991) noted:

> *... inefficiency is widely regarded as a central problem in the health and population sectors in Pakistan... (especially) in the public sector. Resources are misallocated, in part because no investigation has typically been made into the cost-effectiveness of various options. Political influence and leakage of equipment and supplies further distort public sector allocations. Poor management and centralised financial, administrative and management authority reduce the efficiency of facility-level staff services.*

Table 1

Expenditure on Health for Selected Developing Countries

Country	% of Total Govt. Expenditure on Health	% of GNP Spent on Health
Pakistan	1	0.002
Bangladesh	4.8	0.007
Nepal	4.7	0.009
India	1.6	0.003
Sri Lanka	4.8	0.014
Indonesia	2.4	0.005
Egypt	2.8	0.011
Philippines	4.2	0.008

Source: *World Development Report, 1991-92.*

The major policy debate, in this context, facing the public health system in Pakistan, should then address the problem of re-allocating the limited resources among inputs in the most economical way. In other words, given the limited available budget, should the public health department (*PHD*) go for a more infrastructure programme (development expenditure) or hire more health personnel (recurring expenditure) or a combination of both so that there is a real improvement in the health services? Or, even more importantly, what type of wage policy should it adopt if more personnel are needed to work in rural health centres (*RHCs*), basic health units (*BHUs*) and hospitals, especially when there are shortages of competent nurses and doctors in the country?

Furthermore, if, historically or even in the future, there is more emphasis on development expenditure i.e., rapid expansion of *BHUs* or *RHCs* especially under the recently initiated *Social Action Programme* (*SAP*),[1] what implications will this programme have on recurring budget? Will the present health system in the country be able or capable of handling such an accelerated expansion of facilities? That is, given the present institutionally fixed wages and inelastic supply of health personnel, will the health system be able to attract enough doctors or nurses? All these public policy issues are critical for the *sustainability* of the basic health programme in Pakistan and they cannot be addressed in isolation. In this context, a general optimisation framework for the *Public Health System* (PHS) is needed if the *efficient cost-effective* input-output linkages consistent with the resource (budget) constraints and institutionally fixed wage policies are to be established.

In view of the above considerations, therefore, the objective of this study is:

(a) To develop a general optimisation framework in order to address the policy issue of *cost-effectiveness* and *efficiency* for the public health system;
(b) based on the above theoretical framework, to identify and forecast an optimal mix of *cost-effective* inputs (doctors, nurses, paramedics), outputs (outdoor and indoor patients) wages and expenditure requirements (both recurring and development) of *PHS* covering up to the end of the *Perspective Plan* period, 2002-2003;[2] and

[1]In response to donor agencies' insistence, the Government of Pakistan recently initiated an accelerated *Social Action Programme* (SAP), whereby the policy-makers have committed to revamp public expenditures in certain key areas namely, primary education, basic health, public health, population planning etc.

[2]In order to undertake long-run planning, Government of Pakistan prepared a fifteen-year perspective plan covering a period between 1988-89 to 2002-2003, wherein it stipulated targets for all key economic and social variables including basic health.

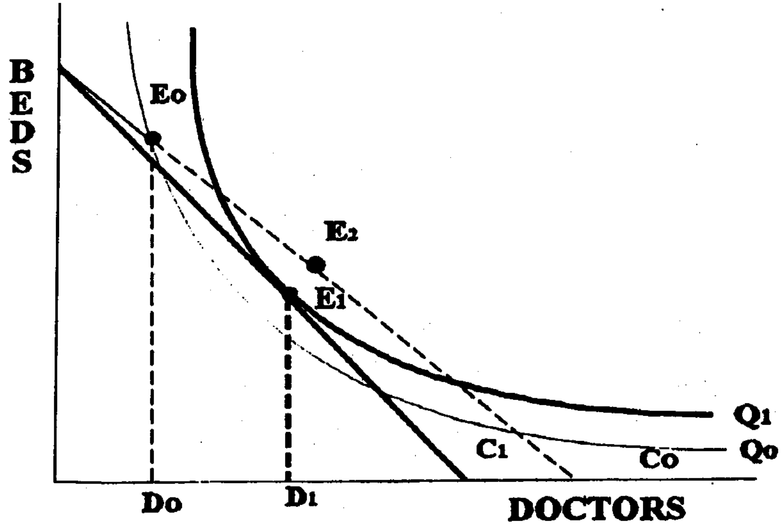
(c) to estimate and compare the extent of *efficiency gains/loss* for *PHS* under alternative assumptions.

The organisation of rest of the chapter is as follows. Section 2 develops a simple optimisation model for the public health system and then formulates propositions under alternative assumptions (e.g., constrained strategy etc.). In order get an intuitive insight, we first illustrate the model using a diagram and then in Section 2.1 and 2.2 the analytical derivations and conditions of the model to achieve positive efficiency gains for health facilities are presented. Section 3 presents the discussion on simulation results and ensuing policy implications. Before discussing the simulation results, we also report parameter estimates for input supply and production functions for Punjab in Sections 3.1.1. The discussion on *ex-ante* simulation results are presented in Sections 3.1.2 while the policy conclusions are given in Section 4.

2. DIAGRAMMATIC APPROACH OF A SIMPLE OPTIMISATION PUBLIC HEALTH MODEL

In the following, we develop a theoretical optimal allocation model for the health sector under institutionally fixed wage rates for health professionals. In order to keep the analysis simple and, at the same time, realistic, we assume that the health facility produces output [e.g., total patients treated (*Q*)] using two inputs, namely, infrastructure [represented by beds (*B*)] and medical professionals [e.g., doctors (*D*)]. In addition, the supply of personnel is assumed to be inelastic as there does not exist an unlimited inflow of doctors and, especially at the institutionally set low wage rates, many of these health professional are reluctant to offer their services to the public health system. The public sector is assumed to be able to set the wage rates of doctors and other government personnel (supposedly below the market rate) because their share (in terms of total expenditure) within the health sector is very large. Thus, the public sector, represented by the provincial health department (*PHD*) in this case, can be labelled as a *monopsonist*. Due to limited public funds allocated to the health and other social sectors, the basic task of *PHD* (acting as a *monopsonist*) in this context is to maximise the output of the health facility subject to the available budgeted resources (*TC*). Assuming a well behaved production technology, *c* as the unit cost of a hospital bed (*B*) and as the fixed wage rate for doctors, the optimal health input allocation problem of *PHD* can be illustrated with the help of a diagram as given in Figure 1.

