Indira Gandhi
National Open University
School of Social Sciences

MEC-108
Economics of Social Sector and Environment

Block

3

ECONOMICS OF HEALTH

Expert Committee

Prof. Joyashree Roy
Jadavpur University
Kolkata

Prof. S. P. Singh
I. I. T. Roorkee
Roorkee

Prof. S. Sandhya
Central University of Hyderabad
Hyderabad

Prof. Duraisamy
Madras Institute of Development Studies
Chennai

Prof. Padmaja Mishra
Utkal University
Bhubaneshwar

Dr. Subrata Mandal
Ambedkar University
Delhi

Prof. Saumen Chattopadhyay
ZHCES, Jawaharlal Nehru University
New Delhi

Prof. Gopinath Pradhan
School of Social Sciences
Indira Gandhi National Open University
New Delhi

Prof. Narayan Prasad
School of Social Sciences
Indira Gandhi National Open University
New Delhi

Prof. Kaustuva Barik
School of Social Sciences
Indira Gandhi National Open University
New Delhi

Sh. Saugato Sen
School of Social Sciences
Indira Gandhi National Open University
New Delhi

Prof. B. S. Prakash **(Convener)**
School of Social Sciences
Indira Gandhi National Open University
New Delhi

Course Coordinator : Prof. B.S. Prakash

Block Preparation Team

Unit No.	Resource Person (Unit Writer)	Format and Language Editing (Units 7 to 9)	Editor (Content) (Units 7 to 9)
7	Dr. Satarupa Bandyopadhyay Assistant Professor Department of Economics Bethune College, Kolkata	Dr. B.S. Prakash Professor of Economics SOSS, IGNOU, New Delhi	Prof. S. Sandhya Central University of Hyderabad Hyderabad
8	Prof. Sushil Kr. Haldar Department of Economics Jadavpur University, Kolkata		
9	Prof. Arijita Dutta Department of Economics University of Calcutta,Kolkata		

Material Production

Mr. Manjit Singh
Section Officer (Publication)
School of Social Sciences
IGNOU

Cover Design

M/s ADA Graphics
New Delhi

July, 2016

ISBN-

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by Director, School of Social Sciences.

Lasertypesetted at Graphic Printers, 204, Pankaj Tower, Mayur Vihar, Phase-I, Delhi-110091.

Printed at :

BLOCK INTRODUCTION

The present block is focused on Economics of Health. It has three units. As in Block 2 for education, here also **Unit 7** focuses on Demand for Health Services. Beginning with an introduction on conceptual aspects of some important health indicators, the unit presents a discussion on the linkage between health indicators and economic development. Focusing in particular on the role of economics in health sector, the unit then explains the central argument behind the 'efficiency wage hypothesis' and its relationship to human development. The importance of externalities of health to economic development is then discussed. Theories of relevance to demand for health like Becker's theory of behaviour and Grossmann's human capital model are then explained. The unit concludes by a discussion on the 'supply factors affecting the demand for health'. In this, the role of physicians, pharmaceutical companies and health insurance are discussed.

Unit 8 is on 'Supply of Health Services'. Beginning with a distinction between the 'healthcare services' and 'health services', the unit proceeds to explain the difference between 'medical care' and 'healthcare'. The coexistence of different markets for health services in an economy is then discussed. Focused on some of the important specific cases of such market conditions, the unit then makes an analytical presentation of the conditions required for the establishment of optimal services of healthcare. The supply of health services crucially depending on its production, and the efficiency frontier for the production of health services being different for different production technologies, the unit then discusses how the optimal mix of healthcare inputs can be used for producing a fixed amount of medical care. Since providing healthcare services by the public sector for the achievement of 'social optimum' is most important, the unit makes an analytical presentation of the optimality conditions required to be established for achieving this. Finally, keeping in view the limitations of public healthcare and the importance of PPPs in providing the required volume of healthcare services, the role of 'public private partnership' initiatives (i.e. PPPs) is outlined.

Unit 9 is on 'Determinants of Health Services'. The unit discusses the determinants of demand for and supply of healthcare services separately. Under the former, factors like income & health, poverty & malnutrition, other socio-economic determinants of health and healthcare finance are discussed. In the latter, factors like: (i) prices & wage, (ii) availability of adequate health workers, (iii) considerations of technical efficiency, (iv) pharmaceutical pricing and (v) government policy are discussed. The three units in the block, taken together, are thus aimed at giving a composite view of Health Economics with a focus on conceptual and analytical issues involved.

4th Blank

UNIT 7 DEMAND FOR HEALTH SERVICES

Structure

7.0 OBJECTIVES

After going through this unit, you will be able to:

- evaluate whether health is a consumption or an investment good;
- describe the various indicators of measuring health;
- discuss the different health indicators along with its measurement issues;
- bring out the linkage between health and economic development;
- discuss the significance of 'efficiency wage' in contributing to health and productivity of workers;
- explain the concept of externality in health;

- comment on why health is both a public and a merit good;
- distinguish between the concepts of private demand and social demand for health;
- explain the role of health in economic development.
- explain how 'Becker's Theory of Behaviour' is useful in explaining the consumer behaviour by incorporating health as an important variable;
- derive the conditions required under the Grossman model for investment in health to be optimum; and
- discuss the various supply side factors which influence the demand for health.

7.1 INTRODUCTION

Health, or more specifically, good health, is the most primary requirement for the wellbeing of the human race. In the preamble of the 1746 Constitution of the World Health Organisation, health has been defined as *'a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity'*. The enjoyment of the highest attainable standard of health has been acknowledged as one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition'. To remain healthy has therefore always been a matter of daily concern for every member of any household, because, ill health, apart from making us suffer physically and mentally, acts as a hindrance to other sources of wellbeing also. For instance, it may stop us from going to school, to work, or even to movies, thereby disturbing our other social activities which are both our rights and duties towards the society.

Although health is looked at as a state of physical and mental wellbeing with absence of diseases, economists take a radically different view in this regard. They view health as a commodity, a durable good, or a type of capital which provides a flow of services [produced from a stock of health (capital)] consumed over a person's entire life time. Each individual is assumed to be endowed with a given stock of health at the beginning of his life. Over the period, the stock of health depreciates with age which may be reinforced by investments in medical services. Death occurs when an individual's stock of health falls below a critical minimum level. Naturally, the initial stock of health, and the rate of depreciation, vary from person to person and depend on many economic and non-economic factors, some of which are uncontrollable. For instance, a person has no control over the initial stock of health allocated to him at birth, but medical services can compensate for many deficiencies to some extent. The rate at which health depreciates also depends on many factors, such as the individual's age, physical structure, lifestyle, environmental factors, and the amount of medical care consumed. For instance, the rate at which health depreciates in a person diagnosed with high blood pressure is likely to depend on the amount of medical care received, environmental factors (does he or she have a stressful occupation?), and lifestyle (does the person smoke or have a weight problem?). All these factors interact to determine the person's stock of health at any point of time, along with the pace at which it depreciates.

People therefore desire good health as a commodity both for consumption and investment purposes. Good health is considered as a durable consumption good because it yields utility by improving the quality of life. The investment element of health is explained by the relationship between health and time. If one is in good health, (s)he is less sick and therefore will have more healthy days available in the future to work and enhance income

or pursue other activities, such as leisure. It is looked at by the economists from the same perspective as education. Just like a person invests in education to increase the potential to access to a higher wage, a person invests in health to increase the likelihood of having more healthy days to work and generate income.

7.2 HEALTH INDICATORS

Since health is not tangible in nature, the concept of remaining in good health differs from person to person, depending on their social and cultural background. For instance, a poor illiterate slum dweller may consider himself *quite* healthy under same poor health conditions as compared to a well-educated upper-middle class person. This poses the question whether the measure of health status could be different for individuals and the population as a whole. In other words, measurement of health actually depends both on quantity as well as quality of life. However, some indicators are commonly used as measures of health status at both micro and macro level. Before defining those indicators, one should understand the broad categories of the health indicators i.e. exactly what is measured and why. An indicator is a measure that is used to demonstrate change in a situation, or the progress in, or results of, an activity, project, or programme. Any health project or programme, or more generally, the health system of a country can be evaluated by these indicators. Conceptually, health indicators can be divided into four broad categories depending on which stage of the programme/system is being evaluated.

Input Indicators: They measure the quantity, quality, and timeliness of resources — human, financial and material, technological and information — provided for an activity/ project/programme. They also include the policy environment i.e. national policies and legislation with regard to health. Examples of this type of indicators are: number of doctors per thousand population, number of hospitals for 1 lakh population, access of mothers and children to local healthcare, birth attendance by trained personnel, etc.

Process Indicators: It is difficult to isolate the effects of programmes on outcomes. Hence programmes are evaluated in terms of intermediate outputs or process indicators. These indicators basically measure *quality* of health activities i.e. the progress of activities in a programme/project and the way these are carried out. The process indicators of health program address operational issues and questions that can be answered with programme level data and measures. These indicators may enable policy makers and programme managers to assess and improve the health services (e.g. number of participants in the pulse polio programme in a month, number of institutional deliveries during one year, etc.).

Output/Outcome Indicators: They measure the output/outcomes of several programmes. Examples are IMR, CDR, expectation of life at birth, cause specific morbidity and mortality rates.

Impact Indicators: They measure the change in achievements in health status due to the effect of interventions/actions which have occurred over a long-term (e.g. measurable change in quality of life, reduced incidence of diseases, increased income for women, reduced mortality, etc.)

Besides the above, there are non-health socio-economic indicators which measure variables like population structure, educational distribution, magnitude and nature of employment, standards of water/sanitation/pollution levels, etc. which affect the health system indirectly. We now define some of the most widely used health indicators internationally.

7.2.1 Mortality Indicators

Mortality means death. Mortality rate is the number of deaths due to a disease divided by the total population. If there are 25 lung cancer deaths in one year in a population of 30,000, then the mortality rate for that population is 83 per 100,000. The different kinds of mortality indicators are:

Crude Birth/Death Rate (CBR/CDR): CBR/CDR is the number of live births/ deaths occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of that area during the same year.

CBR/CDR = Number of live births ÷ deaths mid-year total population * 1000

Infant Mortality Rate (IMR): IMR is the number of deaths of children under one year of age per 1000 live births. The infant mortality rate correlates very strongly with, and is among the best predictors of, state failure. IMR is therefore also a useful indicator of a country's level of health or development, and is one of the components of the *physical quality of life index* (PQLI). The method of calculating IMR often varies widely between countries, and is based on how they define a live birth and how many premature infants are born in the country.

Maternal Mortality Rate (MMR): MMR is the annual number of female deaths per 1 lakh live births from any cause related to or aggravated by pregnancy or its management (excluding accidental or incidental causes).

Proportional Mortality Rate (PMR): PMR is the number of deaths from a specific cause in a specific period of time per 100 deaths from all causes in the same time period.

Life Expectancy (LE): LE is a statistical measure of how long a person (or organism) may live, based on the year of their birth, their current age and other demographic factors including gender. At a given age, life expectancy is the average number of years that is likely to be lived by a group of individuals (of age x) exposed to the same mortality conditions until they die. The most commonly used measure of life expectancy is life expectancy at age zero; that is, at birth (LEB), which can be defined in two ways: *cohort* LEB defined as the mean length of life of an actual birth cohort of all individuals born in a given year (this can be computed only for cohorts that were born many decades ago and all their members have died) and *period* LEB defined as the mean length of life of a hypothetical cohort assumed to be exposed (from birth till death) to the same mortality factors or conditions observed in a given year.

Life expectancy is by definition an arithmetic mean. It can also be calculated by integrating the survival curve from age 0 to positive infinity (or equivalently to the maximum lifespan, say 'omega'). For an extinct or completed cohort (say all people born in year 1850), it can be calculated by averaging the ages at death. For cohorts with some survivors, it is estimated by using mortality experience in recent years. Such estimates are called period cohort life expectancies.

Illustration: Let there be 21 intervals of age, with the start of the interval taken as 'zero' and end point of interval taken as 100 as:

$X_0 = 0, X_1 = 1, X_2 = 5, X_3 = 10, ..., X_{21} = 100$

Let $M = (M_i)_{21\times1}$ be the vector of mortality probabilities denoting the probability of dying between age X_{i-1} and X_i, for $i = 1, ..., 21$. Now, to find out the life expectancy at birth, one must calculate the total years lived in each interval. In other words, the LE

at 'i', LE_i, is:

$LE_i = (X_i - X_{i-1})p_i + a_i d_i$

where, $X_{i\partial 1}$ and X_i are the ending and starting points of each interval, p_i is the percentage of total population that lives on to the i + 1 interval, a_i is the average number of years lived in an interval by an individual who dies in the same interval and d_i is the percentage of total population that dies in the interval (X_{i-1}, X_i). We calculate p_i and d_i as:

$p_i = \Pi_{j=1}^{i}(1 - M_j)$ with $p_0 = 1$ and $d_i = p_{i-1}M_i$

Now, the life expectancy at birth is: $LE = \sum_{i=1}^{21} LEi$

Since estimating the average years lived in an interval by accounting for those who have died is difficult, it would be simple to assume:

$a_i = (X_i - X_{i-1})/2$

This implies that on an average, people die in the middle of the interval. Obviously, this is a bit presumptuous and leads to small errors in calculation.

7.2.2 Morbidity Indicators

Morbidity means illness. It is measured by two indicators: prevalence rate and incidence rate. A person can have several co-morbidities simultaneously (e.g. ranging from Alzheimer's to cancer to traumatic brain injury). Prevalence is a measure used to determine the level of morbidity in a population like determining a person's likelihood of having a disease. The number of prevalent cases is the total number of cases of disease existing in a population. A prevalence rate is the total number of cases of a disease existing in a population divided by the total population in a particular area. For instance, if a measurement of cancer is taken in a population of 40,000 people, of whom 1,200 were recently diagnosed with cancer and 3,500 are living with cancer, then the prevalence of cancer is 4700/40000 = 0.1175 (or 11,750 per 100,000 persons). Incidence rate is used as a measure of disease that allows us to determine a person's probability of being diagnosed with a disease during a given period of time. Incidence is, therefore, the number of newly diagnosed cases of a disease. An incidence rate is the number of new cases of a disease divided by the number of persons at risk for the disease. For instance, if over the course of one year, five women out of a total female study population of 200 who did not have breast cancer at the beginning of the study period, are diagnosed with breast cancer, then the incidence of breast cancer in this population is 5/200 = 0.025 (or 2,500 per 100,000 women-years of study).

Health Status Indicator: These are the measurement of the health status of a given population conveyed by using various indices (including morbidity, mortality). For instance, Low Birth Weight (LBW) is an indicator defined as less than 2,500 gm of a newly born child (whose weight is measured within the first hour of its life). The incidence of low birth weight in a population is defined as the percentage of live births that weigh less than 2,500 gm out of the total number of live births during the same time period. Other health status indicators include number of hospital visits due to injury, etc.

7.2.3 Burden of Disease Indicators

At the individual level, the term 'burden of disease' refer to the overall impact of diseases and injuries. At the societal level, it refers to the economic costs of diseases. The term 'global burden of disease' (GBD) refers to an indicator published by World Bank in its

World Development Report (1993). It measures the total loss of health resulting from diseases and injuries. Disease burden is thus the impact of a health problem as measured by financial cost, mortality, morbidity, etc. It is often expressed in terms of quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs), both of which quantifies the number of years lost due to disease (YLDs).

Quality Adjusted Life Year (QALY): The QALY was invented in the 1970s and has become an internationally recognised indicator since the mid-1990s. QALY is the arithmetic product of life expectancy combined with a measure of the quality of life-years remaining. The calculation is relatively straightforward: the time a person is likely to spend in a particular state of health, weighted by a utility score from standard valuations. In such valuation systems, '1' equates perfect health and '0' equates death. Since certain health states are characterised by severe disability and pain, they are regarded as worse than death and hence are even assigned negative values.

If a healthcare intervention provides perfect health for one additional year, it would produce one QALY. Likewise, an intervention providing an extra two years of life at a health status of 0.5 would also equal one QALY. This effect when related to cost is expressed as 'cost per QALY'. For instance, if a new treatment gives an additional 0.5 QALYs and the cost of the new treatment per patient is Rs. 5,000, then the cost per QALY would be Rs. 10,000 (i.e. 5,000/0.5).

Disability-Adjusted Life Year (DALY): DALY is an alternative tool (which emerged in the early 1990s) as a means of quantifying the burden of disease. It sums up years of life lost (YLL) due to premature mortality and years lived in disability/disease (YLD). YLL are calculated as the number of deaths at each age multiplied by the standard life expectancy for each age. YLD represents the number of disease/disability cases in a period multiplied by the average duration of disease/disability weighted by a disease/disability factor. For instance, a woman with a standard life expectancy of 82.5 years, dying at age 50, would suffer 32.5 YLL. If she additionally turned blind at aged 45, this would add 5 years spent in a disability state. If the weight factor is 0.33, then it results in 0.33 x 5 = 1.65 YLD. In total, this would amount to 32.50 + 1.65 = 34.15 DALYs.

For DALYs, the scale used to measure the health state is inverted to a 'severity scale', where '0' equates perfect health and '1' equates death. The weight factors are age-adjusted to reflect social preference towards life years of a young adult (over an older adult or young child). Furthermore, they are discounted with time, thus favouring immediate over future health benefits.

Check Your Progress 1 [answer the questions in about 100 words within the space given]

1. Why is 'health' regarded as both a consumption as well as an investment good?

   .....

   .....

   .....

   .....

2. State the four categories into which 'health indicators' are classified? How are the 'output' indicators different from the 'impact' indicators?

   .....

.....

.....

.....

3. Which mortality indicator is regarded as useful for assessing country's level of health or development? How is it defined?

.....

.....

.....

.....

4. How is life expectancy defined? Distinguish between cohort LEB and period LEB.

.....

.....

.....

.....

5. Distinguish between QALYs and DALYs. How are they measured?

.....

.....

.....

.....

7.3 HEALTH INDICATORS AND ECONOMIC DEVELOPMENT: LINKAGE

At a micro level, ill health could lead to loss of income, more expenditure on health, even indebtedness, loss of property and wealth and medical poverty. It may also lead to loss of years in education, thereby reducing the quality of human resource. At macro level, the consequences are lower labour resource and productivity, lower GDP (with even larger share of GDP going to health expenditure), lower capital formation, etc. Serious illness is a major reason of poverty in countries with inadequate public health services and insurance coverage. Reports reveal that in the countries hardest hit by malaria or AIDS, the progress on development had ceased altogether (WHO, 2000). Till 1970s, economic growth and economic development was treated synonymous and it was thought that per-capita income, the indicator of economic growth, would indicate the development as well. A sharper distinction between growth and development was introduced in 1970s, with growth considered to merely show the performance of the economy and development identified with the end or the desired outcome of growth. Development in this sense reflected the distribution of benefits of growth across different sections of the society with concerns of equity duly addressed. An even more comprehensive indicator of economic growth with development was introduced in the early 1990s in which the larger issue of quality of life was given due weightage. This issue was first raised in the World Development Report of 1991 which stated that: *the*

challenge of development is to improve the quality of life which goes beyond mere economic growth.....it encompasses better education, higher standard of health and nutrition, less poverty, a cleaner environment, more equality of opportunity, greater individual freedom and a richer cultural life. The transition from 1970s to 1990s, thus saw the replacement of physical quality of life index (PQLI) [which was developed in 1979 by taking into account the three variables of life expectancy at age 1, infant mortality rate and literacy], by the more comprehensive human development index (HDI). The latter (i.e. HDI) was also based on three related aspects but with wider and extended connotations viz. (i) *longevity* measured by life expectancy at birth, (ii) *knowledge* measured by the weighted average of adult literacy (two thirds) and combined primary, secondary and tertiary gross enrolment ratio (one third) and (iii) the *standard of living* measured by per capita real income adjusted for purchasing power parity. Its focus was thus on the ends of development (longevity, knowledge and material choice) rather than on the means of development. HDI is also used to measure the relative position of development (inter-regional or sub-national and international) on a more comparable basis.

Measurement of HDI: In the 2010 Human Development Report, the computation of HDI was further improved by modifying in terms of the following three dimensions:

- A long and healthy life indexed by 'life expectancy at birth';
- Good education measured by the 'mean years of schooling' and 'expected years of schooling'; and
- A decent standard of living measured by the GNI per capita (PPP US$)

The construction of the three indices was thus made as:

1. Life Expectancy Index (LEI) $= \dfrac{LE - 20}{85 - 20}$

2. Education Index (EI) $= \dfrac{MYSI + EYSI}{2}$ where

$$MYSI = \frac{MYS}{15} \quad \text{and} \quad EYSI = \frac{EYS}{18}$$

3. Income Index (II) $= \dfrac{\ln(GNIpc) - \ln(100)}{\ln(75{,}000) - \ln(100)}$

The HDI was then arrived at as the geometric mean of the above three normalized indices. That is:

$$HDI = \sqrt[3]{LEI \cdot EI \cdot II.}$$

[**Note**: LE is life expectancy at birth. MYS is mean years of schooling (i.e. years that a person 25 years-of-age or older has spent in schools). EYS is expected years of schooling (i.e. years that a 5-year-old child will spend in schools throughout his life). GNIpc is gross national income at purchasing power parity per capita. Fixing the maximum at $75,000 means that for countries with GNI per capita greater than $75,000, only the first $75,000 contributes to human development. By this, the higher income is prevented from dominating the HDI value. Currently, we have only 4 countries with GNI pc above the cap – Liechtenstein, Kuwait, Qatar and Singapore.]

Health and development are thus closely interrelated with a potential to influence each other strongly. In recognition of this fact, health has been accorded a distinct status in the measurement of the level of development of an economy. Health contributes to the process of human capital formation making a major contribution to raising the productivity of labour. Due to these strong forward and backward linkages, investment in health increases the returns on other investment in human capital (such as in education) by contributing to ones' learning abilities. The net result is a positive return in terms of increased productive years of life.

7.4 ROLE OF ECONOMICS IN HEALTH SECTOR

The role of economics in health sector entails how economic principles work in determining the health sector outcomes. For an individual desiring to enjoy good health, health is like a production function where good health is a functional outcome of several factors. At society level, in both its micro and macro perspective, it is about efficient allocation of the limited resources to have an optimal level of health. Health economics therefore addresses the various economic theories by considering health as a commodity.

Four tenets of neo-classical welfare economics (viz. utility maximization, individual sovereignty, consequentialism and welfarism) are of particular importance in this context. Utility maximization, is essentially a behavioural assumption while the latter three are normative assumptions concerning who is in the best position to judge welfare and the type of information required for judging the goodness of a resource allocation. Utility maximization holds that individuals choose rationally – that is, given a set of options, an individual can rank the options and choose the most preferred among them as per the defined notions of consistency. Individual sovereignty asserts that individuals are the best judges of their own welfare i.e. any assessment of individual welfare should be based on a person's own judgement. It rejects paternalism – a notion that a third party may know better than the individuals themselves on what is best for them. Consequentialism holds that any action, choice or policy must be judged exclusively in terms of the resulting, or consequent, effects. In other words, outcome, not process, matters. Welfarism is the proposition that the 'goodness' of any situation (e.g. resource allocation) should be judged solely on the basis of the utility levels attained by individuals in that situation (i.e. excluding all non-utility aspects of the situation).

Culyer (1989, 90) attempted to develop an alternative extra-welfarist framework that embodies the centrality of health as the outcome of concern. Building on Sen's notion of extra-welfarism, Culyer argues that normative evaluation should focus on the characteristics of people, including non-utility characteristics. Ill-health creates a need for healthcare, which restores a person's health or forestalls a worsening health. By this reasoning, extra-welfarism integrates the two key concepts viz. the concept of need (as opposed to demand) and health (as opposed to utility) as a final outcome of concern. This alternative approach calls for optimum solutions (i.e. equilibria) different from the neoclassical framework. With changed demand and utility functions, an individual's optimum level of work effort will change. So is the producer's neoclassical profit maximising behaviour which also will change. The concept of efficiency wage hypothesis becomes relevant in this regard.

7.4.1 Efficiency Wage Hypothesis

The **efficiency wage** hypothesis argues that wages, at least in some markets, are decided in a way that is not market-clearing. Specifically, it points to the incentive for managers to pay their employees more than the market-clearing wage in order to increase their productivity or efficiency, or reduce costs. The increased labour productivity

and/or decreased costs pays for the higher wages. Paul Krugman explains how the efficiency wage theory comes into play in a real society. The productivity $E(w)$ of individual workers being a function of their wage w, their total productivity is the sum of the individual productivities. Hence the sales (V) of the firm to which the workers belong becomes a function of employment (L) and the individual productivity. The firm's profit P therefore is:

$$P = V\left(LE\right) - wL. \tag{7.1}$$

If we assume that higher is the wage of the workers, higher would be the individual's motivation for higher productivity, we have: $\frac{dE}{dw} > 0$. If the employment is to be so chosen that the profit is maximised, under the optimising condition, we should have:

$$dP = \frac{\partial V}{\partial\left(LE\right)} LdE - Ldw, \tag{7.2}$$

$$\text{i.e. } \frac{dP}{dw} = \frac{\partial V}{\partial E}\frac{dE}{dw} - L. \tag{7.3}$$

The gradient $\frac{\partial V}{\partial E}$ (i.e. the slope) is positive because higher the individual's productivity, higher is the sales. Since $\frac{dP}{dw}$ cannot be negative (because of the optimising condition), we have $dP/dw \geq 0$. This means that if the firm increases their wage, their profit also would increase up to an optimum level. Thus, the efficiency wage theory motivates the owners of the firm to raise the wages for increasing the profit of the firm.

The term 'efficiency-wages' (or 'efficiency-earnings') was first introduced by Alfred Marshall to denote the wage per efficiency unit of labour. Thus, in Marshallian efficiency wages would make employers pay different wages to workers for different levels of efficiency. The modern use of the term also refers to the idea that higher wages would increase the efficiency of the workers through various channels making it worthwhile for the employers to offer wages that exceed a market-clearing level. The concept of efficiency wage makes room for the extra welfarist approach in health sector. That is, extra payment to the workers as fringe benefit can increase the level of her/his health and hence productivity.

7.4.2 Health and Human Development

The right to good health has long been recognised as a fundamental part of human rights for a dignified life. The 1948 Universal Declaration of Human Rights mentioned health as a part of the right to an adequate standard of living. The 1966 International Covenant on Economic, Social and Cultural Rights reiterated this right. Since then, many other international human rights treaties have contributed to recognizing this right. At national level, each country has ratified at least one international human rights treaty recognizing the right to health. The UN Committee on Economic, Social and Cultural Rights adopted in the year 2000 a General Comment on the Right to Health. It set out the right to health, extending it to not only the timely and appropriate healthcare but also to its underlying determinants like access to safe and potable water and adequate sanitation, adequate supply of safe food, nutrition, housing, etc. According to this Comment, the

right to health contains four elements viz. (i) *availability* i.e. functioning public health and healthcare facilities, goods and services, as well as programmes in sufficient quantity; (ii) *accessibility* i.e. health facilities, goods and services accessible to everyone with four overlapping dimensions viz. (a) non-discrimination, (b) physical accessibility, (c) economical accessibility (affordability) and (d) information accessibility; (iii) *acceptability* i.e. all health facilities, goods and services must be culturally appropriate and be sensitive to gender and life-cycle requirements; and (iv) *quality* i.e. health facilities must be scientifically and medically appropriate and of good quality. Such a focus on health has contributed to changing the definition of development by including non-economic factors like health and education being considered as critical components of development. Three out of eight Millenium Development Goals (2000) also are related to health. By 2015, the world has achieved 47 percent reduction in maternal mortality and 49 percent reduction in child mortality. The Sustainable Development Goals (2015) also emphasise improved nutrition, healthy life for all and availability and sustainable management of water and sanitation for all as its critical components.

Check Your Progress 2 [answer the questions in about 100 words within the space given]

1. How was the focus of HDI fundamentally different from that of the PQLI?

   .....

   .....

   .....

   .....

2. How has the computation of HDI been markedly improved in the more recent years?

   .....

   .....

   .....

   .....

3. In what way the Culver's approach incorporates 'health' as an important dimension of welfare?

   .....

   .....

   .....

   .....

4. State the overlapping concerns in the Marshallian efficiency wage concept and Krugman's 'efficiency wage' theory.

   .....

   .....

   .....

   .....

5. What are the four elements of health incorporated in the General Comment of UN Committee in the year 2000?

.....

.....

.....

.....

7.5 EXTERNALITIES IN HEALTH

Any spill over effect of consumption or production is called externality. If the consumption of any good by an individual has direct effects on the utility of another individual, it is said to give rise to an externality with the market mechanism not leading to a Pareto-optimal allocation (such as in a situation of perfect competition). In the case of health goods, two types of positive externalities can be there. It can either: (i) directly improve the health of the other individual (called physical externality), or, (ii) simply lead to increased satisfaction for the other individual (called psychological externality). Physical externalities arise in the context of communicable diseases. If individual *i* seeks treatment or tries to prevent oneself from contracting such a disease, the probability of individual *j* getting infected is reduced. Thus, in a two person case involving *i* and *j* only, a Pareto optimum could easily be brought about by *j* voluntarily subsidizing the vaccination costs of *i*. In reality, externalities are spread over many people, giving rise to their collective good property. A 'collective good' (also called 'public good') being characterized by non-rivalry in consumption, a consumer *i* obtaining a unit of good *h,* say vaccination against a communicable disease, would affect several other individuals positively without a diminishing effect on *i*'s enjoyment of the good. Further, such collective goods also have non-excludability characteristics (i.e. nobody can be barred from using them even though she or he has not contributed to their provision). Psychological externalities refer to instances where most people would be distressed when seeing others die from starvation or lack of medical care so long as their suffering is not caused by their own actions.

7.5.1 Health: Public Good or Merit Good

The above discussion conveys that some component of healthcare is characterised by the properties of non-rivalry and non-excludability. Such considerations suggest that compulsory vaccination to prevent communicable diseases should be provided by the government using the general tax base. In other words, healthcare has a component of public good in it. But some parts of it like hospital services, physician services, medicines, etc. cannot be treated as public good as they are characterised by neither non-rivalry nor non-excludability. The part of the healthcare which possess the character of public good are in almost all cases provided and distributed by the government. This is evidenced by the public health systems which exists in all the countries. This is where the concept of a *merit good*, introduced by Richard Musgrave, becomes important. A merit good is a commodity which an individual or society should have on the basis of a concept of need, rather than the ability and willingness to pay. The concept lies behind many economic actions of governments which are not performed for financial reasons or for supporting incomes (e.g. tax rebates). Examples include the provision of food stamps to support nutrition, the delivery of health services to improve quality of life and reduce morbidity, subsidized housing, etc. In the healthcare sector, merit goods which provide services universally to everyone comes close to the concept of primary goods found in the works of philosophers like Rawls in discussions of 'social

inclusion'. On the 'supply' side, it is argued that there should be more of implicit redistribution (via the provision of primary healthcare services), rather than explicit redistribution through income. Alternatively, it is argued that society in general may be in a better position to determine what individuals need than the individuals themselves. It is in this sense that some economists argue that though health is not entirely a public good, healthcare in general should not be totally left to the market and should be provided and distributed by the government.

7.5.2 Distinction Between Private Demand and Social Demand for Health

Due to the positive externality characteristics, the social marginal benefit in health exceeds the private marginal benefit in healthcare market i.e. the society is better off when its people are in good health. Consuming healthcare services can generate external benefits to society (like preventing communicable diseases) and easing the emotional costs of those who live with or take care of the sick. Healthy individuals also are more productive which can lead to gains for society. In a market with externalities, the private market outcome does not coincide with the socially optimal outcome. The existence of external benefits from the consumption and production of healthcare services leads to an outcome in which the socially optimum amount of healthcare is greater than the level achieved in a private market. Figure 7.1 illustrates this situation.

The external benefit is shown as the vertical difference between the market demand curve d_p and the social demand curve d_s. Because individuals do not take external benefits into account when making consumption decisions, the market demand curve intersects supply at a lower quantity than the social demand curve. The equilibrium private output, q_p, is less than the socially optimal output q_s, which indicates a market failure. One of the consequences of market failure is that not all socially beneficial transactions are realized. The root of the problem is that external benefits are not received by those who consume the service, and thus they are not willing to pay for it. This is one reason why the government and other institutions provide subsidies for healthcare coverage and for healthcare research and development.

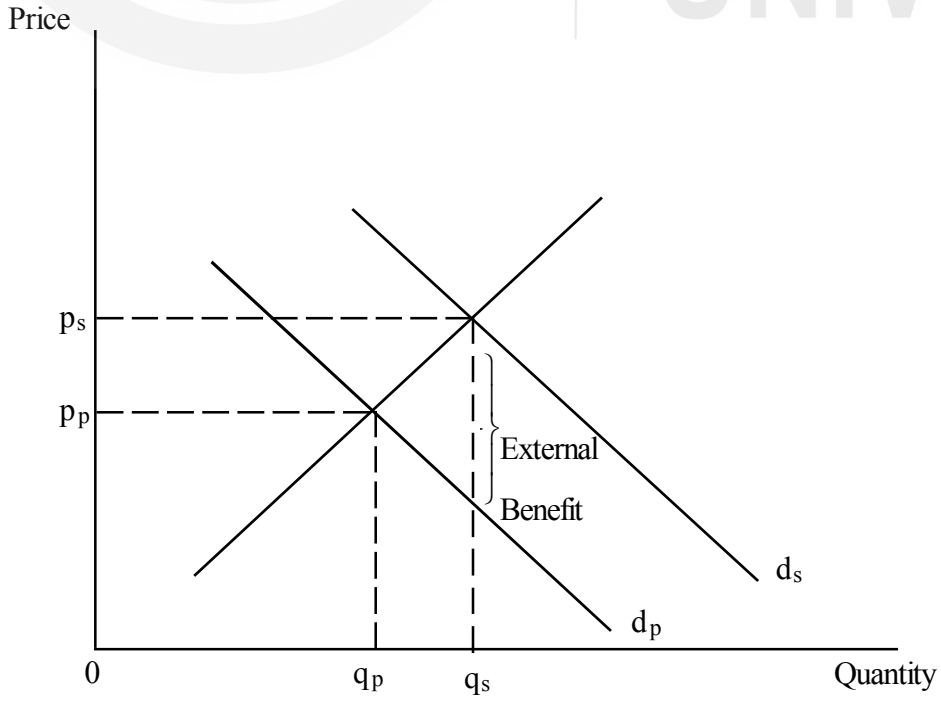


Figure 7.3: Private Demand and Social Demand for Health

7.6 ROLE OF HEALTH IN ECONOMIC DEVELOPMENT

Health had never been considered important in early development planning because it was viewed as a consumption rather than investment commodity. It later acquired its importance when it was looked at as *human capital*. Human factors, such as the level of health of the labour force, have been shown to be of great importance, relative to the accumulation of physical capital, in explaining growth in industrial countries as well as in the developing world. Expenditures on health, as well as on education, are considered as investments in human resources contributing to productive capacity. A healthier labour force will suffer less disability and will work more effectively and more steadily resulting in increased productivity and rising per capita income.