FIGURE 1
OPTIMAL ALLOCATION OF HEALTH INPUTS WITH INSTITUTIONAL FIXED WAGE RATES



Essentially, the diagram consists of isoquant (Q) and isocost ($^{\bullet}$) curves for various combination of *Bs* and *Ds*. With institutionally fixed wages for doctors and unit price per beds (c), the initial isocost line is given by C_0 with slope (=-w/c) in Figure 1. It is, however, important to note that **not all points** (particularly E_0, on the broken portion of C_0) on this isocost are feasible. Given the inelastic supply of doctors, the number of professionals offering their services will simply not **exceed** D_0. Therefore, the initial equilibrium in this case will be established at E_0 as shown in Figure 1, which is neither *efficient* nor *optimal* even though the available budgeted resources are completely utilised. Here obviously, more beds and fewer doctors are employed to produce an *inefficient* Q_0 level of output [i.e., patients treated].

On the other hand, however if PHD adopts a more *flexible wage policy* and allows the wage rate of doctors to increase from $^{\bullet}$ to w_1, then the availability of doctors will increase from D_o to D_1. It is interesting to note that even though the *isocost* line has pivoted inwards (due to higher wages) form C_0 to C_1, the output (patients treated) however, has increased from Q_0 to Q_1. In fact, the new equilibrium level attained at E_1 is at a higher *isoquant* which is both *optimal* as well as c*ost effective*.[3] The above simple example leads us to make the following proposition that:

[3]It should be noted that E_2 on the old *isocost* line C_0 is also an equilibrium point for a fixed monosonistic wage rate which is both *efficient* and at a higher output level. However, to attain such a level of output, the producer (PHD) must be able to exert enough monopsony power to hire the required level of doctors. Here we argue that such a situation will be difficult if not impossible to achieve particularly when there are alternatives available to superior qualified health professionals within (in the private sector) and outside the country.

Proposition 1: *It is possible for the public health policy-maker to be cost efficient and, at the same time, achieve a higher output (in terms of more patients treated) even with the same available budget provided the public sector adopts a flexible wage policy (close to the market wage) for health professionals (be it doctors, nurses or paramedics).*

2.1 Optimisation Conditions under a Monoposonistic Public Health System

It was argued earlier that the *monopsonistic* equilibrium output, such as E_2 in Figure 1, will not be feasible under an institutionally fixed wage system. However, in order to explore alternatives (other than flexible wages) under which the *monopsonist PHD* may achieve equilibrium at a higher output level, we explicitly need to derive the *first order conditions* (*FOCs*) of the present health system. By comparing these *FOCs* with those of the standard *competitive* producer we may be able to establish the extent of divergence between them and thus propose the condition(s) which will lead to an improvement in the output for *PHD.*

Assuming the supply of doctors as a function of wages [D=$D(w)$] and output [patients treated (Q)] being produced by doctors and beds, the optimisation problem of the monoposonistic *PHD* will simply entail the maximisation of the following:

$$\text{Max:} \quad Q = Q\lfloor D(w); B \rfloor; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (1)$$

$$\text{subject to:} \quad \overline{C} = wD(w) + (m+c)B; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (2)$$

where m is the unit recurring cost per bed. Other variables in the above model are as defined earlier. Writing the *lagrangian* function of the above, we get:

$$\wp(w, B, \lambda) = Q|D(w); B| + \lambda|C_0 - wD(w) - (m+c) \quad \ldots \quad \ldots \qquad (3)$$

where λ is the *Lagrangian* multiplier. The maximisation of the above *Lagrangian* will yield the following three first order conditions (*FOC*):

$$\frac{\partial \wp}{\partial w} = \frac{\partial Q}{\partial D} \bullet \frac{\partial D}{\partial w} - \lambda \left[w\frac{\partial D}{\partial w} + D \right] = 0; \quad \ldots \quad \ldots \quad \ldots \qquad (4)$$

$$\frac{\partial \wp}{\partial B} = \frac{\partial Q}{\partial B} - \lambda(c+m) = 0; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (5)$$

$$\frac{\partial \wp}{\partial \lambda} = \overline{C} - wD - (c+m)B = 0; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (6)$$

From the first two *FOCs*, we have

$$\frac{\frac{\partial Q}{\partial D}}{\frac{\partial Q}{\partial B}}=\frac{w\frac{\partial D}{\partial w}+D}{(c+m)\frac{\partial D}{\partial w}}=\frac{w}{(c+m)}\left[1+\frac{D}{w\frac{\partial D}{\partial w}}\right] \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

If we define the marginal rate of technical substitution (*MRTS*) between doctors and beds as:

$$MRTS_{D,B}=\frac{\partial Q/\partial D}{\partial Q/\partial B}; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

and the elasticity of supply of doctors with respect to wages ( ) as:

$$\xi_D^S=\frac{w}{D}\bullet\frac{\partial D}{\partial w}; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

then the above optimisation problem of the *monopsonistic PHD* will require that

$$MRTS_{D,B}=\frac{w}{(c+m)}\left[1+\frac{1}{\xi_D^S}\right]. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

The above optimisation condition simply states that, for the *monopsonistic PHD*, $MRTS_{D,B}$ should not only be equated to the ratio of relative prices [*w/Cc+m)*], as is the case for the standard competitive producer, but it should now be equated to a weighted relative prices of doctors and beds. The weight factor simply includes the supply elasticity of doctors ( ). It is interesting to note that, depending on the magnitude of ( ), the above optimisation condition for *monopsony* can be equivalent to that of the competitive producer. This leads us to make our second proposition that:

Proposition 2: *The optimal condition (s) for the monopsonistic producer (PHD) can be equivalent to those of the competitive producer as long as the price (wage rate) elasticity of supply of the monopsonistic input (doctors) is infinitely elastic.*

Whether or not the above proposition is valid in the case of *PHD* is a matter of empirical investigation and we relegate this to the next section where we report our results on estimated elasticities for inputs.