Better health raises per capita income also through a number of other channels. One way is by altering decisions about expenditures and savings over the life cycle. The idea of planning for retirement occurs only when mortality rates become low enough for retirement to be a realistic prospect. Rising longevity in developing countries has opened a new incentive for the current generation to save — an incentive that can have dramatic effects on national saving rates. Although this saving boom lasts for only one generation and is offset by the needs of the elderly once population aging occurs, it can substantially boost investment and economic growth rates while it lasts. Another channel is by encouraging foreign direct investment. Investors shun environments where the labour force suffers a heavy disease burden. Endemic diseases can also deny humans access to land or other natural resources as economies adjust gradually to their steady-state output level over time. In this case, countries that have high levels of health but low levels of income may experience relatively faster economic growth as their income adjusts. Evidence from cross-country growth regressions suggests that the contribution of better health to economic growth is large. Once countries reach their steady-state level of income, growth slows. Indeed, among all such well-established influences as the initial level of income per capita, geographical location, institutional environment, economic policy, initial level of education, and investments in education, the initial health of a population has been identified as one of the most robust and potent drivers of economic growth. For instance, Bloom, Canning, and Sevilla (Harvard University) found that one extra year of life expectancy raises the steady-state GDP per capita by about 4 percent.

Value of lost output due to illness: At the macroeconomic level, the impact of illness on gross domestic product (GDP), both now and in the future, is of great relevance. This is measurable and gives a clear idea about the extent to which health influences economic growth and development. A macroeconomic approach to assessing the impact of ill-health should be concerned with establishing the aggregate impact of disease and injury across different economic agents on three areas related to economic welfare: *non-health consumption possibilities, leisure time and health status*. Most economic impact studies at the societal level have focused on gross domestic product (GDP), which represents only market related consumption opportunities. While this has a clear meaning, it is important to note that GDP also includes expenditure on health goods and services. So this component should be omitted from the GDP so that the focus of analysis is directed towards properly establishing the present value of discounted aggregate flows of current and future consumption of non-health related goods and services linked to disease. Key channels through which disease or injury can impact on macroeconomic performance or output include increased health expenditures, labour and productivity losses, and reduced investment in human and physical capital formation.

Several modelling approaches have been used to estimate the aggregate economic impact of disease. One of them is the cost-of-illness or human capital approach, which combines 'direct costs' (medical care, travel costs, etc.) and 'indirect costs' (the value of lost production because of reduced working time) into an overall estimate of economic impact on society, often expressed as a percentage of current GDP. Although the cost-of-illness approach focuses on the societal impact of disease or injury, it provides only a partial picture of the true macroeconomic impact of disease as it fails to consider the contribution of depleted capital accumulation, investment in human capital, and demographic change to diminished economic growth.

Check Your Progress 3 [answer the questions in about 100 words within the space given]

1. What is 'externality'? Distinguish between 'physical externality' and 'psychological externality'.

   .....

   .....

   .....

   .....

2. Why is it argued that the provision of 'healthcare' should not be entirely left to the private sector?

   .....

   .....

   .....

   .....

3. State some important ways by which better health contributes to the society.

   .....

   .....

   .....

   .....

4. Do you agree that the expenditure on health goods and services should be discounted from the GDP estimates? Why?

   .....

   .....

   .....

   .....

7.7 DEMAND FOR HEALTH VERSUS TRADITIONAL DEMAND FUNCTION

Though every individual demands good health, construction of a demand function for health is not easy. This is because, first of all, health is not a tangible variable. It is a physical and mental state of well being, and hence, cannot be cardinally measured. Secondly, it is subject to interpersonal variation i.e. the concept of good health can vary from person to person. Thirdly, since every individual is a producer as well as a consumer

of good health, the demand for and supply of health is generated within the human body. So market for health actually works within an individual's body structure. In view of these factors, demand for health cannot be explained by the traditional theory of demand.

7.7.1 Becker's Theory of Behaviour

Gary Becker (1965) proposed a revised theory of demand where he: (i) incorporated time, or more specifically leisure, as a determinant of utility, and resultantly of demand, and, (ii) considered household to act both as the producer and the consumer. With this, he opined that the consumer's demand for commodities is actually a *derived demand*. In its broad outline, his approach views commodities as primary objects of consumer choice from which utility is directly obtained. These commodities are produced by the consumer himself/herself through the productive activity of combining the purchased market goods and services with some of the household's own time. In this framework, all market goods are inputs used in production processes of the non-market sector. The consumer's demand for these market goods is a derived demand analogous to the derived demand by a firm for any factor of production. Formally, therefore, the household's utility function is:

$$U = U(Z_1, Z_2, \ldots, Z_n), \tag{7.4}$$

where Z_i, stands for both the services and the quantity of the commodity Z_i produced by the household using a vector of market goods x_i, and a second vector of quantities of its own time t_i. Therefore, the production function of Z_i is:

$$Z_i = Z_i(x_i, t_i; E), \tag{7.5}$$

where E is a vector of variables representing the environment in which the production takes place. The utility function is maximized subject to the production function constraints and a constraint on the household's available time:

$$T = t_w + \sum_{i=1}^{n} t_i, \tag{7.6}$$

as well as the usual income constraint:

$$I = \sum_{i=1}^{n} p_i x_i, \tag{7.7}$$

where t_w and t_i are the household's time spent in the labour market and in producing Z_i respectively and p_i and x_i are the price and quantity of the commodity-input used in producing Z_i. The time and money income constraints can be collapsed into a single resource constraint on the household's full income S as:

$$S = wT + V = \sum_i (wt_i + p_i x_i) \tag{7.8}$$

where w is the wage rate, assumed to be constant, and V is the household's non-wage income. The importance of this concept of full income is that it embodies both the time and money income constraints and its magnitude is independent of the fraction of time the household chooses to allocate to income-earning activities. The utility function (7.4) is maximized subject to the constraints of the production function (7.5) and full income constraint (7.8). The Lagrangian for this is expressed as:

$$L = U(Z_1, Z_2, \ldots, Z_n) - \ddot{e}[\sum_i (wt_i + p_i x_i) - S] \tag{7.9}$$

The first order conditions for maximization with respect to the commodities imply:

$$\frac{MU_i}{MU_j} = \frac{w\frac{dt_i}{dZ_i}+p_i\frac{dx_i}{dZ_i}}{w\frac{dt_j}{dZ_j}+p_j\frac{dx_j}{dZ_j}} = \frac{\pi_i}{\pi_j} \quad , \qquad (7.10)$$

The ratio of the marginal utilities of any two commodities Z_i, and Z_j i.e. MU_i/MU_j, will be equal to the ratio of their marginal costs, π_i / π_j, where the derivatives in equation (7.10) are marginal input-output coefficients. These marginal costs are also the shadow prices of Z_i, which are determined by the prices of market goods and time, and by the productivity of each in producing Z_i. Similarly, equation (7.9) can be differentiated with respect to all factors of production to determine their optimal use as:

$$\frac{\frac{\delta U}{\delta Z_i}\frac{\delta Z_i}{\delta f_{ik}}}{\frac{\delta U}{\delta Z_j}\frac{\delta Z_j}{\delta f_{jl}}} = \frac{MU_i MP_{ik}}{MU_j MP_{jl}} = \frac{pf_{ik}}{pf_{jl}} \quad , \qquad (7.11)$$

Where f_{ik} is the factor k (either goods or time) used in producing Z_i and f_{jl} is the factor 'l' used in producing Z_j. When both factors are used in the same production function (i = j), the condition reduces to equality of the ratio of marginal products to the ratio of the factor prices. Alternatively, if k = 1 (i.e. if the same factor input is used in several production functions), equation (7.11) implies that the factor will be allocated among commodities to equalize the utility value of its marginal product in the production of different commodities. Changes in environment E, may affect factor prices and the input coefficients and thereby alter a commodity's relative price $ð_i/ð$, where ð is an index of all commodity prices. It may also affect the general price level ð itself by raising or lowering the average $ð_i$ as a whole. A change in the average price of all commodities is comparable to a change in the household's cost of living or to a change in its opportunity set. So full money income, S, can be converted into full 'real' income by dividing full money income by the general price level i.e. S/ð.

This new approach of derived demand has been successful in explaining various human demand behaviour related to non-market events and has since been applied to explain the same. One such category of applications stems from the close relationship between this framework and the growing human capital literature. With the consumer's own time introduced into the analysis of consumption behaviour, the productivity of his time, and hence the human capital embodied in the individual consumer, becomes an important object of analysis. Thus, not only the productivity of non-market time affect consumer behaviour, the effect of human capital on non-market production is itself an important source of the yield on investments in human capital. Hence, the analysis of the incentive to invest in education, health, migration, search and so forth, in principle, incorporates the 'consumption returns' for the non-market benefits accruing to the investment. This approach is thus helpful in explaining the consumer behaviour by incorporating health as an important variable.

7.7.2 Grossman's Human Capital Model of Demand for Health

If one is to accept health as an asset capable of being produced, then the consumption of health capital gives reason to view its reduction owing to aging and lifestyle, resulting in a net change in the capital stock of 'health' over time. Investment is achieved by the input of (curative) medical services and by one's own effort in preventive and maintenance

efforts. The return on the stock of health capital can therefore be alternatively viewed as spending less time in habits leading to bad health. Healthy time gained can increase utility directly or indirectly due to higher labour income and thus higher consumption. Rational individuals will maximize utility by the optimal management of their stock of health over life time. Grossman (1972) investigated this dynamic optimisation problem using optimal control theory as follows. Consider an individual with a planning horizon of two time periods. During each period, she or he experiences a non-negative amount of sick time t^s, which if is lower, the health stock H would be larger. In other words, healthy time constitutes the non-tradable return of the unobserved stock of health. The individual derives positive utility from consumption goods X while deriving disutility from sick time $t^s(H)$. The utility function defined over these arguments is thus time independent (i.e. the marginal rate of substitution between sick time and consumption does not change with aging). Future utility is discounted by a subjective factor $\beta \leq 1$. Thus, the individual maximizes the discounted utility U specified as :

$$U = U[t^s(H_0), X_0] + \beta\ U[t^s(H_1), X_1], \tag{7.12}$$

with $$\frac{\delta U}{\delta t^s} < 0, \frac{\delta^2 U}{\delta (t^s)^2} > 0, \frac{\delta U}{\delta X} > 0, \frac{\delta^2 U}{\delta X^2} < 0, \frac{\delta t^s}{\delta H} < 0$$

The crucial component of the Grossman specification is the equation that defines the change in the health stock over time. On the one hand, health capital depreciates at a rate δ, causing health to worsen over time. On the other hand, the individual can increase health capital by investing I. This involves the purchase of medical services M or spending t^I units of time on preventive effort. We thus have:

$$H_1 = H_0(1-\delta) + I(M_0, t^I), \tag{7.13}$$

with $$\frac{\delta I}{\delta M} > 0, \frac{\delta^2 I}{\delta M^2} < 0, \frac{\delta I}{\delta t^I} > 0, \frac{\delta^2 I}{\delta (t^I)^2} < 0$$

Equation (7.13) constitutes a constraint that will enter the individual's maximization problem. However, there is not only a change in health over time but also in wealth and wisdom (i.e. skills). In particular, savings S_0 achieved in the first period are available for consumption in the second period. Savings yield interest r to become $(1+r)S_0$. By equation (7.13), investment in health occurs during the initial period only.

Abstracting from health insurance, if 'p' denotes the price of medical care M, then healthcare expenditure amounting to pM has to be financed out of labour income or initial wealth A_0, or w_0 denoting the wage in the initial period. Since consumption must be positive in both periods, and total time available can be normalized at 1 in both periods, altogether the following budget constraint holds after discounting for the present value:

$$A_0 + w_0(1-t^s(H_0) - t^I) + \frac{w_1 + (1-t_1^s(H_1))}{R} = pM + cX_0 + \frac{cX_1}{R} \tag{7.14}$$

To solve this maximization problem, we consider the Lagrangian function,

$$L\left(H_1, t^1, M, X_0, X_1\right) = U\left[t^s\left(H_0\right), X_0\right] + \beta U\left[t^s\left(H_1\right), X_1\right] +$$

$$\mu\left[H_0\left(1-\delta\right) + I\left(M, t^1\right) - H_1\right] +$$

$$\lambda\left[A_0 + w_0\left(1 - t^s\left(H_0\right) - t^I\right) + \frac{w_1\left(1 - t_1^s\left(H_1\right)\right)}{R} - pM - cX_0 - \frac{cX_1}{R}\right]$$

The Lagrangian multipliers μ, $\lambda > 0$ indicates the extent to which a relaxation of the pertinent constraints would improve the achievement of the overall objective as measured by discounted utility. First-order conditions for an initial optimum are obtained by setting the derivatives with respect to all decision variables equal to zero. The derivatives with respect to μ and λ simply ensure that the constraints (7.13) and (7.14) are satisfied. Treating H_0 as given, the first order conditions are:

$$\frac{\partial L}{\partial H_1} = \beta, \frac{\partial U}{\partial t^s}\frac{\partial t^s}{\partial H_1} - \frac{\lambda}{R} w_1 \frac{\partial t^s}{\partial H_1} - \mu = 0 \tag{7.15}$$

$$\frac{\partial L}{\partial t^I} = \mu \frac{\partial I}{\partial t^I} - \lambda w_0 = 0 \tag{7.16}$$

$$\frac{\partial L}{\partial M} = \mu \frac{\partial I}{\partial M} - \lambda p = 0 \tag{7.17}$$

$$\frac{\partial L}{\partial X_0} = \frac{\partial U}{\partial X_0} - \lambda c = 0 \tag{7.18}$$

$$\frac{\partial L}{\partial X_1} = \beta \frac{\partial U}{\partial X_1} - \frac{\lambda}{R} c = 0 \tag{7.19}$$

These conditions can be made easier by first dividing equation (7.16) by equation (7.17) to get:

$$\frac{\partial I / \partial t^1}{\partial I / \partial M} = \frac{w_0}{p} \tag{7.20}$$

Secondly, dividing equation (7.18) by equation (7.19) we get,

$$\frac{\partial U / \partial X_0}{\partial U / \partial X_1} = \beta R \tag{7.21}$$

Equation (7.15) can be rewritten by solving (7.19) for λ / R as :

$$-\beta \frac{\partial t^s}{\partial H_1}\left[\frac{w_1}{c}\frac{\partial U}{\partial X_1} - \frac{\partial U}{\partial t^s}\right] = \mu \tag{7.22}$$

Using (7.17) and (7.18), we obtain,

$$\mu = \frac{\partial U / \partial X_0}{\partial I / \partial M}\frac{p}{c} \tag{7.23}$$

Substituting (7.22) in (7.23) yields:

$$-\beta \frac{\partial t^s}{\partial H_1}\left[\frac{w_1}{c}\frac{\partial U}{\partial X_1} - \frac{\partial U}{\partial t^s}\right] = \frac{\partial U / \partial X_0}{\partial I / \partial M}\frac{p}{c} \tag{7.24}$$

The above condition requires the marginal utility of an investment in health to be equal to its marginal cost. In the Grossman model, health and wealth constitute two interrelated assets whose values are optimally managed over time by the individual. In the case of health, the marginal utility of holding an extra unit of the stock has a consumption and an investment component. Their sum has to be equal to the marginal cost associated with holding an additional unit of health stock.

7.8 SUPPLY FACTORS AFFECTING DEMAND FOR HEALTH

The unique characteristics of healthcare as a commodity lead to a complete lack of consumer sovereignty making room for the third party to enter into the healthcare market. Consequently, the supply side factors play a significant role in influencing the demand behaviour where physicians, pharmaceutical companies and insurance companies are the key players.

7.8.1 Role of Physicians

Due to lack of knowledge about the commodity, consumers do not know how much or what medicine they should consume, what kind of treatment they should avail or whether the treatment taken will cure them. This makes the physician the first agent to decide upon diagnosis, treatment, prescription, and referral to other providers of medical services (e.g. specialists, hospitals, pharmacists). Consequently, the relationship between supplier (physician) and consumer (patient) will play a key role in the existence of supplier-induced demand in the healthcare market.