2.2 Efficiency Gain/Loss for the Monopsonistic PHD Producer

Based on a simple model constructed earlier we can also analytically compute and compare the *efficiency* gain/loss (in terms of output) of the *monopsonistic PHD* under alternative scenarios. Suppose that, at a given point in time, the actual allocation of inputs by *PHD* is such that the size of the infrastructure in the form of beds (B) is greater than that of the optimal *monopsonist.* In the following, we can demonstrate that halting or reducing the growth of infrastructure building and optimally reallocating the resources towards health professionals may lead to an improvement in the health facilities. In fact, what is more interesting now is that the *monopsonist PHD* with fixed beds (named as *constrained monopsonit*) will be able to achieve an even higher level of output as compared to the standard monopsonist.

If we define O_c^* and O_s^*, respectively, as the optimal level of outputs for the *constrained* and *standard monopsonist*, then the expression for efficiency *gain/loss* (E) can be written as:

$$E = \left[\frac{O_c^*}{O_u^*} - 1\right] \bullet \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \qquad (11)$$

Obviously, a positive value for E will lead to an efficiency gain in favour of the *constrained monopsonist* and *vice-versa.* By substituting the optimal values for O_c^* and O_s^* (obtained from the optimisation model) in the above equation, we can obtain an explicit expression for E.

If the supply elasticity and the share of doctors (π^D) and nurses ((π^B) are assumed to be fixed in magnitude as a simplifying assumption and the production function is assumed to be *Cobb-Douglas*, then the two optimal outputs *(standard* and *constrained)* and their corresponding input demand functions obtained through the optimisation procedure can be written as :[4]

Standard Monopsonist

$$O_s^* = A_o (D_s^*)^{\alpha} (B_s^*)^{\beta}; \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \qquad (12)$$

$$B_s^* = \frac{\Pi^B}{(m+c)} (\overline{C} + cB_{-1}) \; ; \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \qquad (13)$$

$$D_s^* = \left[\Pi^D (\overline{C} + cB_{-1})\right]^{\frac{Y}{1+Y}} \; ; \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \qquad (14)$$

[4]Note that the case of constrained monopsonist, since the number of beds are fixed at $\overline{B}$ (B_{-1}), the budget constraint will now be changed to $\bullet = wD + m\overline{B}$.

Constrained Monopsonist

$$O_c^* = A_o(D_c^*)^{\alpha}(\overline{B})^{\beta}; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (15)$$

$$D_c^* = (\overline{C} - m\overline{B})^{\frac{\gamma}{1+\gamma}}; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (16)$$

where γ is the supply elasticity of doctors and α and β, respectively, are the output elasticities of doctors and beds. Substituting these optimals, we get the following explicit expression for E which contains entities that are either exogenous [e.g., · and $\overline{B}$] or fixed in values [e.g., share of doctors (π^D), m, and c]:

$$E = \left[\frac{(\overline{C} - m\overline{B})}{\Pi_D(\overline{C} + cB_{-1})}\right]^{\alpha} \bullet \left(\frac{\overline{B}}{B^*}\right)^{\beta}; \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (17)$$

Thus, for E to be positive (implying positive *efficiency* gains), the above equation should be written as:

$$\left(\frac{wD_c^{\ *}}{wD_s^{\ *}}\right) \rangle \left(\frac{B_s^{\ *}}{\overline{B}}\right)^{\frac{\beta}{\alpha}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (18)$$

This leads us to make our third proposition that:

Proposition 3: *Constrained monopsonist PHD with fixed infrastructure (beds) may have positive efficiency gains (E>0) relative to those of a standard monopsonist as long as the fixed number of beds (B) exceeds the optimal (B_s^*) and, at the same time, the total allocation of expenditure on doctors by the constrained monopsonist is greater than that of the standard case.*

These three propositions have important implications for public health policies and they are summarised below:

(a) In general, given limited budgeted resources available, a proper ***cost effective*** allocation between ***recurring*** (health professional) and ***development*** expenditures (infrastructure) can lead to higher efficiency gains for health facilities.

(b) In the case where *PHD* acts as a *monopsonist* and there is an inelastic supply of health professionals, adopting a ***flexible wage policy*** rather than institutionally set wages may not only attract more of these professionals but it is also possible to attain a higher output for the health system which will be both ***cost effective*** and ***efficient***.

(c) In the even that the existing allocation of ***infrastructure*** (beds) exceeds the optimal level, than a ***policy of consolidating*** the infrastructure and diverting the limited available resources towards recurring outlays can lead to ***efficiency gains*** which, in fact, will **exceed** those of the optimal cost effective producer.

In order to test these theoretical propositions and also to obtain the size and magnitude of these efficiency gains for *PHD*, next section, as an example, develops an empirical simulation model based on combined provincial data for Pakistan.

3. EMPIRICAL RESULTS FOR PUBLIC SECTOR HEALTH

This section reports simulation forecast results, both historical and *ex-ante*, based on optimisation strategies subject to available resources for the public health system in Pakistan. Estimated optimal mix of expenditures on inputs in this context will be *efficient* and also *cost effective*. Thus, a comparison of the actual available data on health output with the forecasted values will enable us to compute the *efficiency gain/loss* for alternative optimal health policies. In addition, the simulation forecast results will provide us with an opportunity to test the validity of the propositions made in the previous section.

Long-term forecasts generated by the model covering up to the end of perspective plan period 2002-03 will be useful for policy-makers in establishing the optimal requirements of physical health inputs (e.g., doctors, nurses, paramedics and beds), expenditure allocations (both development and recurring) and the wage policy under alternative health strategies.

3.1. Simulation Results for the Public Health System

In this section, we report estimated simulation results based on an extended model developed for the public health system. The basic parameters (e.g., elasticities, output share, etc.) used for simulation purposes, in this study, were estimated employing health sector data on Punjab[5]. Since there is a plethora of numbers generated by the simulation model, in order to describe these numbers more effectively, only the broad policy oriented results are discussed in the main text.

[5]A discussion on data sources, definition of variables, and ensuing problems and anomalies are reported in a technical *Appendix*, which is available on request the authors.

3.1.1. *Estimated Wage Elasticities and Production Functions*

Estimated *ordinary least squares* (OLS) parameter values of the supply function for health professionals based on *RHFs* and *UHFs* of Punjab are given in Table 2. The supply function for each professional included its respective wage rate and the number of registered professionals.

Table 2

Estimated Supply Functions of Health Personnel

	Rural Health Facilities (RHFs)		Urban Health Facilities (UHFs)		
Variables	Doctors	Paramedics	Doctors	Nurses	Paramedics
Intercept	–15.450 (–3.572)*	–8.108 (–8.585)*	0.295 (0.202)	–3.917 (–3.182)*	–1.587 (–2.149)*
Doctors Nominal Wage	1.472 (1.761)*	–	0.659 (2.258)*	–	–
Nurses Nominal Wage	–	–	–	0.253 (0.440)	–
Paramedics Nominal Wage	–	1.078 (3.239)*	–	–	0.498 (2.909)*
Registered Doctors	0.759 (1.849)*	–	0.127 (0.780)	–	–
Registered Nurses	–	–	–	0.9792 (1.956)*	–
Registered Paramedics	–	0.715 (2.021)*	–	–	0.271 (2.737)*
R^2	0.969	0.987	0.949	0.932	0.976
DW-test	1.690	1.570	2.386	1.837	2.44
Observations	1977–92	1977–92	1977–92	1977–92	1977–92

Notes: 1. Numbers in Parentheses are *t*-values.
2. Paramedics in RHFs include Nurses and Paramedics.
3. Asterisk indicates significance of the estimated parameters at 10 percent or less level of significance.

Since the function used is in logarithmic form, the parameters estimated in this context simply represent elasticities. For instance, in case of doctors in *RHFs*, the estimated parameter value of 1.47 in Table 2 represent the doctors' supply elasticity with respect to wage rates. Analysing the results in Table 2, it is interesting to note that health professionals in *UHFs* have inelastic wage elasticities (less than one). This implies that these professionals, in most cases, will be less than enthusiastic in offering their services in response to a wage rate increase. These wage elasticities (less than one). This implies that these professionals, in most cases, will be less than enthusiastic in offering their services in response to a wage rate increase. These results have important implications in terms of interpreting the *propositions* made earlier. Obviously, with an inelastic supply of health professionals, an institutionally fixed wage policy will not be copious enough to attract them towards the public health system.

Table 3 reports the estimated parameter values for health output (patients treated) production functions of both *RHFs* and *UHFs*. These estimated regression results in Table 3 indicate that, for both *RHFs* and *UHFs*, the role of infrastructure (in terms of large parameter values for beds as compared to other inputs) in the production process is more dominant than other remaining inputs, namely, health professionals. Analysing the health problem on the basis of production function alone, in this case, may suggest that resources be diverted towards building of infrastructure in order to improve health facilities. However, it should be noted that a production function approach may tell us only about the technical efficiency of the inputs.

Table 3

Estimated Production Functions for RHFs and UHFs

Variables	Rural Health Facilities (RHFs)	Urban Health Facilities (RHFs)
Intercept	1.373	–6.449
	(5.205)*	(–12.97)*
Doctors	0.129	0.353
	(0.929)	(8.024)*
Paramedics	0.360	0.290
	(3.095)*	(4.836)*
Hospital Beds	–	0.701
		(6.194)*
RHFs	0.561	–
	(1.948)*	
Registered Paramedics	0.715	–
	(2.021)*	
R^2	0.995	0.999
DW-test	2.297	1.961
Observations	1977–92	1977–92

Notes: 1. Numbers in parenthesis are *t*-values.
2. Paramedics in RHFs include Nurses and Paramedics.
3. Asterisk indicates significance of the estimated parameters at 10 percent or less level of significance.

Whether a *technically efficient* input is also *cost-effective* from the entire health system point of view can be established if the problem is analysed within a broader optimisation framework where the available cost constraints are also considered. Inputs derived from the optimisation approach will not only be *optimal* but, more importantly, they will be *cost-effective.* Distribution of health inputs through this optimisation principle will ensure the *allocative efficiency* of the health system. In the next section, we first discuss the *historical* optimisation simulation results and then present the *ex-ante* results.

3.1.2. *Ex-ante Forecast and Efficiency Gain/Loss (1992–2003)*

Based on the estimated supply and output elasticities from Tables 2 and 3, respectively, and exogenously given profile for total cost and total stock of registered health professionals, we can now directly compute the optimal expansion paths for inputs, and output for both *UHFs* and *RHFs* using the reduced form expressions derived from the optimisation problem.[6]

Tables 4–6 show that there were more beds for urban hospitals than they should have had in 1992 if the system had operated under *a cost-efficient* approach. A pertinent public policy question that may arise in this context is what should the *PPHD* do or what options are available to them so that there is a real improvement in the system in terms of *positive efficiency gains*? Obviously, one thing *PPHD* possibly *cannot* or *should not* do, is to dismantle its existing infrastructure. In fact, in future, the growth in infrastructure building programmes (in terms of new hospital beds) undertaken by *PPHD* should be slowed down or perhaps halted at the current level. Adopting such a policy will enable the *PPHD* to achieve the optimal mix of inputs in due course of time so that with each year gone by the population growth and the normal wear and tear of capital will gradually bring the existing number of beds closer to the optimal value.

In the following we present *ex-ante* simulation results, relying on two strategies: (a) *standard strategy* based on normal optimisation approach adopted earlier; and (b) *constrained strategy* based on an optimisation problem in which the beds in *UHFs* are now fixed at the *baseline* year 1992 level until they become equal to those of the *standard strategy.* The simulation results for *health personnel* and *infrastructure, wage structure* and *efficiency gain/loss* based on both *standard* as well as *constrained strategies* are reported in Tables 4–6.