The most striking feature of the physician-patient relationship is the patients' *incomplete information* about true needs. Therefore, they delegate the choice of treatment to the better-informed doctor while deciding merely on whether or not to follow the doctor's advice and comply with his or her prescriptions. Even in this decision, patients are not completely sovereign, as the patient's confidence in the doctor's competence is an important precondition for successful treatment. This means that the demand curve actually mirrors the *decisions of suppliers* and not of consumers of healthcare. In the limit, therefore, the demand for medical services is supplier *determined*. This fact remains without consequences as long as the doctors act as *perfect agents of their patients* i.e. choose what patients would have chosen had they possessed the necessary technical knowledge. It becomes a problem as soon as doctors' decisions on behalf of their patients are influenced by their own interests. If doctors vary the information provided to their patients systematically, with an increase in physician density, so as to secure their own benefits, then supplier-*determined* demand becomes supplier-*induced* demand (SID) (Figure 7.2). Full insurance coverage of patients is assumed here so that the demand does not depend on price. D_0 denotes the 'primary demand' i.e. the volume of services that meets the ideal standard of medical care given the number of patients, prevalence of health conditions and the decisions of patients to contact the doctor. p_0 is the regulated fixed unit price, assumed to equate primary demand and supply of medical services, given the initial physician density (i.e. the S_0 curve). Now if, due to an increase in the number of physicians, the supply curve shifts outward (to S_1), and providers had no direct influence on demand (or they act as perfect agents of their patients), the demand curve would remain stable. At the regulated price, the quantity of services actually transacted would remain constant at $D_0 = M_0$. In this case, providers would produce fewer services than they would like to at this price i.e. there is underutilization of their capacity at the level of $M_1 - M_0$. If, however, physicians do *not* act as *perfect* agents, but advise their patients medically unnecessary procedures/

medicines to pursue their own interests, or marginally effective or totally ineffective services, then $M_1 = S_1(p_0)$ of supply *will actually be 'demanded'*. In this case, the demand curve would shift from D_0 to D_1 i.e. to the point where the quantity demanded equals the supply S_1. According to the 'SID' hypothesis, the amount of services transacted in the market is exclusively determined by supply as desired by the physicians. The 'demand' required to match the supply is artificially created (induced) through the advice to patients, whether medically appropriate or not.

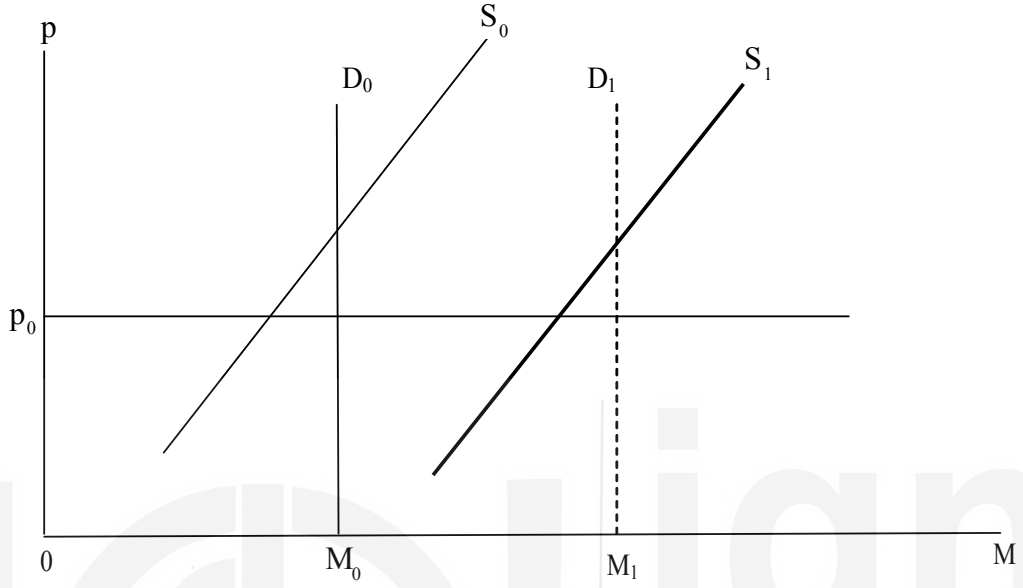


Figure 7.2 : Supplier-Induced Demand as a Response to an Increase in Physician Density

7.8.2 Role of Pharmaceutical Companies

Drugs play an important role in the healthcare market. Here patients cannot buy the drugs from the pharmaceuticals directly as in other commodities: a third party is involved i.e. the physicians who can influence the quantity of drugs to be purchased by the patients. Due to the uncertainty of occurrence of ill health and asymmetric information between patients and physicians, the market for medical care creates a scope for the pharmaceutical companies (suppliers) to enjoy more market power. Of all the goods and services traded in the market economy, pharmaceuticals are perhaps the most contentious. They carry a moral weight that most privately traded goods do not due to a belief that people have a right to healthcare. Since innovation accounts for most of the cost of production, the price of drugs is much higher than their cost of manufacture, making them unaffordable to many poor people. Moreover, promotional cost of drugs is so high that it constitutes a considerable portion of the drug price. Firms protect the intellectual property (IP) of drugs suing those who try to manufacture and sell the patented drugs cheaply. Due to these reasons, pharmaceutical companies are widely regarded as vampires who exploit the sick and ignore the sufferings of the poor. At local level, the asymmetric information about the drug keeps the buyers (patients) at receiving end. The unhealthy nexus between the private practitioners and the representatives of the pharmaceutical companies often leads to consumption of marginally necessary, or totally unnecessary drugs, making it another example of supplier induced demand. This again leads to welfare loss.

7.8.3 Role of Health Insurance

Due to asymmetric information between patients and physicians, many of the choices individuals make as healthcare consumers, as well as providers, involve a substantial

amount of uncertainty. Just as, for an individual consumer, many medical illnesses occur randomly, and therefore the timing and amount of medical expenditures are uncertain, from the healthcare provider's perspective, patient load and types of treatment are unknown before they actually occur. Since they are unpredictable, there is a substantial degree of risk. Since most people generally dislike risk, they are willing to pay some amount of money to avoid it. This is where the health insurers enter the market as a third party provider.

Consumers actually purchase a pooling arrangement when they buy a policy from an insurance company that helps to mitigate some of the risk associated with potential losses. Standard statistical theory suggests that the expected value, m, of a distribution of outcomes, such as losses, can be computed as the sum of the weighted values of the outcomes, *Li*, with the probabilities, p_i, serving as the weights. The expected loss, therefore, equals:

$$\mu = p_1L_1 + p_2L_2 \qquad (7.25)$$

However, since people are also concerned about the variability of the expected loss, it stands to reason that a distribution of likely outcomes involves greater risk when more variability exists around the expected value. Since the variance increases when the actual outcomes, L_i, are further away from the expected outcome μ, both the expected loss and its variance (or standard deviation) can be thought of as measures of risk. Generally speaking, more risk is associated with a higher expected loss and when the distribution of the expected loss, or standard deviation, exhibits wider variability. Risk aversion occurs when people receive disutility from taking on additional risk and are therefore willing to pay to avoid it or are made to pay to accept it. But when people face the same distribution of outcomes, a pooling arrangement is not about reducing the expected loss; rather it is about reducing the standard deviation or variability of the loss. In addition, the variability of the expected loss decreases as more individuals with similar individual loss distributions join a pooling arrangement. Assuming that the losses are not perfectly correlated, more individuals joining the pooling arrangement help reduce the probability of the extreme outcomes occurring, thereby making the expected loss less variable and more predictable. In other words, consumers typically gain from entering into pooling arrangements as pooling helps reduce the variability of the expected loss. Although the consumers benefit when they enter into a medical expense pool, the individual loss function associated with medical expenses is so heavily skewed toward the left, that only a very few people will actually incur large medical expenses in the absence of insurance. But as more and more risk-averse individuals purchase health insurance, they manage to avoid or transfer some of their risk by replacing their individual loss distributions with the average loss distribution of the group. Since compared to the individual loss distributions, the average loss distribution involves less variability around the expected loss, it results in less risk faced by an individual when engaged in a pooling arrangement.

John Nyman (2003) advanced an alternative argument about why people desire medical insurance. Nyman pointed out that many medical interventions, such as a liver transplant or coronary bypass, cost more than most people hold in terms of their net worth (i.e. value of assets less the value of liabilities). In addition, banks are reluctant to loan out money for a potentially lifesaving medical intervention as they are unsure about its repayment. Thus, conventional insurance theory treats medical insurance as reducing the representative consumer's out-of-pocket expenditure on medical care. The lower out-of-pocket expenditure, in turn, creates a downward movement of the demand curve leading to additional units of medical care being demanded for which their marginal

costs exceed their marginal benefits. This is the situation described as the problem of 'moral hazard'.

Figure 7.3 brings out the economic reasoning behind this concept of moral hazard where the typical consumer's demand, *d*, for medical care is downward sloping. The line MC, shown as a straight line, reflects the marginal cost of delivery of medical care, assumed to be constant with respect to the amount of medical care produced and determined in the market. For the uninsured individual, the consumer equilibrium occurs where MC and demand curve intersects at a price and quantity, say p_1 and q_1. The amount of medical care consumed is considered efficient because for every unit between the origin and q_1, willingness to pay or marginal benefit, as revealed by demand, never falls below MC. However, if the consumer purchases full insurance coverage, complete insurance coverage lowers the consumer's out of-pocket expenditure bringing it below the demand curve (i.e. from p_1 to 0). At this point of consumer equilibrium, because each additional unit of medical care between q_1 and q_0 generates more costs than benefits at the margin, q_0 amount of medical care is consumed with a welfare loss. In other words, if they had to pay the full price for the medical care i.e. without insurance coverage, they would not have purchased the additional units (q_0—q_1). These additional units of medical care may reflect extra visits to doctors or longer stays in hospitals than medically necessary. Thus, conventional theory treats the transition from uninsured to insured status as resulting in the consumption of frivolous or unnecessary medical care.

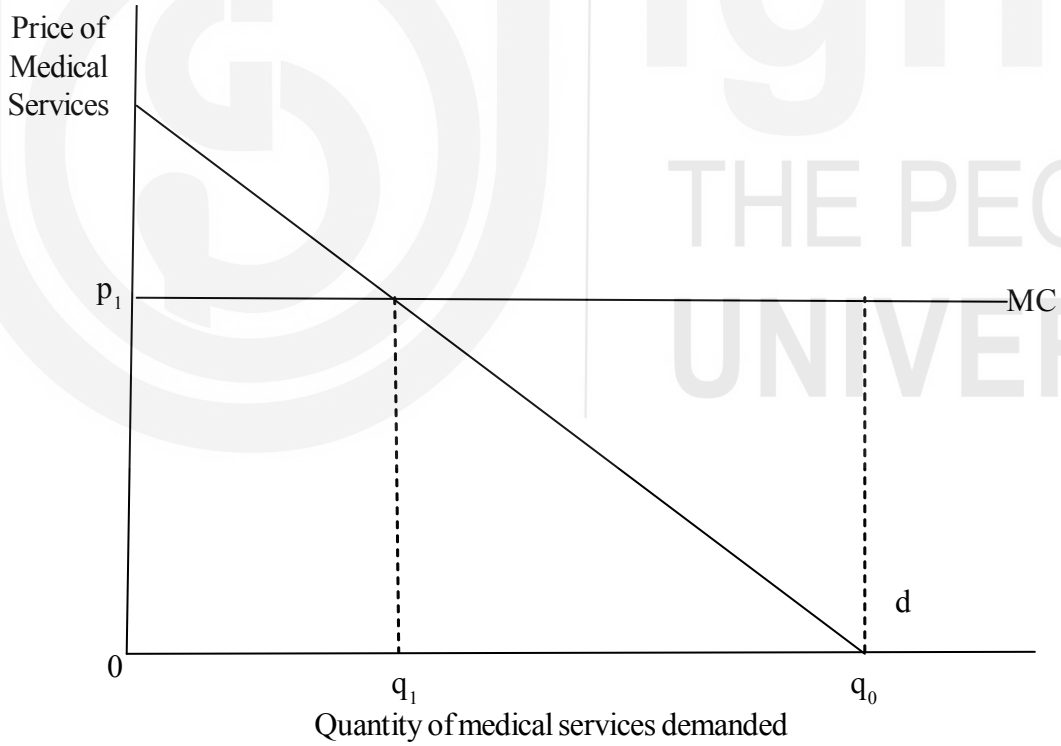


Figure 7.3 Moral Hazard in Insurance Market

Check Your Progress 4 [answer the questions in about 100 words within the space given]

1. State the four reasons why the traditional theory of demand cannot be used to explain the demand for health.

   .....

   .....

.....

.....

2. What is the fundamental difference introduced by Becker in improving the approach to describing the consumer demand behaviour?

.....

.....

.....

.....

3. How has the approach of 'derived demand' been useful in explaining the consumer/ human demand behaviour related to non-market events?

.....

.....

.....

.....

4. State the discounted utility function specified by Grossman. What is a crucial component of this specification?

.....

.....

.....

.....

5. What is Supplier Induced Demand in the context of healthcare? Give examples from major segments of players to illustrate how the marginalised and the poor in the society suffer due to distorted market conditions.

.....

.....

.....

.....

6. How does more risk averse people joining a health insurance plan results in reduced risk being faced by individual insurers?

.....

.....

.....

.....

.....

7.9 LET US SUM UP

Health, as a commodity, has some unique characteristics. For this reason, neoclassical demand function cannot explain the health behaviour in both micro and macro perspective. Health status or level of health is not also easily measurable. For this reason, various health indicators are used to make it quantifiable. Characteristically, health has both the components of a consumption good and an investment good and that is why an individual is considered to both produce and consume health. In light of this, the concept of *derived demand* is more appropriate for healthcare. Health had never been considered important in early development planning because it was viewed as a consumption rather than an investment commodity. But it later acquired its importance when it was looked at as *human capital*. Human factors, such as the level of health of the labour force, have been of great importance, relative to the accumulation of physical capital, in explaining the growth in industrial countries as well as in the developing world. A sick population may lead to more health expenditure, loss of productivity and income, thereby reducing the growth rate of the economy. That is why, economic development is argued to be interlinked with health and education and is now measured by human development index. However, from society's point of view, health, and more specifically primary health, has in it a public good component since it has positive externality on others. Demand for health is also characterised by asymmetric information between the consumer and the supplier leading to *supplier induced demand*. Supply side players like physicians and other service providers as well as pharmaceutical companies may and generally influence the demand for healthcare services more than the required thus leading to welfare loss. Health is also argued to have merit good property which differentiates the private demand from social demand. These characteristics actually call for the state intervention in healthcare market. This is the reason for existence of public health systems, though in different forms and extent, in every economy.

7.10 KEY WORDS

Human Capital : *Human capital* is a collection of resources — knowledge, talents, skills, abilities, experience, intelligence, training, judgment, and wisdom possessed individually and collectively in a population. These resources are the total capacity of the people that represents a form of wealth which can be directed to accomplish the goals of the nation or state.

Supplier-Induced Demand : A hypothesis about firm behaviour that propose that physicians and other service-providers pursuing their own economic self-interests, taking advantage of the asymmetry of information about medical care, persuade their patients to consume more medical care than is necessary. It is defined as the capacity of a supplier (physician) to exploit his informational advantage to induce a higher level of demand for the good or service than what the consumer would have demanded if both the physician and the patient had the same amount of information. In healthcare market, supplier-

induced demand translates to the capacity of the physician to induce patients to demand healthcare that they do not necessarily need.

Merit Good : Is a commodity which an individual or society should have on the basis of some concept of need, rather than the ability and willingness to pay.

7.11 SOME USEFUL BOOKS AND REFERENCES

1) Balarajan Y., Selvaraj S and Subramaniam S.V. (2011), *Healthcare and Equity in India,* Online pub. Lancet, 2011, Feb 5; 377(9764): 505–515.

2) Baru R., et. al. (2010), Inequities in Access to Health Services in India: Caste, Class and Region, *Economic and Political Weekly*, Vol XIV No 38.

3) Baru R.V. and Bisht R. (2010), *Health Service Inequities as Challenge to Health Security,* Oxfam India Working Papers Series IV.

4) Chiang E., *Healthcare Economics Global Networks*, Stone G.W (ed.), Core Economics (ebook) , www.worthpublishers.com

5) Hurley J. (2000), *An Overview of the Normative Economics of the Health Sector,* Cuyler A. J. And Newhouse J.P., (eds.), Handbook of Health Economics (Vol I), Elsevier, Netherlands.

6) International Monetary Fund (2004), *Health and Development,* Washington DC.

7) Santerre R.E. and Neun S.P. (2010), *Health Economics: Theories, Insights and Industry Studies (5th ed.),* South-Western Cengage Learning, USA.

8) World Health Organisation (2009), *WHO Guide To Identifying the Economic Consequences of Disease and Injury,* Geneva.

9) Zweifel P., Breyer F and Kifmann M. (1997), *Health Economics (2nd Ed.),* Springer, OUP.

7.12 ANSWERS/HINTS TO CYP EXERCISES

Check Your Progress 1

1) See 7.1 and answer.

2) See 7.2 and answer.

3) See 7.2.1 and answer.

4) See 7.2.1 and answer.

5) See 7.2.2 and answer.

Check Your Progress 2

1) See 7.3 and answer.

2) See 7.3 and answer.

3) See 7.4 and answer.

4) See 7.4.1 and answer.

5) See 7.4.2 and answer.

Check Your Progress 3

1) See 7.5 and answer.

2) See 7.5.1 & 7.5.2 and answer.

3) See 7.6 and answer.

4) See 7.6 and answer.

Check Your Progress 4

1) See 7.7 and answer.

2) See 7.7.1 and answer.

3) See 7.7.1 and answer.

4) See 7.7.2 and answer.

5) See 7.8.1 & 7.8.2 and answer.

6) See 7.8.3 and answer.

UNIT 8 SUPPLY OF HEALTH SERVICES

Structure

8.0 OBJECTIVES

After going through this unit, you will be able to:

- characterise the basic features of health services;
- distinguish between healthcare and medical care;
- state the different types of health services;
- indicate the difference between the private and public supply of health services;
- derive the average revenue productivity of physicians under the Pauly-Redisch model;
- discuss the equilibrium conditions of a price discriminating physician;
- differentiate between health and healthcare production in the short run;
- explain how the 'efficiency frontier' for the production of health services varies for differing production technologies; and
- show that achievement of cost minimisation subject to output constraint is identical to achieving output maximisation subject to cost constraint.

8.1 INTRODUCTION

Health inequality, health burden and health system widely vary across the countries in the world. Low and middle income countries account for 24 percent of world's GDP but 90 percent of world's disease burden. It is estimated that India is losing more than 6 percent of her GDP annually due to premature deaths and preventable illnesses (World

Bank, 2010). With the total expenditure on health at about 4.2 percent of GDP, and the share of public expenditure at about 1.1 percent of GDP, India remains among the five countries with the lowest public health spending in the world. Out of pocket expenditure is therefore very high (about 70 percent of total health expenditure) and hospitalization for major illnesses is a major cause of indebtedness, especially for those living below the poverty line. Thus, providing adequate healthcare to mitigate such risks is one of the main policy objectives in developing countries. In the developed countries households are protected from high medical expenditure by requisite health insurance coverage or through a tax funded health system. But the situation is completely different in case of developing countries. Even though the coexistence of public and private healthcare institutions is accepted in all countries including developing countries, given that the objectives and functions of private and public healthcare institutions differ (private healthcare institutions having profit motive), the role of government in formulating policies for regulation is critically important. This is particularly because of the problem of moral hazard (i.e. recommending costly tests and treatment even where it is not required) and adverse selection (i.e. many poor getting left out of the public medical care and the non-poor being included) due to which government intervention is crucially needed. The joint venture of public-private partnership (PPP) model of healthcare institutions is suggested to correct such market anomalies to an extent. Against this background, the present unit deals with the issue of supply of health services from the point of view of establishing optimal conditions for its efficient delivery.

8.2 HEALTH SERVICES

In any healthcare system, service provision is the output from the combined inputs that goes into its production. The combined inputs are provided by different categories of skilled persons such as doctors, nurses, technical and non-technical health workers, procurement and distribution of pharmaceutical supplies, etc. While increased inputs should lead to improved service delivery, it does not always hold due to lack of efficient management (or coordination failure) of healthcare services. Ensuring availability of health services that meets a minimum quality standard along with the ease of securing access to them are the two key elements of an effective healthcare service. According to WHO the other key characteristics that should govern any efficient health service delivery are: comprehensiveness, accessibility, coverage, continuity, quality, person-centeredness, coordination and accountability & efficiency. In particular, the characteristic of comprehensiveness is taken to imply the coverage of a range of health services appropriate to the needs of the target population including preventive, curative and rehabilitative services under the broad umbrella of 'health promotion activities'.

Healthcare and medical care are two distinct terminologies. Although used interchangeably, from the perspective of health economics, the two terms are not identical. Healthcare is a much broader concept compared to medical care. Medical care is needed when a person becomes sick or ill. But healthcare should ideally start at an early age focused on life style habits like healthy diet, regular physical exercise, avoidance of excessive alcohol and smoking, etc. Medical care is provided by the doctors to the patients whereas healthcare can be provided by every individual to himself/herself. The concept of integrated system of healthcare services defined by WHO encompasses a 3-tier system of preventive healthcare. In this, primary prevention is defined as prevention of health problems before they occur (e.g. by health education, antenatal care, disease prevention, immunization, etc.). Secondary prevention refers to early detection of diseases and intervention which is also known as curative healthcare. Tertiary prevention refers to correction and prevention of deterioration, rehabilitation and terminal care.

In most of the countries in the world, both private and public healthcare services co-exist. However, the role of private healthcare is fundamentally different from public healthcare in terms of the objectives, functions and management. In view of this, in a democratic as well as planned economy, healthcare is mainly provided by publicly financed institutions rather than private insurance payments. The main difference between the public and private healthcare system is that, besides most of the services of public hospitals being largely free of cost, due to 'non-rivalry (in consumption) and non-excludability properties' of public healthcare, the number of patients per doctor is far higher in public hospitals than that in a private hospital. This impacts on the quality of healthcare services provided by public hospitals.

Supply of healthcare services are complex because they are often multi-product and interdependent. Such interdependency influences the complexity of efficient supply of healthcare services due to factors like: (i) labour time of multitude of trained professionals such as physicians, surgeons, specialists, nurses, medical technicians, pharmacists, etc.; (ii) diagnosis of diseases through procedures and testing; (iii) nursing care and hospital services; (iv) emergency services like ambulances and special intensive care units; and (v) supply of pharmaceutical products. The potential of any one of these critical services placing the patient's life in danger is what makes the health services mutually interdependent and complex in its character. In other words, healthcare service is not homogeneous like any other service bought and sold in a market. In addition, consequent to the entry of big private hospitals (run on corporate style) in the healthcare market, there is a significant shift in the market structure influencing the delivery of healthcare service in the urban markets. Although they cater to the more elite, allowed by a conscious decision by the government to reduce the gap between the demand and supply for quality healthcare services, it makes the market structure dichotomized bringing into fore the glaring gap between the standard of services provided by the publicly and privately managed health institutions. Since the ultimate outcome of healthcare very much depends on the joint-efficiency of interdependent services, the specialization of input markets goes to suggest the supply-side complexities in the ultimate delivery of efficient health service. To understand the issue of supply of health services, we therefore proceed to consider two specific situations in this section. In the first, we shall take a look at the 'organization of healthcare market' which brings into focus the government's role in establishing the policy linkages for the operation of different agents covering both the public and the private sector players and institutions. We shall then turn explicitly towards the physician-centric role of healthcare services in the health sector. This is because since both the public and the private sector operate jointly, so as to bring about competition in services provided on the one hand and also bridge the huge gap in the healthcare services demanded by the population and the inability of the public sector to exclusively provide the same on the other, we need to focus on what drives the efficiency considerations of the scores of physicians operating in the market. In this, from the point of health economics, we shall particularly take a closer look at how the equilibrium price charged by the private physicians is arrived at depending on the nature of market conditions in which they function.

8.2.1 Organisation of the Healthcare Market

The linkage among the various interdependent agents involved in healthcare services is organically depicted by Narciso (2004) [Figure 8.1]. The different agents of healthcare market are: national health authority (NHA), national public health service, healthcare providers such as physicians, hospitals, health insurance companies, consumers of health services (viz. patients), pharmaceutical companies, etc. The NHA of a country plays a significant regulatory role by formulating the national health policy. In particular, what

type of health intervention is to be given priority is determined by NHA. Besides directly controlling the public healthcare system, by deciding on the expansion of the public healthcare system NHA also determines whether a patient is to be given public healthcare service or not. In doing this, it takes into account the economic conditions of the population and the type of diseases considered important to be focused for public health delivery (e.g. polio eradication, leprosy control, etc.). In other words, keeping in mind the distribution of population, geographical spread and the incidence of different disease frequencies, NHA is supposed to set up different public healthcare institutions in an economy. NHA also directly controls the pharmaceutical industries who produce drugs as well as medical equipments by playing an important role in the determination of prices of medicines and healthcare instruments.

Pharmaceutical companies supply their products to pharmacies as well as physicians. Thus, the agency relationship between the physicians and patients provides the incentive to promote the products of pharmaceutical companies with the objective of influencing their choice of medicines. If a country allows the law of substitution in the healthcare market, the pharmacists would be allowed to sell the medicines different from the one prescribed by the physician. This means that under the substitution law, when a doctor prescribes a branded drug, the pharmacy is entitled to sell a generic version of that medicine to the patient directly. The private health insurers also play an important role in the healthcare market by contracting the payment schemes between the physicians and the patients.

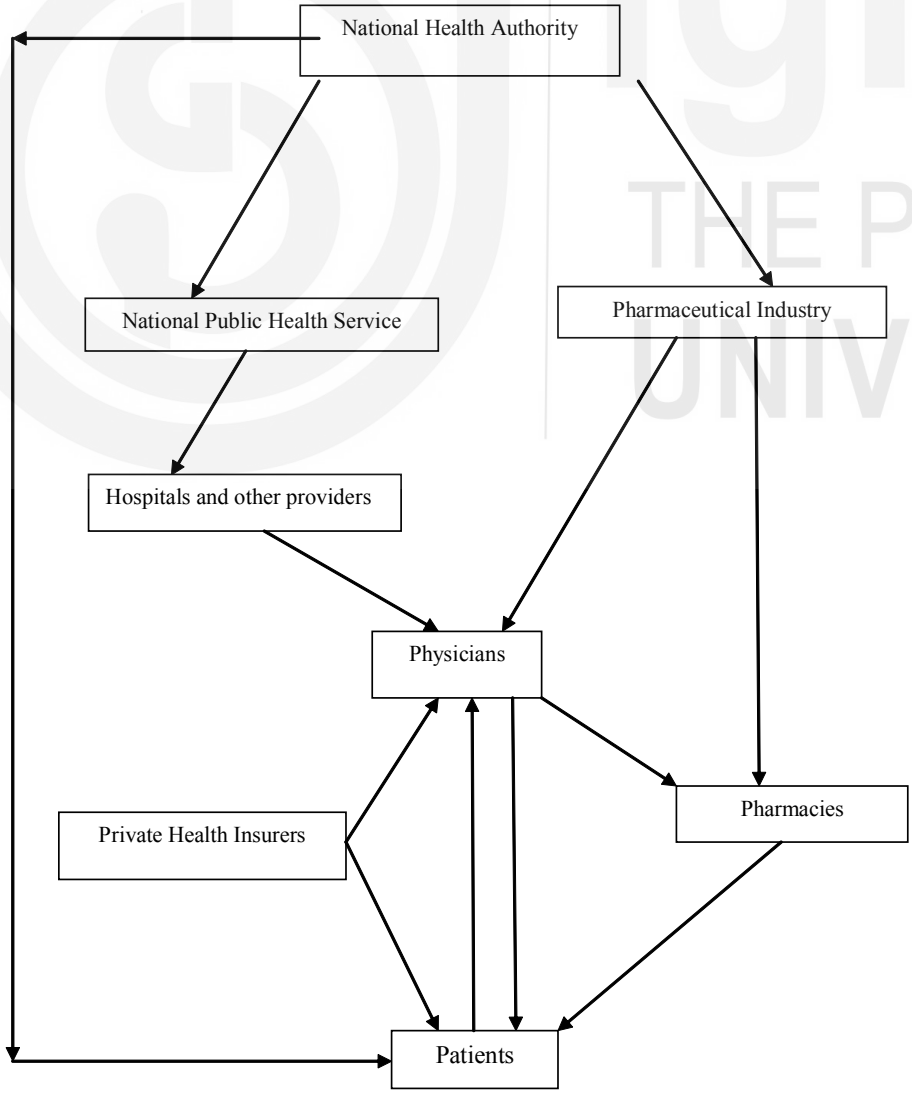


Figure 8.1: Organization of Healthcare Market

Economists view the provision of physician services as a production process involving a multitude of inputs aside from the physician's labour. These include the labour inputs of others like nurses, physician assistants, technical staff, etc. along with various non-labour inputs such as office space, medical instruments and diagnostic equipment. In light of this, it is important to distinguish the physician labour as a strategic input in the production of medical services. Standard economic theory dealing with markets for normal goods and services cannot be applied in case of healthcare market because of the following reasons:

1. the assumption of perfect information is not valid in case of physician's healthcare market;

2. a physician might charge higher fees from patients belonging to higher income group than those with lower income group for getting same quality of services. Here, we need to assume that the price differences are positively related to quality differences as otherwise it amount to 'price discrimination' which, under perfect competition conditions, the physicians do not have the market power to impose;

3. medical fees charged by the physicians vary to a large extent across different geographical areas. It is generally argued that if competitive forces operate, medical fees charged would tend to be lower;

4. if the prices of physician's services are set administratively, either by the government or by a private insurer, the quantity (i.e. number of patients attended by the physician) of physician's services tend to increase (i.e. moral hazard due to greed taking over).

The market for physician's services is therefore monopolistically competitive rather than perfectly competitive. In view of this, unlike in competitive markets, every physician has his own downward sloping demand curve. This is because search among doctors is costly for patients and therefore if a patient is satisfied with the services of a physician, the patient generally does not move to other physician unknown to him//her. Not only such information tend to spread among the consumers (i.e. patients), a patient may not leave his/her physician even if the physician charges higher price for healthcare service. In view of this, price differentials among physicians in a monopolistic market can persist for a long time. A second reason why the physicians face a downward sloping demand curve is that people may be willing to pay more for the convenience of visiting a physician who is geographically closer to their location. Thus, to attract patients coming from distant places, physicians may sometimes choose to offer their services at a discount. Given these factors, at what level the physician's fees/price would stabilize, under conditions of differing markets, is an important aspect of Health Economics on which the literature is abundant. In view of this, this is the next issue to which we shall now turn. We, however, need to note that the efficiency considerations of the government health services are different from those of the private physicians i.e. the former's is driven by the objective of meeting the social optimum as opposed to the private optimum in case of the latter.

Check Your Progress 1 [answer the questions in about 100 words within the space given]

1. State the basic characteristics of a well functioning healthcare delivery system. Among these, what does the term comprehensiveness connote?

   .....

   .....

.....

.....

2. Distinguish between healthcare and medical care. What does the term 'integrated system of healthcare services' connote?

.....

.....

.....

.....

3. Why is the supply of healthcare services viewed both as complex and interdependent? What are the factors that contribute to these twin characteristics of healthcare services?

.....

.....

.....

.....

4. What role does the agency relationship play between the physicians and the pharmaceutical companies in the healthcare market?

.....

.....

.....

.....

5. State any three reasons as to why the principles of standard economic theory cannot apply to healthcare market.

.....

.....

.....

.....

6. Why is it that the price differentials in the healthcare market among private physicians tend to continue to prevail for a long time?

.....

.....

.....

.....

8.2.2 Determination of Equilibrium Price for Physicians

Monopolistic competition explains the existence of price differential among physicians in a particular geographical area but cannot explain why physicians' fees are higher in areas with more physicians relative to population. One plausible reason is due to differences of input prices of the physicians. For instance, rent is higher in main city or town compared to sub-urban or rural areas. Thus, if a physician opens a clinic or hospital at a central place of a city, it raises the average total cost of the physician. It is also argued that patient's search cost is higher in large cities than in smaller cities which

in turn makes the demand curve of individual physicians less elastic. This is so because if the physician were to rise his prices, under conditions of less information, patients would face higher search costs and hence other factors remaining constant, price is higher when the demand curve facing the individual physician is less elastic. Figure 8.2 shows the equilibrium price of a physician under conditions of monopolistic competition where the marginal revenue equals the marginal cost (MR = MC). Figure 8.3 shows the physician's short run equilibrium in which he earns the 'economic profit'. In the long run, however, the physician earns just the 'normal profit' since he tends to set the price at a point where the price is equal to the 'average total cost' (ATC) [Figure 8.4]. In fact, if the price set by the physician is at a point where it is equal to 'minimum average total cost' (MATC), it would indicate the point of his 'productive efficiency'. However, if the average total cost is above the market price, then the physician will incur a loss equal to the 'average total cost' minus 'the market price' multiplied by the quantity produced. He can still minimise his losses by producing that quantity where the marginal revenue equals marginal cost (i.e. P = MC, the point where the physician can attain 'allocative efficiency'). Points 2 and 3 of Figure 8.4 are sub-optimal because at point 2, he will incur losses and will leave the market whereas at point 3, the very fundamental principle of a firm's equilibrium is violated (since MR ≠ MC). Technically, therefore, at points 2 and 3, the productive and the allocative efficiency are not achieved by the physician.

In some cases, a physician will be in an advantageous position to continue earning economic profits, even in the long run. For instance, a physician can have an excellent location, relative to other locations in the area, which will always give him an advantage over other physicians in that market. Or, a physician may have distinct and costly medical equipment that gives him an edge preventing competition from others. In such cases, physicians have some degree of market power which would allow them to price their services above the competitors' price.