It should be noted that the future data under the headings of ***actual*** in Tables 4-6, do not exist in published documents. They are, however, generated in this study by simply taking the annual compound growth rates of the past five to ten years of actual historical data on these variables. The underlying behaviour in projecting

[6]The detailed reduced form for the optimisation problem is available from the authors on request.

Table 4

Health Personnel and Physical Infrastructure: Ex-ante Forecasts for Standard and Constrained Optimisation

Variables \ Year	1992	1994	2003	1994	2003
	Urban Health Facility (UHF)				
	Baseline	Standard		Constrained	
Doctors					
Actual	23,5528	26,620	45,989	26,620	45,989
Optimal	22,044	24,842	24,533	24,499	42,533
%Change	–6.31%	–6.68%	–7.51%	–7.97%	–7.51%
Nurses					
Actual	10,981	12,990	27,421	12,990	27,421
Optimal	15,634	18,249	36,609	18,120	36,609
%Change	42.37%	40.48%	33.51%	39.49%	33.51%
Paramedics					
Actual	31,350	34,372	54,572	34,372	54,572
Optimal	33,458	37,462	62,305	37,016	62,305
%Change	6.72%	8.99%	14.17%	7.69%	14.17%
Beds					
Actual	50,138	54,800	81,028	54,800	81,028
Optimal	43,196	47,480	73,038	50,138	73,038
%Change	–13.85%	–13.36%	–9.86%	–8.51%	–9.86%
	Rural Health Facility (RHF)				
Doctors					
Actual	10,083	11,913	25,001	11,913	25,001
Optimal	9,605	11,755	29,179	11,516	29,179
%Change	–4.74%	–1.33%	16.71%	–3.33%	16.71%
Paramedics					
Actual	20,900	25,718	64,820	25,718	64,820
Optimal	22,726	27,273	61,988	26,782	61,988
%Change	8.74%	6.05%	–4.37%	4.14%	–4.37%
RHF					
Actual	1,274	1,442	2,499	1,442	2,499
Optimal	1,465	1,656	2,879	1,599	-2,879
%Change	14.99%	14.84%	15.21%	10.89%	15.21%

Table 5

Wage Structure for Health Facilities: Ex-ante Forecasts for Standard and Constrained Optimisation

Variable \ Year	1992	1994	2003	1994	2003
	Urban Health Facility (UHF)				
	Actual	Standard		Constrained	
Doctors					
Actual	54,303	61,208	103,954	61,208	103,954
Optimal	49,862	58,070	115,286	56,858	115,286
%Change	–8.18%	–5.13%	10.90%	–7.11%	10.90%
Nurses					
Actual	24,438	28,376	55,091	28,376	55,091
Optimal	30,801	34,630	58,680	33,678	58,680
%Change	26.04%	22.04%	6.51%	18.68%	6.51%
Paramedics					
Actual	17,975	20,703	38,753	20,703	38,753
Optimal	16,155	18,936	38,702	18,505	38,702
%Change	–10.13%	–8.53%	–0.13%	–10.62%	–0.13%
	Rural Health Facility (RHF)				
Doctors					
Actual	42,114	47,620	82,041	47,620	82,041
Optimal	51,389	55,107	75,463	54,316	75,463
% Change	22.02%	15.72%	–8.02%	14.06%	–8.02%
Paramedics					
Actual	16,651	20,240	48,274	20,240	48,274
Optimal	23,049	25,207	37,698	24,786	37,698
% Change	38.42%	24.54%	–21.91%	22.46%	–21.91%

Note: Wages are reported in current Pakistani rupees (Rs).

these actual data is predicated on the presumption that the *Public Health Department* continues to follow the future course of actions based on their health policies and practices of the recent past in matters concerning setting wages, hiring health professionals, annual development plans, etc. Comparing these actual data with the simulated ones generated by our optimisation model will not only enable us to examine the extent of divergence between optimal and sub-optimal health inputs within and across different health facilities but, more importantly, the above analysis will allow us to compute the size of *ex-ante efficiency gain/loss* of the optimal approach by simply calculating the yearly percentage deviation between optimal and actual outputs of a given health facility.

Table 6

Efficiency Gain/Loss for Health Facilities:
Ex-ante Forecasts for Standard and Constrained Optimisation

Variables \ Year	1992	1994	2003	1994	2003
	Urban Health Facility (UHF)				
	Actual	Standard		Constrained	
Health Output					
(Composite Index)					
Actual	13,182	17,703	65,288	17,703	65,288
Optimal	14,288	19,145	71,713	19,572	71,713
Efficiency (gain/loss)	8.39%	8.15%	9.84%	10.56%	9.84%
	Rural Health Facility (RHF)				
Health Output RHFs					
(Patient Treated)					
Actual	35,027	41,351	86,448	41,351	86,448
Optimal	38,055	44,948	95,450	43,703	95,450
Efficiency (gain/loss)	8.65%	8.70%	10.40%	5.69%	10.40%
	Combined Health Facilities				
Total Health Output					
(Composite Index)					
Actual	20,664	29,411	74,288	26,154	7.429e+09
Optimal	22,422	31,882	81,794	28,322	
Total Efficiency (gain/loss)	8.51%	8.40%	10.10%	8.29%	10.10%

Considering 1991-92 as the ***baseline*** year, future projections are made up to the end of the Perspective Plan period (2002-03). In addition, forecasts for all monetary variables are measured in current rupees. Having computed the actual projected data on health inputs (between 1992-93 to 2002-03) based on the methodology explained above, the corresponding actual health output variable is generated using the underlying estimated production technology parameters reported in Table 3.[7]

Focusing on the board policy issues, the *ex-ante* simulation results reported in Tables 4-6, reveal several interesting facts that have important implications for *public Health Policies* in Pakistan. The salient features of the results are summarised below:

Infrastructure

- Actual number of beds in *UHFs* in the baseline year 1991-92 appeared to be greater than what an *optimal optimisation strategy* would stipulate by over 13 percent. However, by the end of the *perspective plan* period (2002-03), this figure may come down to about 10 percent as shown in Table 4. It is

[7]Forecasting the actual health outputs based on the estimated production function, reported in Table 3, will not be unreasonable as the predictive power of these equations are very high (99 percent).

important to note that **if the *public Health Department (PHD)* were to adopt a policy of slowing down the growth of health infrastructure (in terms of beds) and re-allocate the existing resources then our results suggest that it is possible for the PHD to attain an *efficiency gain* of about 11 percent for *UHFs* in 1994 as reported in Table 6.**

- Furthermore, the re-allocation of resources from urban health development expenditure may take place not only towards other urban recurring activities (e.g., more medicine and health personnel) but, interestingly enough, these resources can now be diverted towards rural health facilities. In fact, our analysis suggests that some of these funds could be transferred towards the development of rural infrastructure (in terms of either upgrading *BHUs* to *RHCs* or outright construction of new *BHUs*).

Health Personnel

- In terms of health personnel, the optimisation model predicted a *faster* growth for doctors in *RHFs* during the entire plan period (about 11.1 percent p.a.), which is to be induced by raising their salaries by a ***substantial*** amount particularly in the earlier years (over 15 percent in 1993) as shown in Table 5. However, once enough doctors are attracted towards *RHFs*, the model predicts modest salary increases towards the latter part of the plan for the health system to be *cost-effective.*
- As for the doctors in the *UHFs*, *cost-effective* strategy suggests roughly the same number of doctors as the actual but with significantly higher salaries by the end of the plan period by about 11 percent as shown in Table 5. This result can be rationalised since, in urban areas, enough opportunities and alternatives are available to the doctors (in private hospitals or clinics). Thus, to improve urban health facilities in the long run, *PHD* should hire fewer good-quality doctors and pay them well.
- If *PHD* were to pursue its wage policies for nurses at the existing pattern, then the optimisation model predicts severe shortages of nurses to the extent of over forty-two percent in 1992 as reported in Table 4. Shortages will persist even at the end of the plan period 2002-2003. This excruciating situation of nurses in the present set-up in Pakistan is due to a combination of many factors. Obviously, one of the important reasons is the present low salary structure and, in this regard, the **optimisation model suggests an immediate boost in nurses' salary by over 26 percent** as shown in Table 5. Another reason is the social stigma attached to this profession particularly in the public sector health system in Pakistan. It needs to be emphasised that, even in the Defence Medical Services in Pakistan, nurses are not only paid reasonable salaries but, more importantly, they are accorded an honourable status in the military hierarchy. For instance, a

nurse would begin her career in the Armed Forces as a commissioned officer while, in the public health system, the starting level is only at the BSP grade 11 that is far below the category of an officer (BSP-17).

Expenditure Pattern

- The critical public policy issue that deserves some discussion in this context is the availability of funds to meet the future recurring expenditure obligations for the health system. Our optimisation model predicts recurring outlays of over 5 and 2.5 times that of development expenditures for *UHFs* and *RHFs*, respectively, in 1992-93. These figures are expected to remain high until the end of the *Perspective Plan period.* Many plans formulated in the past emphasised the establishment of infrastructure in terms of opening new hospitals, *RHSs*, and *BHUs* without giving due consideration towards the recurring expenditures [e.g. health personnel and other inputs (medicine, x-ray films, etc.)]. **What is crucial for the long-term sustainability of the Public Health System is the commitment from the policy-makers for provision of steady inflow of funds to meet recurring expenditure.**

Efficiency Gains

- A crucial part of our analysis in this study is to investigate whether by becoming more *cost-effective* it is possible for the *Public Health Departments* to improve their provision of basic health facilities. The optimisation results in Table 6 suggest that about ten percent efficiency gains can be made in *UHFs* and *RHFs*. In fact, our simulation results in Table 6 clearly supports the *proposition* made earlier that, by reallocating resources within *UHFs* from development to recurring expenditures, reasonable *efficiency gains* of over ten percent were attained in 1994. **The *efficiency gains* for the combined health system were also in the order of eight to nine percent.**

4. CONCLUSION AND POLICY RECOMMENDATIONS

Pakistan's share of total health expenditure to gross national product has never exceeded 0.8 percent per annum [*Economic Survey 1992-93*], which is significantly lower than many of its neighbouring countries in the region. If the future is any reflection of past history, then one does not expect substantial public funds to be forthcoming and diverted towards this sector in the immediate or medium term future especially when the country is already experiencing large increasing budgetary deficits. Prudent public policy research in this context, based on a realistic pragmatic approach, should then be geared towards an investigation into measures to improve the present *Public Health System (PHS)* through an *efficient*, *cost-effective* reallocation of health inputs within the existing limited budget. This study has examined these health policy issues within the context of an optimisation framework for *PHS* and forecasted future (up to 2002-03) *efficient* optimal mix of health inputs (doctors, nurses, paramedics), outputs (patients treated for urban and rural health facilities), expenditures (development and recurrent) and wage policies (health personnel) under alternative scenarios. Comparing the projected actual health outputs (based on historical growth rates) with those of optimal values, efficiency gains were also computed. Based on these simulation results, this study makes policy recommendations for *PHS* and some of the important ones are summarised below:

- Our optimisation model predicts an excess build-up of infrastructure (measured in terms of beds) in the present urban health facilities while, in the *rural* areas, there is a paucity of *RHCs* and *BHUs*. **We recommend that the growth of health infrastructure building in the urban areas be slowed down in the short-run (two to three years) and some of the resources be reallocated toward the rural sector either in terms of building new BHUs or upgrading the existing RHCs.**
- Our optimal forecasts suggest that both more nurses and doctors should be hired in *UHFs* and *RHFs*, respectively, during the entire plan period though, in terms of percentage, the demand for nurses will exceed that of doctors in *RHFs*. **We recommend that not only attractive wage policies be formulated for these personnel but, more importantly, like the Armed Forces, the status of nurses in the Public Health System be elevated by giving them higher BSP.**
- There is a shift in focus from development towards recurring expenditure as predicted by the optimisation model. **It is, therefore, recommended that for every rupee of development expenditure incurred, PHD *must* plan or keep provisions for recurring outlays.** This is critical from the point of view of long-term sustainability of *PHS*.