8.2.3 Price Discrimination in Conditions of Dual Market

The asymmetry of information between the patient and the physician ensures that the latter determines the patient's medical needs. In this context, 'supplier induced demand' is defined as the capacity of a supplier (physician) to exploit his informational advantage to induce a higher level of demand for the good or service than what the consumer would have demanded if both the physician and the patient had the same amount of information. In healthcare market, supplier-induced demand translates to the capacity of the physician to induce patients to demand healthcare that they do not necessarily need. In the alternative scenario i.e. when this advantage is not available to the physician, he would be acting as the 'price taker' (as opposed to his role as the 'price setter' in the former case). In such a situation of dual markets i.e. one in which he is the 'price setter' and the other in which he is the 'price taker', there is an optimal amount of patients that he can treat from both the markets. The case of first market is similar to patients who have a medical insurance although in this case the physician has to accept the insurer's payment as payment in full. The physician therefore has the option to decide up to what number he can accept patients with insurance coverage and reject others. In other words, in situations like this, the physician has two demand curves (Figure 8.5). The demand curve faced by the physician in the first market is D_p with the corresponding marginal revenue curve MR_p. In the insured market (i.e. where the patients have health insurance coverage), physician faces a horizontal demand curve which is equal to the marginal revenue curve. The elasticities of two markets being different, the physician can discriminate prices for his/her healthcare services. If we assume that there is an administrative cost in the total cost component of the physician,

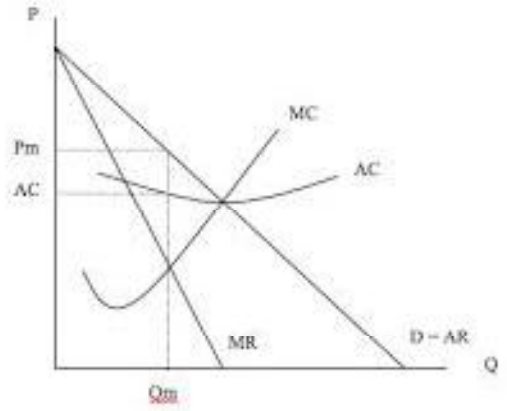


Figure 8.2: Equilibrium of a Monopolist

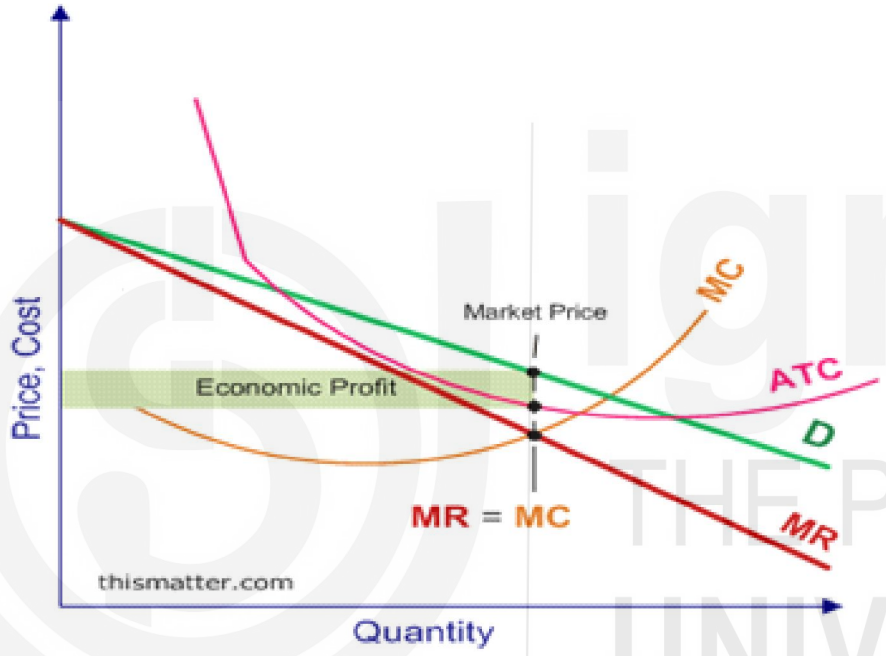


Figure 8.3: Monopolistic Competition: Short Run Economic Profit

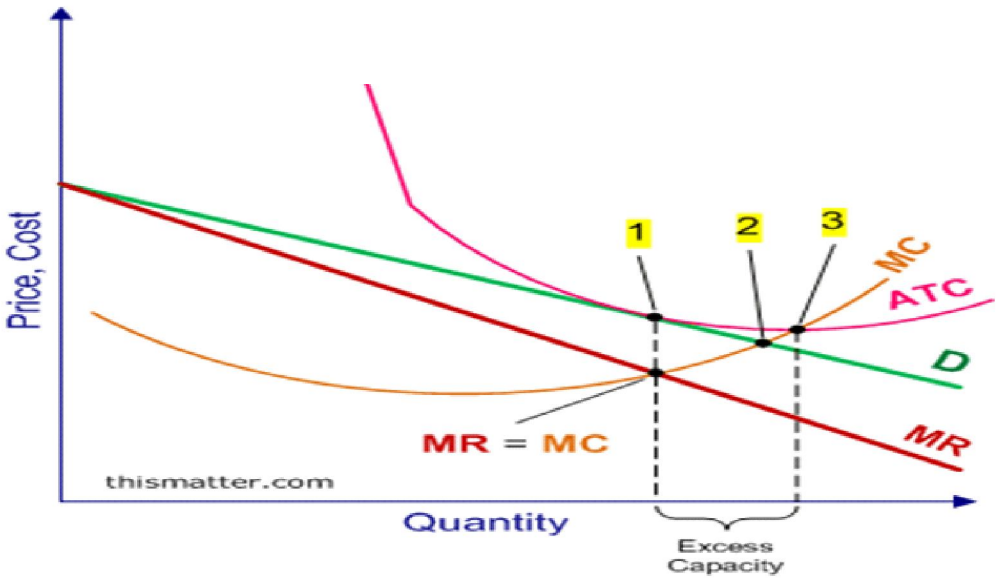


Figure 8.4: Monopolistic Competition: Long-Run Equilibrium

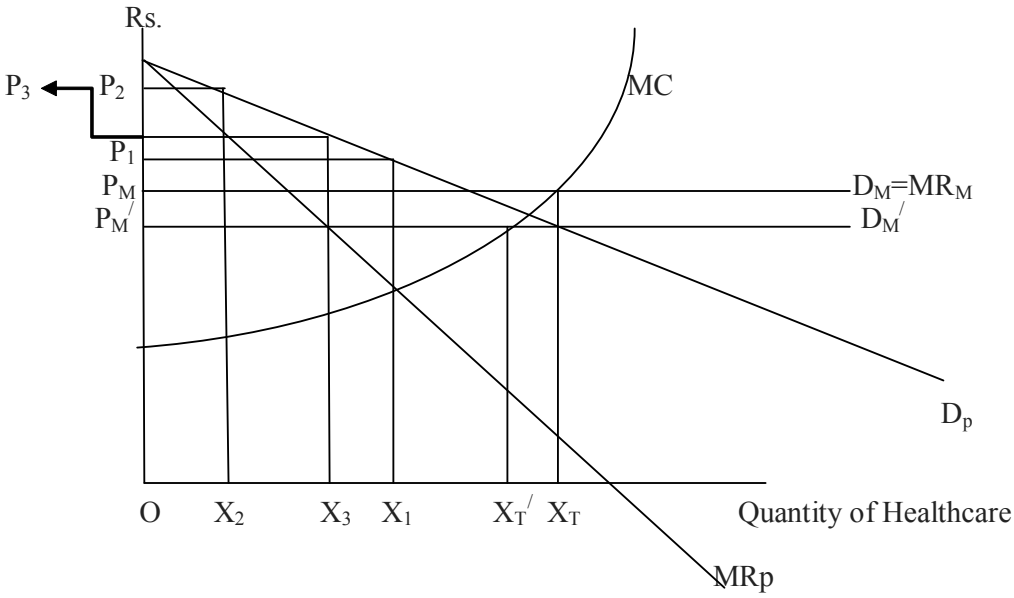


Figure 8.5: Equilibrium for a Price Discriminating Physician

like the cost of collecting payment from the insurer ($\hat{g}$) [i.e. in cases of delays in payment where the physician has to remind the insurer], with the corresponding component for the non-insured cases indicated by $\breve{g}$, we need to subtract the per unit administrative cost from the respective prices in the two markets. Further, other than this type of administrative cost, there is also a common cost function for the physician which applies to both the markets. These costs include the imputed value of physician time and the cost of non-physician labour and capital inputs. Assuming that the total cost of healthcare service in both the markets is identical, we have MC as same for both the markets. Hence, the equilibrium price (P_1) for the healthcare service (X_1) is determined by equating MR_p and MC in the first market. In the 2nd market, where the physician is a price taker, the demand curve is infinitely elastic. Thus, $P_M - \hat{g} = MR_M$ (where the sub-script M stands for the insurance payment). Since the physician accepts the patients in the decreasing order of profitability, the physician starts with patients in the first market, accepting the patients up to (X_2, P_2), as determined by equating MR_M and MR_P. From this point onwards, the physician accepts patients from the 2nd market until he reaches the quantity X_T. Thus, the number of units he accepts from the price-taking market is (X_T-X_2) with the rest of the units coming from the first market.

If we assume that the price falls to $P_M/$ in the 2nd market, the total output from this market also falls to $X_T/$ and the output from the first market rises to X_3 i.e. as determined by $D_M/ = MR_p$. In this case, the output in the first market rises consequent to a fall in the market price from P_2 to P_3. That is, output in the 2nd market falls from (X_T - X_2) to ($X_T/ - X_3$). The main argument here is that patients covered by the health insurance market lose some access to the physician after the fall in the price. With the total physician output falling for the insured market, the change in output due to this decrease is allocated to persons in the 2nd market. Patients in the first market enjoy the benefit of price cut.

8.2.4 Optimality Conditions in the Presence of Quality Variable

While most of the hospitals are 'not for profit' run by the public sector, there has also been a significant presence of private sector corporate style hospitals being run 'for

profit'. The latter mainly signifies 'quality' service which the former due to heavy user pressure is unable to provide in most of the cases. Here, we first consider a hospital that maximises its profit and then the case of 'not for profit' hospital. Assuming once again that the hospital faces downward sloping demand curve for its services, the relationship between the quantity of healthcare services produced (x) and the price (p) charged for it, can be expressed as:

$$p = f(x), \frac{dp}{dx} = f'(x) < 0, \quad \text{..... (8.1)}$$

Let $\pi = p(x)x - c(x)$ where π stands for profit and c(x) is the total cost of the hospital services. First order condition for profit maximization requires that:

$$\frac{d\pi}{dx} = 0 \Rightarrow \frac{dp}{dx}x + p(x) - \frac{dc}{dx} = 0 \Rightarrow MR = MC \quad \text{..... (8.2)}$$

The optimal quantity of healthcare services (x*) at an equilibrium price (p*) is determined by the intersection between MR and MC. In reality, however, physicians face a tough competition because in an area the number of physicians are numerous whereas in case of hospital (especially big hospitals), the number of hospitals are extremely few. Thus, a good private hospital can enjoy the monopoly power but physicians practising in small clinics in the suburbs, in general, cannot.

The profit function of the hospital that provides quality healthcare service can be written as:

$$\pi = p(x, y)x - C(x, y) \quad \text{.....(8.3a)}$$

$$\pi = p(x, y)y - C(x, y) \quad \text{.....(8.3b)}$$

where y stands for the level of quality and C is the total cost which is positively related to both x and y. Since the hospital cannot increase both x and y simultaneously, in order to maximise profit, the firm will first choose x = x* for a given level of y. Similarly, in respect of quality of healthcare, the firm will then choose y = y* that ensures profit maximisation for a given level of quantity. First order condition for profit maximisation with respect to x and y, yields $\frac{\partial \pi}{\partial x} = \frac{\partial \pi}{\partial y} = 0$. That is:

$$\frac{d\pi}{dx} = x\frac{dp}{dx} + p(x, y) = \frac{dC}{dx} \text{.....(8.4)}$$

$$\frac{d\pi}{dy} = y.\frac{dp}{dy} + p(x, y) = \frac{dC}{dy} \text{.....(8.5)}$$

Optimal conditions for the supply of healthcare service by the hospital indicate that in equilibrium, the hospital will choose $x=x^*$ and $y=y^*$ level of quantity and quality respectively that ensures equality between MR and MC. An important implication of this case is that hospitals will provide different levels of quality which will be higher when the price increase associated with an increase in quality is higher. In other words, $\frac{dp}{dy}$ is higher and the added cost of producing higher quality is lower, that is $\frac{dC}{dy}$ although positive is relatively low.

The objective of a public hospital is, however, not to earn profit but rather to maximise its utility (U) which is comprised of quantity (x) and quality (y) of healthcare services. Assuming a hospital maximises $U(x,y)$ subject to a budget constraint, we can explain the equilibrium conditions by assuming that the quality of healthcare is fixed at y_o (Figure 8.6). We have two average cost curves AC_L and AC_H, one for lower level of quality of healthcare and the other for higher level quality. Similarly, there exists two demand curves (viz. D_L and D_H) - one for lower quality and other for higher quality. Figure 8.7 shows that when the quality of healthcare is allowed to vary from L to H (i.e. lower to higher), the equilibrium point shifts from e to e* (i.e. from lower quality healthcare to higher quality healthcare). Consequently, the equilibrium amount of healthcare increases from x* to x**. Since here both the demand and average cost curves are defined for a fixed quality level, when quality changes, the demand and cost curves also shift provided the patients (consumers of healthcare) are willing to pay a higher price for a higher quality level. The demand curve will shift rightward as quality of healthcare increases and the average cost curve will move upwards for every quality improvement.

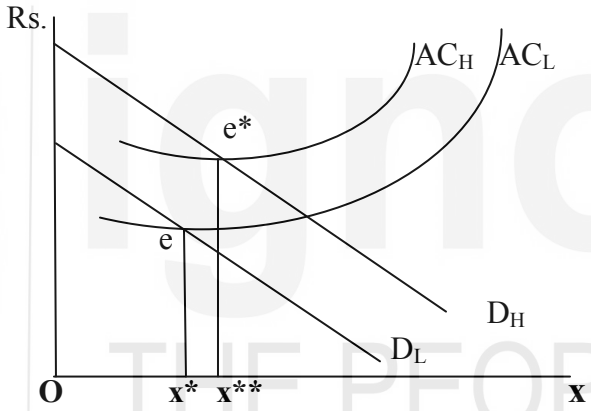


Figure 8.6: Equilibrium for a Public Hospital with Quality

8.2.5 Optimality Under Physicians' Cooperative

Pauly and Redisch (1973) developed a model of hospital services in which physicians work as a cooperative (i.e. hospitals like poly-clinics that are commonly seen in cities). The model estimates the average revenue of physicians under certain assumptions made for the cooperative venture. These are:

(1) The patient would receive two bills, one for the hospital charges and the other for the physician's services;

(2) The patient is concerned with the total bill to be paid to the hospital;

(3) Demand for hospital stay depends on price charged by the hospital per day with the length of stay remaining fixed;

(4) Total price of the hospital stay is P and X is the duration of stay measured in days. Therefore, $P = P(X)$, where P includes payment to the hospital (P_H) and the physician;

(5) The hospital earns zero profit i.e.

$$\pi = 0, TR = TC, P_H X = wL + cK \quad \text{.....(8.6)}$$

where w is the wage rate for hospital employees, L is the number of persons employed other than physicians, c is the price of capital and K is the amount of capital used by the hospital.

(6) Total revenue is distributed between the hospital and physicians. Therefore:

$$PX = P_H X + Y_M M \quad \text{..... (8.7)}$$

where Y_M is the mean income per physician of the hospital and M is the number of physicians. Therefore, Y_M can be written as:

$$Y_M = \frac{(P - P_H)}{M} X \text{ , ... (8.8)}$$

Using (8.6), (8.8) can be rewritten as:

$$Y_M = \frac{PX - wL - cK}{M} \text{ ..... (8.9)}$$

Thus, quantity X is produced by three inputs: physicians engaged in hospital, non-physician labour and capital. Therefore, X can be written in a functional form as:

$$X = X(M, L, K) \text{ ..... (8.10)}$$

with $P = P(X)$, such that $\frac{dP}{dX} < 0$

The prime objective of this model is to maximise income per physician from the production of X. The collective decision of the hospital's physicians is to maximise Y_M. Therefore, hospital will choose M, L and K in such a way that doctor's income (Y_M) is maximised.

First order condition to maximise Y_M with respect to L yields: $\frac{\partial Y_M}{\partial L} = 0$, Thus, from equation 8.9, we have:

$$\frac{\partial Y_M}{\partial L} = \frac{P\frac{\partial X}{\partial L} + X\frac{\partial P}{\partial X}\frac{\partial X}{\partial L} - w}{M} = 0 \text{ .....(8.11)}$$

Equation (8.11) can be alternatively written as:

$$\frac{\partial X}{\partial L}\left(P + X\frac{\partial P}{\partial X}\right) = w \quad \text{.....(8.12)}$$

Since $MP_L = \frac{\partial X}{\partial L}$ and $MR = P + X\frac{\partial P}{\partial X}$ the left hand side of equation (8.12) represents marginal revenue productivity of labour (MRP_L). Therefore, in equilibrium, the hospital will hire labour (L*) up to that level where MRP_L is equal to wage rate (w). In the same way, the optimal amount of capital (K*) is determined by equating MRP_K = c and the optimal size of physician is determined by equating $\frac{\partial Y_M}{\partial M} = 0$

Therefore, from equation (8.9), we have:

$$\frac{\partial Y_M}{\partial M} = \frac{P\frac{\partial X}{\partial M} + X\frac{\partial P}{\partial X}\frac{\partial X}{\partial M}}{M} - \frac{P\,X - cK - wL}{M^2} = 0 \quad \text{.....(8.13)}$$

Multiplying throughout by M and rearranging the terms, we get:

$$\frac{\partial X}{\partial M}\left(P + X\frac{\partial P}{\partial X}\right) = \frac{PX - cK - wL}{M} = Y_M \quad \text{.....(8.14)}$$

Left hand side of equation (8.14) represents the marginal revenue productivity of physician and the right hand side is the income per doctor i.e. the average revenue productivity of physicians at the hospital.

Profit maximisation is not always the sole objective of all the private hospitals although there are hospitals whose objective is to maximise profit. The above analysis, therefore, suggests that a 'non-profit hospital' has an objective function defined as a linear combination of its profit and non-profit expressed in terms of either the quantity or quality or mix of both. If the hospital is run by a Trustee Board (of a religious or minority group), then the objective function of the hospital is to provide medical services to the poor uninsured individuals. In such hospitals, the specific religious group/ community are given priority in respect of waiting time, free medical bed or concession to the patients of the group.

Check Your Progress 2 [answer the question in about 100 words within the space given]

1. Under conditions of 'monopolistic competition', in the short run, when does a private physician earn his 'economic profit'?

   .....

   .....

   .....

   .....

2. In the context prevailing in question 1 above, distinguish the terms 'productive efficiency' and 'allocative efficiency'. Also, identify the levels at which the private physician would set his market price.

   .....

   .....

   .....

   .....

3. What is meant by 'supplier induced demand'? How does it translates to the patient's disadvantage in the healthcare market?