It is important to note that all this reallocation of resources is feasible within the projected actual budget and, interestingly enough, it will also lead to *efficiency gains* in the order of 8 to 10 percent for the entire *Public Health System*.

REFERENCES

Pakistan, Government of (Various Issues) *Economic Survey*. Islamabad: Finance Division, Economic Advisors Wings.

Punjab, Government of (Various Issues) *Demand for Grants Current Expenditure*.

Punjab, Government of (Various Issues) *Punjab Development Statistics*. Lahore: Bureau of Statistics.

The World Bank Report (1991) *Pakistan Health Sector Study: Key Concerns and Solution*. New York: The World Bank Report.

World Development Report (1991) *The Challenge of Development*. New York: Oxford University Press.

Comments

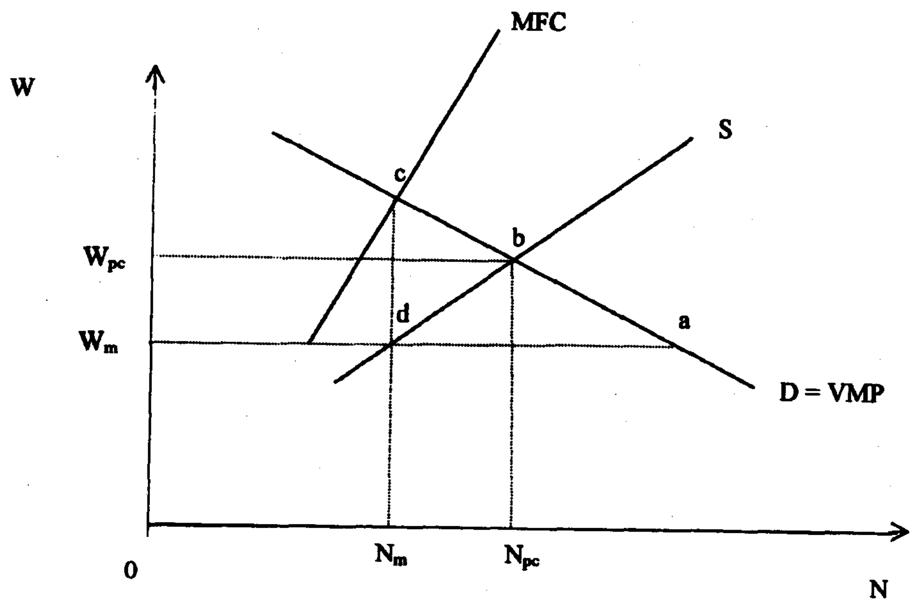
I commend the authors for attempting to analyse rigorously the economics of the public health sector in Pakistan. The main conclusions are that channelling more resources to personnel, nurses, and doctors and less to infrastructure for the same total cost will result in more output of about 8–10 percent. I do not necessarily disagree with the authors' conclusions, but I recommend that we should treat these conclusions with extreme caution and as highly tentative. This is because the study suffers from several conceptual problems and employs highly questionable assumptions and data set.

The paper starts out by lamenting on the meager resources that are devoted to the public health sector. Statistics on health as a percentage of GNP for some neighbouring and other countries are provided with the purpose of showing the dismal state of relative expenditures on health in Pakistan. At this point I was expecting that the authors would go on to forcefully argue for more expenditures to be allocated to public health. Instead, the authors' objective turns out to be of finding whether the health sector is economically efficient with the meager resources that it gets.

The framework employed is the neoclassical optimisation problem whereby the health sector maximises output with respect to the various inputs subject to a budget constraint. Remember we are talking about the public sector here and not the private sector. The equilibrium solution to this problem determines the optimal amounts of health personnel and output in each time period. While the framework employed is more elaborate, it is the same one that I used in my paper entitled, "Optimal Resource Utilisation in Pakistan Agriculture" presented at the third PSDE Meetings, Khilji (1986). There are some methodological differences though among these papers. My paper focussed on the private agricultural sector and not on the public sector. It allowed equilibrium to be reached over several periods instead of assuming that equilibrium is reached in each period as in this paper. My paper allowed the output elasticities, with respect to the various inputs, to vary over time instead of the constant elasticities assumed in this paper.

In terms of efficiency, the authors are never clear of what type of inefficiency they are referring to. According to the quotation from the World Bank study which they present on the first page, inefficiency is due to, "Political influence and leakage of equipment and supplies....... Poor management and centralised financial, administrative and management authority..." These notions of inefficiency are different than what the authors model. The paper talks about allocative inefficiency due to the monopsonistic power of the public health sector, which is assumed rather than proved. Employing isoquant and isocost curves in Figure 1, the authors painstakingly explain that a fixed wage for doctors below a market determined wage would bring in fewer doctors. This will result in less output for the same cost. All this is familiar to undergraduate students in economics. There are many ways the same point is normally made. Price and wage ceilings result in misallocation of resources. This is a proposition that we teach our first year students and is not as novel as is presented in Proposition 1 of the paper. However, the question arises, what is the source of inefficiency? Is it an institutionally fixed wage or is it the monopsonistic power of the public health sector. Instead of worrying about isoquants and isocosts, which are rather cumbersome, the point can be equally made by using the familiar supply and demand diagrams. Figure 1 below does just that.

FIGURE 1: Monopsony in the Health Sector



The figure presents the market for doctors. D is the demand curve by the public health sector for doctors and S is the supply of doctors at various wage rates W. The wage W_{pc} is the market clearing wage where supply is equal to demand. It would result in N_{pc} amount of doctors hired per period of time. Since D is also the Value Marginal Product (VMP) of Doctors = Price* Marginal Product of Doctors, N_{pc} amount of doctors coincides with D_1 in Figure 1 of the paper. Output would be Q_1 which is optimal. The problem with a monopsonist is that he faces an upward sloping supply curve. This means that when he hires additional doctors, he has to raise the wage, not only for the additional doctor, but also for all doctors in his employment. This gives rise to a marginal factor cost (MFC) for him which is higher than the supply price of doctors. In order to maximise output for a given cost, he will equate the VMP of doctors not to the wage rate but rather additional cost of hiring them which is MFC. This occurs at point c in Figure 1. The wage that corresponds to MFC at point c is W_m which is what the monopsonist will pay At this wage only N_M doctors will be willing to work. N_M doctors corresponds to D_0 in Figure 1 of the paper where the wage rate < VMP of doctors resulting in sub-optimal amounts (from the economy's stand-point) of doctors and output. The reason here for allocative inefficiency in the economy is the monopsonistic power of the producer.

The point of presenting all this is to show that both institutionally fixed wages below the market clearing wages or monopsony elements lead to inefficient allocation of resources. For example, Wm could be either institutionally fixed or can be the result of the monopsonist's optimisation behaviour The effects will be the same. The paper moves back and forth between these two. Moreover this is a far cry from what the World Bank study had in mind about leaks, political influence, and poor management. These types of inefficiencies would be hard to put in the optimisation framework employed in the paper.

I wander why the paper wastes so much time proving Propositions 2 and 3. We all know that if a monopsonist faces an infinitely elastic supply curve he would not be, by definition, a monoposonist but a perfect competitor in that factor market. Proposition 3 is also well known. If you are stuck with a fixed factor higher than the optimal amount than keeping it fixed and increasing other inputs will result in higher output than a situation where you start with an optimal combination of inputs. This has to do with higher marginal products of the relatively scarce factors.

Let me turn now to the econometrics, solution algorithm, and simulations in the paper. Unfortunately the paper is not very forthcoming about any of these matters. In fact it does not even elaborate on the data employed except to state that it is the health sector data on Punjab. The other thing we know is that this data is for the period 1977–1992.

Essentially there are two outputs, one is an index of patients treated in Rural Health Facilities, and the other is an index of patients treated in Urban Health Facilities. These outputs are related to inputs through production functions which are statistically estimated with inputs as doctors, paramedics, hospital beds, and registered paramedics. Given these estimated production functions and an exogenous budget to be allocated for a year, constrained optimisation of output leads to demand equations for medical personnel for each year. The supply equations of doctors nurses and paramedics are statistically estimated. Equating these factors' supplies to their demands for a particular year provides equilibrium solutions for the amounts of factors employed and their wage rates for that year. Substituting these solution values of inputs into the respective production equations generates values for the two outputs for the year. This solution algorithm is repeated each year based on a new budget allocation. The solution algorithm calculates equilibrium amounts of inputs employed, wage rates, and outputs each year.

In standard forecasting there are two types of simulations. These are *ex-post* and *ex-ante* simulations. Historical simulations (or *ex-post* simulations) involve the comparison of predicted values of inputs and outputs in the past with the actual values of these variables The purpose is to check on the forecasting accuracy of the estimated model. *Ex-ante* simulations (or forecasts) involve projecting the endogenous variables into the future based on expected values of all exogenous variables. Since the future has not happened, it is not possible to check whether the *ex-ante* forecasts are accurate or not.

The paper performs *ex-ante* simulations with a twist. The twist is that it also projects forward, from 1992 to 2003, what the likely values of the endogenous variables (inputs, wage rates, and outputs) are going to be based on past growth rates. These are compared with the values generated from the solution algorithm. This is absolutely bizarre! The questions is that, are not the estimated input elasticities, and supply equations based on historical values of the same endogenous variables? One can understand the use of the word optimal in the *ex-ante* simulations. It can be rationalised as follows: Given the existing structure of production, budget allocation to health, input supply equations, and assuming that the public health sector wants to maximise output, then the optimal amounts of inputs, wage rates, and outputs are the ones that are reported as optimal in Tables 4, 5, and 6. Actual has no meaning here, except for 1992, since the values are not observed.

If the purpose were to show efficiency gains, would not *ex-post* simulations be more relevant? Even if we were to accept the procedure adopted in the paper, we find that the actual amounts of doctors employed are higher than optimal for UHFs and for RHFs (for most years) in Table 4. I thought that the point of the exercise was to show that because wages of doctors were lower than the competitive wages due to monopsonistic elements, less doctors were employed than optimal. Finally, comparison of actual output to optimal output in Table 6 shows efficiency gains in output of the order of 8–10 percent. This is not much given that most of the important parameters are statistically estimated. It would be useful to present the standard errors associated with these efficiency gains to assess their significance. Even if these were not calculable, some qualifications about these point estimates would make sense.

Recently Anne Krueger (1997) remarked that when economists present empirical results that are favourable to one point of view, politicians, whose side is favoured, exploit this result claiming it to be gospel truth. They do not heed all the qualifications, conditions, and assumptions the economists may have employed and stated when presenting their results. Unfortunately no qualifications are put in this paper making it easier for people to claim that the health sector is inefficient. There need to be several qualifications put in this paper. These have to do with the limited data employed, the definition of output, the statistical estimations of coefficients which are assumed to stay constant in the future and their associated standard errors. Additionally, the arbitrary nature of the assumptions about the future course of budget allocations, and expected inputs, wage rates and outputs should be emphasised

While this is an elegant and elaborate piece of work, I am sure the authors would agree that, we should be extremely careful in using the results reported in this study, especially for policy purposes.

Nasir M. Khilji

U.S. Bureau of the Census
and
Economic Adviser
U.S./Saudi Arabian Joint Commission
on Economic Cooperation (JECOR)
Riyadh

REFERENCES

Khilji, N. M. (1986) Optimum Resource Utilisation in Pakistan's Agriculture. *The Pakistan Development Review* 25:4 469–487.

Krueger, A. O. (1997) Trade Policy and Economic Development: How We Learn. *The American Economic Review* March.