   .....

   .....

   .....

   .....

4. Specify the relationship between 'quantity of healthcare produced' and 'price p' and the functional form for profit maximisation. Also, state how the equilibrium price is determined?

.....

.....

.....

.....

5. Under what circumstance does the Pauly-Redisch model apply and what does it basically seeks to show?

.....

.....

.....

.....

8.3 PRODUCTION OF HEALTH

Individuals derive utility from good health. If health or health status of an individual is bad, his/her utility will decline. Health is also a durable good as it depreciates over time. By proper medical treatment, health status of an individual can be improved. A health production function can therefore be written as:

$$H = f(HC, LS, G, EF, O, E) \quad \text{.....(8.15)}$$

where, H = Health, HC = Healthcare, LS = Life Style, G = Genetics, EF = Environment Factors, O = Occupation, E = Education.

8.3.1 Relationship Between Health and Healthcare

A production function of health aims at maximising output that can be produced out of a given combination of inputs. It is a technical relationship relating inputs to output (health) in the most productive way. Note that healthcare (HC) enters as an input into the production of health as in equation (8.15). All medical firms, including hospitals, physician clinics, nursing homes, and pharmaceutical companies, earn revenues from producing and selling some type of medical or healthcare output. Production and retailing activities occur regardless of the form of ownership. A healthcare production function can therefore be written as:

$$HC = f(P, N, NM, K) \quad \text{.....(8.16)}$$

where, HC = Output of healthcare service, P = number of physicians, N = number of nursing staff, NM = number of non-medical staff and K = amount of physical capital. Demand for healthcare is a derived demand; its purpose being to create health. We now differentiate between health and healthcare production using simple graph in the short-run (Figure 8.7).

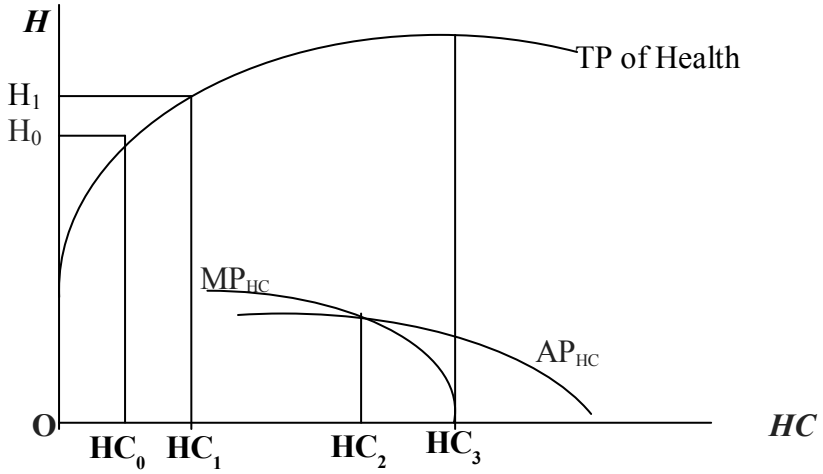


Figure 8.7: Relationship Between Health and Healthcare

In the short-run, we assume that other factors of production are constant and only healthcare varies. We can therefore see that as healthcare (HC) increases, health output (H) also increases but at a diminishing rate. This can be mathematically written as:

$$\frac{\partial H}{\partial HC} > 0, \frac{\partial^2 H}{\partial HC^2} < 0$$

That is, the marginal product of healthcare (MP_{HC} i.e. $\frac{\partial H}{HC}$) is positive but it is diminishing ($\frac{\partial^2 H}{\partial HC^2} < 0$) over time. MP_{HC} is zero when the 'total productivity of health' is maximum i.e. at HC_3 healthcare, MP_{HC}, is zero. Note that to the left of HC_2, $MP_{HC} > AP_{HC}$, whereas to the right of HC_2, $AP_{HC} > MP_{HC}$ but at HC_2, $AP_{HC} = MP_{HC}$. If AP_{HC} ($=\frac{H}{HC}$) is rising, this means that:

$$\frac{dAP_{HC}}{dHC} > 0 \Rightarrow \frac{d}{dHC}(\frac{H}{HC}) > 0 \Rightarrow \frac{HC.\frac{dH}{dHC} - H}{HC^2} > 0 \Rightarrow \frac{\frac{dH}{dHC}}{HC} > \frac{H}{HC^2} or, MP_{HC} > AP_{HC}$$

Similarly, we can verify that when AP_{HC} is maximum, then $MP_{HC}=AP_{HC}$. In the same way, one can prove that when AP_{HC} is falling, $MP_{HC}< AP_{HC}$.

In Figure 8.8, if we assume that pollution level is exogenously given, a higher pollution level reduces the health status. If we assume that the pollution level $P_1<P_0$, then the total output of health shifts upward. An improvement of the pollution level (from P_0 to P_1), i.e. a reduction in the pollution level increases the health from B to D. If the costs of the two measures are assumed to be the same, then policy maker will chose to improve the air quality instead of healthcare. Therefore, in order to carry out the desired economic policy (or for proper allocation of resources), one has to know the health production function as well as the marginal contribution of pollution to total productivity (TP) of health (H) (i.e. cost of reducing the pollution level and cost of improving the level of health).

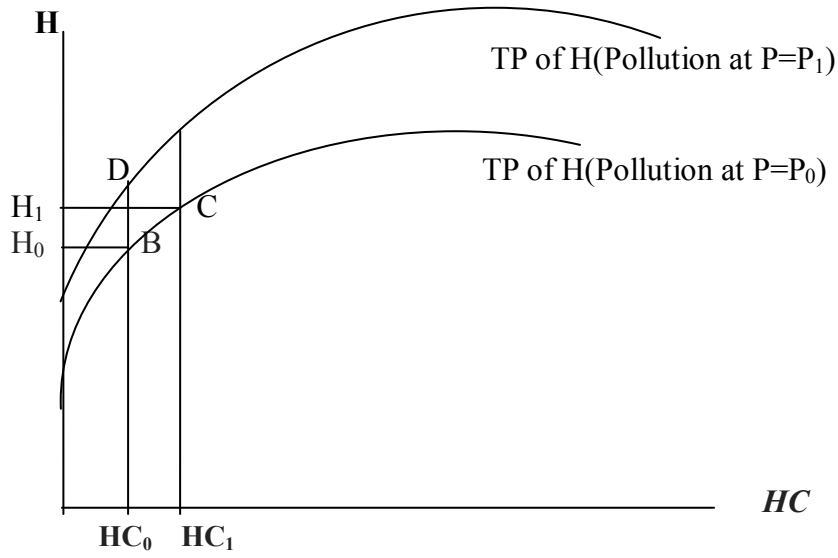


Figure 8.8: Influence of Pollution on Healthcare

8.3.2 Input Substitution and Healthcare Services

Assume that a given level of output of health (H) can be produced by using two inputs X_1 and X_2 with all other factors remaining constant. Graphically, the efficiency frontier is the lower envelope of the set of production possibilities (Figure 8.9) associated with a particular volume of output of health. Therefore, for each output level, there will be the corresponding efficiency frontier. The collection of all these input combinations yields the concept of production function. The efficiency frontier defines a boundary for the set of production possibilities associated with each level of output. The graphical representation of the efficiency frontier is termed as isoquant which represents all possible efficient ways to achieve a certain production level. Figure 8.9 shows three distinct types of isoquants based on three distinct types of production technologies viz. isoquants AA/, BB/ and LL/ associated with three different types of production functions. If we move along OO/ (from O to O/), output of health (H) increases. Therefore, health output produced along the isoquant AA/ is lower than the output produced along the isoquant BB/. Similarly, output produced along isoquant LO/L/ is higher than that of BB/. Such isoquants are also called as equal product curves possessing certain

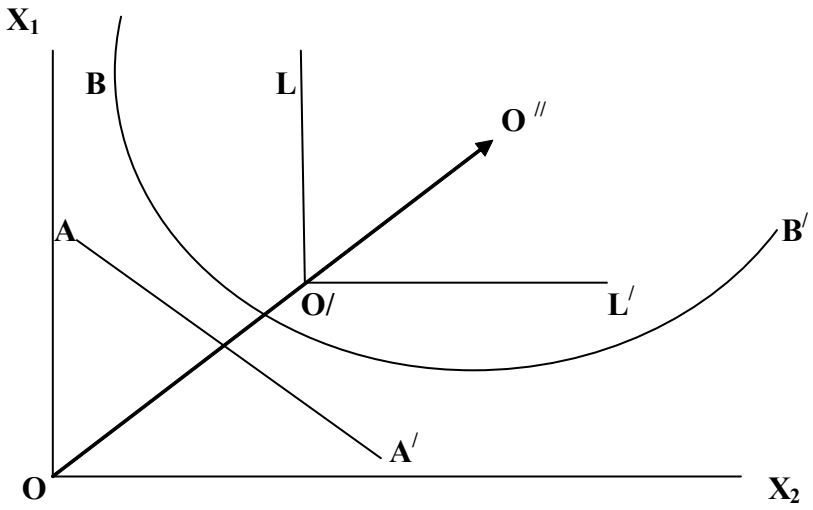


Figure 8.9: Efficiency Frontier for Differing Production Levels

properties. These are: (i) a single isoquant represents a fixed level of output; (ii) two successive isoquants can never intersect each other; (iii) isoquants are convex to the origin; (iv) higher isoquant represents higher level of output, lower isoquant represents lower level of output; (v) if the marginal product of one of the two factors is zero, isoquant is parallel to the axis on which the factor with zero marginal product is represented; (vi) if one of the two factors has negative marginal product the isoquant slopes upwards from left to right; (vii) if both the factors have negative marginal products, the isoquant is concave to the origin; (viii) if X_1 and X_2 are prefect complements to each other, the isoquant is L-shaped which means that the two inputs are always used in fixed proportion and an extra unit of X_1 or X_2 does not add to total output; and (ix) isoquant may be linear having the slope of (-1) when X_1 and X_2 are perfect substitutes.

The marginal rate of technical substitution (MRTS) is defined as: keeping the total output of health desired to be produced as constant, how much of input 1 (X_1) has to be decreased if input 2 (X_2) is increased by one extra unit. In other words, it shows the relation between inputs, and the trade-offs amongst them, without changing the level of total output of health. Technically, MRTS is defined as:

$$MRTS_{X_1,X_2} = (-\frac{dX_1}{dX_2}) = \frac{MP_{X_1}}{MP_{X_2}} \quad \text{.....} \quad (8.17)$$

Here $\frac{dX_1}{dX_2}$ is the slope of the isoquant and MP_{X1} and MP_{X2} are the marginal productivity of X_1 and X_2 respectively. MRTS is thus the negative slope of the isoquant. Equation (8.17) is derived assuming the health production function as $H = H(X_1, X_2)$ for fixed level of health output. Assuming $H = \overline{H}$, one can write: $\overline{H} = H(X_1, X_2)$. Taking the total differential on $\overline{H}$, we have: $d\overline{H} = \frac{\partial H}{\partial X_1}.dX_1 + \frac{\partial H}{\partial X_2}.dX_2$. Along a fixed isoquant, since there is no change in output of health (H), $d\overline{H} = 0$. Thus, we have:

$$d\overline{H} = \frac{\partial H}{\partial X_1}.dX_1 + \frac{\partial H}{\partial X_2}.dX_2 = 0 \text{ or } \frac{dX_1}{dX_2} = -\frac{\frac{\partial H}{\partial X_2}}{\frac{\partial H}{\partial X_1}} = -\frac{MP_{X_2}}{MP_{X_1}}$$

If we assume that along the straight line isoquant AA/, H_0 level of output can be produced, then this can be achieved by using various input combinations of X_1 and X_2. If X_1 and X_2 are further assumed to be perfect substitutes of each other, then the isoquant will be a straight line having a negative slope of -1. This means, if we add one unit of X_1, we must sacrifice the same unit of X_2 in order to remain on the same isoquant H_0. Thus, the isoquant AA/ will touch neither X_1 nor the X_2 axis; if AA/ touches either axis, then H_0 can be produced either by X_1 or X_2 alone.

Along the isoquant BB/, health output is higher than in the isoquant AA/ (BB/ being above AA/). If the isoquant BB/ corresponds to H_1 level of output of health (which can be produced by using various input combinations of X_1 and X_2), then along the isoquant BB/, substitution that takes place between inputs is imperfect. As a result, MRTS tends to zero while diminishing the quantity of X_1 and to infinity while diminishing the quantity of X_2.

Finally, if we assume the health output produced along the L-shaped isoquant $LO'L'$ is H_2 since $LO'L'$ lies above both the isoquants AA' and BB', $LO'L'$ obviously gives higher output. Here, there is no scope of input substitution with the two inputs (X_1 and X_2) being treated as complement to each other i.e. the two inputs are always used in a fixed proportion. In this case, the horizontal segment of $LO'L'$ has a MRTS = 0 and along the vertical segment of $LO'L'$, the MRTS = ∞. This is because X_1 changes but X_2 remains same along LO'. As a result, since, $dx_1>0$ but $dx_2=0$, dx_1/dx_2 = infinity.

8.3.3 Technical Substitution and Elasticity of Substitution

An important characteristic of MRTS (i.e. marginal rate of technical substitution) is that it diminishes as more and more of X_2 is substituted for X_1 [as shown in BB' (but not in AA' and $LO'L'$) in Figure 8.9]. In other words, as the quantity of X_2 used is increased and the quantity of X_1 employed is reduced, the amount of X_1 that is required to be replaced by an additional unit of X_2 so as to keep the output constant will diminish. Isoquants AA' and $LO'L'$ are the extreme cases but the isoquant $LO'L'$ is more realistic in health/healthcare production.

The principle of diminishing MRTS is merely an extension of the law of diminishing returns for the relation between the marginal physical products of the two factors X_1 and X_2. Along an isoquant on BB', as the quantity of X_2 is increased and the quantity of X_1 is reduced, the marginal physical product of X_2 diminishes and the marginal physical product of X_1 increases. Therefore, less and less of X_1 is required to be substituted for every additional unit of X_2 so as to maintain the same level of output. Also, the rate at which the MRTS diminishes is a measure of the extent to which the two factors can be substituted for each other. Thus, the smaller the rate at which the MRTS diminishes, the greater is the degree of substitutability between the two factors. If the marginal rate of substitution between any two factors does not diminish and remains constant, then the two factors are perfect substitutes of each other. This happens along the isoquants AA' (in Figure 8.9). Similarly, along the isoquant $LO'L'$ (which is of fixed coefficient type), the optimum health supply $H = Min(\frac{X_1}{\alpha}, \frac{X_2}{\beta})$, where, α and β are two positive constants. In the Figure 8.9, if the angle $O'' O' L'$ is 45^0 (as shown) then $\alpha = \beta$. In our present case of fixed coefficient type production, the path of production will take place along the locus of line $OO'O''$. Thus, in case of $LO'L'$ isoquant, no substitution is possible between the two factors X_1 and X_2 i.e. two factors of production are perfect complement to each other. In other words, if the two factors are used in fixed proportion, we can avoid wastage of inputs of either X_1 or X_2.

Thus, the MRTS of factor X_2 for factor X_1 declines as more of factor X_1 is substituted for factor X_2 along a standard isoquant (BB' in Figure 8.9). In other words, the MRTS is different for different factor-proportions (i.e. input ratios) used in the production of a health (or healthcare) service. The responsiveness of the proportions in which the factors (or inputs) are used, as there is a movement along an isoquant, may be compared with the change in the substitution possibilities in production as measured by the change in MRTS.

Elasticity of Input Substitution (σ): Elasticity of input substitution is defined as the proportionate change in factors employed to the proportionate change in MRTS in such a way that the output remains the same. It measures the strength of substitution effect. It also measures the relative extent to which one factor will be replaced by the other, whenever there is a change in their relative prices. This means that if X_1 becomes cheaper, the producer will substitute X_1 for X_2.

The rational producer will relatively use more X_2 than X_1 if price of X_2 falls. The rate at which the two inputs (X_1 and X_2) will be substituted for each other will depend upon the MRTS between the inputs and their relative prices. Mathematically, we can define the elasticity of input substitution (σ) as follows:

$$\sigma = \frac{\frac{\Delta(\frac{X_2}{X_1})}{(\frac{X_2}{X_1})}}{\frac{\Delta MRTS_{X_1X_2}}{MRTS_{X_1X_2}}} = \frac{\Delta(\frac{X_2}{X_1})}{\Delta MRTS_{X_1X_2}} \frac{MRTS_{X_1X_2}}{(\frac{X_2}{X_1})} \quad \text{.....} \quad (8.18)$$

where $MRTS_{X_1X_2} = \frac{MP_{X_1}}{MP_{X_2}}$

A producer in equilibrium may replace MRTS by the input price ratio. Thus, the elasticity of substitution can be written as:

$$\sigma = \frac{\Delta(\frac{X_2}{X_1}) \frac{P_{X_1}}{P_{X_2}}}{\Delta(\frac{P_{X_1}}{P_{X_2}}) \frac{X_2}{X_1}} \quad \text{.....} \quad (8.19)$$

where, $MRTS_{X_1X_2} = \frac{P_{X_1}}{P_{X_2}}$

Elasticity of substitution (σ) is always positive varying from zero to infinity depending on the different types of the production possibilities vis-a-vis isoquants. For a fixed coefficient type of production function (e.g. Leontief production function), the elasticity of substitution (σ) is zero. In case of Cobb-Douglas type production function, σ=1 and in case of CES type production function, σ = constant.

8.3.4 Factors of Production of Health and Efficient Use of Resources

A rational producer of health or healthcare tries to minimise the cost for a given level of output or maximise the output for a given level of cost. In case of two factors of production (X_1, X_2), if we know the prices of two inputs, we can specify the total cost function as:

$$C = P_1X_1 + P_2X_2 \quad \text{.....} \quad (8.20)$$

where C = total cost, P_1 = unit price of X_1 and P_2 = unit price of X_2 etc. The production function can be written as:

$$H = H(X_1, X_2) \quad \text{.....} \quad (8.21)$$

Cost minimization subject to output constraint requires that:

Minimize C subject to $H = \bar{H}$. The problem can be solved by using a new function, V as:

$$V = C + \lambda(\bar{H} - H(X_1, X_2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.22)$$

where λ is the Lagrange multiplier. First order condition for minimization of cost (C) subject to output constraint ($\bar{H}$) requires that, $\frac{\partial V}{\partial X_1} = P_1 - \lambda, \frac{\partial H}{\partial X_1} = 0,$

$\frac{\partial V}{\partial X_2} = P_2 - \lambda, \frac{\partial H}{\partial X_2} = 0,$ Therefore, equilibrium condition yields:

$$\frac{P_1}{P_2} = \frac{MP_{X_1}}{MP_{X_2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.23)$$

Similarly, the other objective of the producer is to maximize output (H) for a given level of cost. The objective function is thus:

Maximize H subject to $C = P_1X_1 + P_2X_2$. Therefore, the Lagrangian function can be written as:

$$L = H(X_1, X_2) + \lambda(\bar{C} - P_1X_1 - P_2X_2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.24)$$

The first order condition for maximizing H subject to cost constraint (C=$\bar{C}$) requires that:

$$\frac{\partial L}{\partial X_1} = \frac{\partial H}{\partial X_1} - \lambda P_1 = 0, \quad \frac{\partial L}{\partial X_2} = \frac{\partial H}{\partial X_2} - \lambda P_2 = 0.$$

Thus, in equilibrium, the producer will use inputs in such a way that the price ratio of two inputs should be equal to the ratio of marginal productivity of two inputs. That is:

$$\frac{P_1}{P_2} = \frac{MP_{X_1}}{MP_{X_2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.25)$$

The entrepreneur will therefore be in equilibrium in regard to the use and purchases of the two factors when he is using the quantities of the two factors in such a way that the marginal physical products of the two factors are proportional to the factor prices. Therefore, cost minimisation subject to output constraint is identical to output maximization subject to cost constraint.

How does a firm (hospital) expand his production given the factor prices? In a two inputs model, producer will always follow the condition (8.25) which means that the slope of the isocost lines should be equal to the slope of the isoquants. The slope of the isocost line is derived by taking total differential of equation (8.20), keeping C, P_{X1} and P_{X2} constants. That is:

$dC = P_1dX_1 + P_2dX_2$, Setting $dC = 0$, we have $\frac{dX_1}{dX_2} = -\frac{P_2}{P_1}$

The slope of the isoquant is: $\frac{dX_1}{dX_2} = -\frac{MP_{X_2}}{MP_{X_1}}$

Optimality in connection with the input use is reached if the slope of the isocost line is equal to the slope of the isoquant:

$$\text{i.e. } -\frac{P_2}{P_1} = -\frac{MP_{X_2}}{MP_{X_1}} \text{ or } \frac{P_1}{P_2} = \frac{MP_{X_1}}{MP_{X_2}}$$

The above is same as equation (8.23). This condition is known as the expansion path of the producer. It can also be defined as the locus of the points of tangency between the isoquants and the isocost lines. The expansion path is also known as scale-line of production because it shows how the producer will change the quantities of the two inputs when the scale of production and input prices P_1 and P_2 are increased. The line joining the minimum cost (or its dual i.e. output maximization subject to cost constraint) combinations such as E_0, E_1, E_2, E_3 (Figure 8.10) is called the expansion path because it shows how the input combination with which the entrepreneur produces will alter as the producer expands its level of output (H).

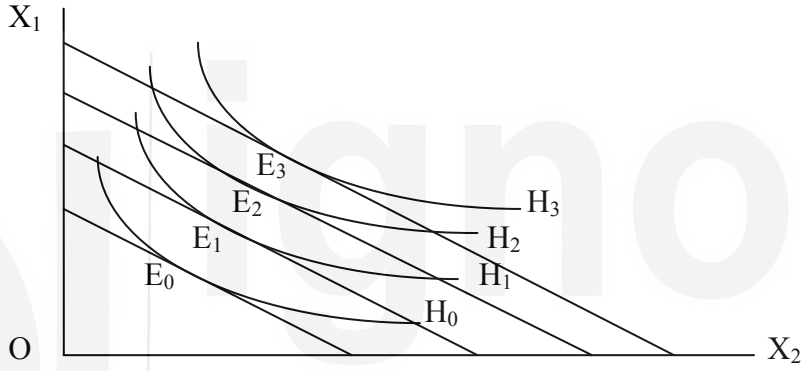


Figure 8.10: Expansion Path of Producer

8.3.5 Estimation of Cost Function from Production Function of Health

We now derive the average and marginal cost functions from the production function of health (H). Cost function can be estimated from production function in the short-run as well as in the long-run. In the short-run the law of variable proportions operate where only one input is considered as variable with the other assumed to be constant. But in the long-run both the inputs would vary and laws of returns to scale would operate.

In the short-run, we assume that the production function (8.22) can be written as $H = H(X_1)_{\bar{X}_2}$ in which input X_2 is fixed at a certain level and only input X_1 varies. We assume that production of health (H) means reduction of infant mortality or increase in infant survival of a particular area. Let X_1 be the input like medical care, X_2 be the input like sanitation or hygiene level, P_1 be the unit price of medical care and P_2 be the unit price of sanitation or hygiene level (both P_1 and P_2 measured in terms of \$/Rs per unit).

Law of variable proportion implies that $\frac{\partial H}{\partial X_1} > 0, \frac{\partial^2 H}{\partial X_1^2} < 0$. The total product of health curve looks like inverted U-shaped which means that as X_1 increases total health increases at a diminishing rate and once it reaches a maximum level, it declines. The corresponding marginal productivity also initially increases, reaches a maximum level

and then sharply declines. In other words, the marginal productivity of X_1 is zero when the total health product is maximum.

The short run average cost (**SRATC**) can be written as:

$$SRATC = \frac{C}{H} = \frac{SRTVC + SRTFC}{H} = \frac{SRTVC}{H} + \frac{SRTFC}{H} = SRAVC + SRAFC$$

.....(8.26)

where SRAVC is the short-run average variable cost and SRAFC is short-run average fixed cost. Graphically, the short-run ATC, AVC, AFC and MC would be as shown in the Figure 8.11.

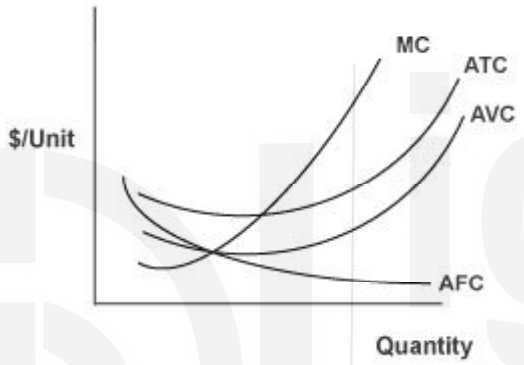


Figure 8.11: Short Run Marginal/Average/Fixed/Total Cost Curves

If we assume that for maintaining a standard hygiene or sanitation level, society has to incur a fixed cost F, then AFC in the short run is: $\frac{F}{H}$. As H increases, AFC declines continuously and it becomes asymptotic to the quantity of health axis (X axis). Now, given P_1, due to law of variable proportions, in the short-run the per unit AVC of health reduces because of higher productivity. It finally reaches to a minimum level whereafter, the per unit AVC rises. MC will pass through the minimum point of AC. In other words, assuming ATC is falling, this means that:

$$\frac{d}{dH}\left[\frac{C}{H}\right] < 0 \Rightarrow \frac{H.\frac{dC}{dH} - C\frac{dH}{dH}}{H^2} < 0 \Rightarrow \frac{MC}{H} - \frac{1}{H}\frac{C}{H} < 0 \Rightarrow MC < ATC \cdot$$

In the same way, we can show that when *ATC* is minimum, *MC=ATC* and when *ATC* rises, $MC > ATC$.

Long-Run Cost Functions from Production Function**:** Here we assume that the production of health or healthcare service is undertaken by the public authority. The technological possibilities of production of health are not identical across the countries in the world with developing countries having to incur a greater amount of cost compared to the developed countries. The inputs (and their prices) of production of good health also differs to a large extent between the developed and developing countries. If the elasticity of substitution between X_1 and X_2 is assumed to be one, then the production

function is of Cobb-Douglas type which exhibits constant returns to scale in the long-run. That is:

$H = ZX_1^{\alpha} X_2^{\beta}$ with $\alpha+\beta=1$. .....(8.27),

Here, Z is a set of exogenously given technological shifter affecting the unit cost of production of health (H) but Z is neutral with respect to the productivities of X_1 and X_2 and their uses. Dispersed population causes higher private as well as public transportation cost in terms of money and opportunity cost of time. Likewise, if a population is comprised of higher share of either older or very younger (aged less than 5 years) age cohort, government has to incur a higher cost of providing healthcare services or to produce good health (like reduction of morbidity or infant mortality) compared to a population with lower share of old age cohort or children.

Now, assuming production of health (H) means reduction of infant mortality or increase in infant survival of a particular area, X_1 is the input like medical care, X_2 is the input like sanitation or hygiene level, P_1 is the unit price of medical care and P_2 is the unit price of sanitation or hygiene level. The total cost (C) function is given by $C = P_1X_1 + P_2X_2$. Minimizing C subject to output constraint ($H = \overline{H}$) yields the Lagrangian:

$$L = P_1X_1 + P_2X_2 + \lambda(\overline{H} - ZX_1^{\alpha} X_2^{1-\alpha}) \quad \text{.....(8.28)}$$

First order condition of minimization of L requires that:

$$\frac{\partial L}{\partial X_1} = P_1 - \lambda.Z.\alpha.X_1^{\alpha-1}.X_2^{1-\alpha} = 0 \quad \text{.....} \quad (8.29)$$

$$\frac{\partial L}{\partial X_2} = P_2 - \lambda.Z.(1-\alpha).X_1^{\alpha}.X_2^{-\alpha} = 0 \quad \text{.....} \quad (8.30)$$

$$\frac{\partial L}{\partial \lambda} = \overline{H} - Z.X_1^{\alpha}.X_2^{1-\alpha} = 0 \quad \text{.....} \quad (8.31)$$

From 8.29 and 8.30, we can write:

$\frac{P_1}{P_2} = \frac{\alpha}{1-\alpha}.\frac{X_2}{X_1}$ or $X_2 = \frac{1-\alpha}{\alpha}.\frac{P_1}{P_2}X_1$. Now, substituting the value of X_2 in equation (8.27), we get:

$$X_1 = \frac{H}{Z}\left[\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right]^{\alpha-1} \quad \text{.....} \quad (8.32)$$

Likewise, $X_2 = \frac{H}{Z}\left[\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right]^{\alpha}$ ..... (8.33)

Substituting the values of X_1 and X_2 in the total cost ($C=P_1X_1+P_2X_2$) function, we get:

$$C = P_1.\frac{H}{Z}\left[\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right]^{\alpha-1} + P_2.\frac{H}{Z}\left[\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right]^{\alpha}$$

$$C = \frac{H}{Z}\left[P_1\left\{\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right\}^{\alpha-1} + P_2\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\right]. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.33A)$$

Hence, the average cost (AC) function can be written as:

$$AC = \frac{C}{H} = \frac{1}{Z}\left[P_1\left\{\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right\}^{\alpha-1} + P_2\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8.34)$$

The marginal cost (MC) function for linearly homogeneous health production function is derived by differentiating C (=Total Cost) with respect to output H:

$$\frac{dC}{dH} = \frac{d}{dH}\left[\frac{H}{Z}\left[P_1\left\{\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right\}^{\alpha-1} + P_2\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\right]\right].$$

Since Z, P_1,P_2, α are constants, we can write:

$$\frac{dC}{dH} = MC = \frac{1}{Z}\left[P_1\left\{\frac{P_1}{P_2}.\frac{1-\alpha}{\alpha}\right\}^{\alpha-1} + P_2\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\right].\frac{dH}{dH}$$

Or,

$$MC = \frac{1}{Z}\left[\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\left(P_1.\frac{P_2}{P_1}\frac{\alpha}{1-\alpha} + P_2\right)\right] = \frac{1}{Z}\left[\left\{\frac{P_1}{P_2}\frac{1-\alpha}{\alpha}\right\}^{\alpha}\left(\frac{\alpha P_2 + P_2 - \alpha P_2}{1-\alpha}\right)\right]$$

Or,

$$MC = \frac{1}{Z}\left[P_1^{\alpha}.P_2^{-\alpha}(1-\alpha)^{\alpha}.\alpha^{-\alpha}P_2(1-\alpha)^{-1}\right] = \frac{1}{Z}\left[\left(\frac{P_1}{\alpha}\right)^{\alpha}.\left(\frac{P_2}{1-\alpha}\right)^{1-\alpha}\right] \ldots\ldots.(8.35)$$

Therefore, in case of linearly homogenous production function, the average and marginal cost is identical which is linearly parallel to the output axis. This is because P_1, P_2, Z and α are constants. Equations 8.35 can be written as:

$$AC = MC = \frac{1}{Z}\left[\frac{P_1}{\alpha}\right]^{\alpha}\left[\frac{P_2}{1-\alpha}\right]^{1-\alpha} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.36)$$

Under perfectly competitive market in the long-run, the equilibrium of a producer is $P_X=AC_X=MC_X$. This corresponds to zero profit because $MR=MC=P_X=AC_X$. Therefore, in the long-run, the equilibrium condition of public healthcare is:

$$\text{LRAC=LRMC=}P_X=MC_X=AR_X=\text{Constant} = \frac{1}{Z}\left[\frac{P_1}{\alpha}\right]^{\alpha}\left[\frac{P_2}{1-\alpha}\right]^{1-\alpha}$$

This is the equation of straight line parallel to the horizontal axis (output axis), if we measure Price/AR/MR/AC/MC vertically. Under perfectly competitive market in the long-run, the equilibrium of a producer is $P_X = AC_X = MC_X$. This corresponds to zero profit because $MR = MC = P_X = AC_X$. If the health production function exhibits constant returns to scale, the AC and MC becomes identical which again equals to price of health output.

Check Your Progress 3 [answer the question in about 100 words within the space given]

1. Specify a production functions of health and healthcare. State the difference between the two in terms of the determinants chosen.

   .....

   .....

   .....

   .....

2. What does an 'efficiency frontier' basically define?

   .....

   .....

   .....

   .....

3. What are isoquants? State any five of their properties.

   .....

   .....

   .....

   .....

4. Define the term 'marginal rate of technical substitution' (MRTS). What does MRTS basically show?

   .....

   .....

   .....

   .....

5. Define the elasticity of input substitution. What does it measure?

   .....

   .....

   .....

   .....

6. What is meant by the 'condition for the expansion path of the producer'? How is it defined?

.....

.....

.....

.....

8.4 PUBLIC-PRIVATE PARTNERSHIP (PPP) IN HEALTH SERVICES

The Public-Private Partnership (PPP) in the provision of health services is defined as a means to bring together a set of actors for the common goal of improving the health of a population based on mutually agreed roles and principles (WHO, 1999). It can also be considered as a form of agreement that entails reciprocal obligation and mutual accountability, voluntary or contractual relationships, sharing of investments and reputational risks, and joint responsibility for design and execution of public ventures to meet social obligations along with a business motivation to sustain. According to European Commission (2003), PPPs are defined as a contractual agreement between the public and the private sectors, whereby the private operator commits to provide public services that have traditionally been supplied or financed by public institutions. Such a joint venture of public and private sectors is suggested to contribute unique strengths, complementing each other, in aiding the complex delivery of healthcare. Studies on governments that contract private organisations in collaborative partnerships have shown positive findings, cost-effective behaviour, better coverage for rural areas, and higher quality of care. In implementing the PPPs in healthcare services, the relatively simple 'design, build, finance and maintain/transfer' models are the most commonly practiced across the countries in the world. However, different governments at the national level are experimenting with more ambitious models such as public-private integrated partnerships (PPIPs) where the private partner is entrusted with additional responsibility for delivering all clinical services in one or more health facilities, including as acute care hospitals, as well as one or more primary care facilities. Such mixed healthcare systems involve private investments and public government finances. Currently, governments in low-income countries spend less than US $34 per citizen for healthcare each year although some of these countries have the largest populations in the world. The haphazard healthcare services available in impoverished nations are due to unregulated private companies, an underfunded public sector and a lack of governmental accountability. By combining private and public healthcare services and insurance, national healthcare can become regulated, transparent, and more equitable. Global health organisations can aid in this endeavour by promoting reform through incentives, and nurturing awareness among the civilian population (Nistar 2010). Therefore, one can expect various benefits from successful PPP model of healthcare. Raman et. al. (2009) point out that PPP in healthcare can: (i) reduce out-of-pocket expenditure of poor people; (ii) increase efficiency of healthcare system by reducing wastage of funds and utilize the resources properly for strengthening the quality of doctors, nurses and non-medical staff; (iii) augment the accountability of the working staff in healthcare institutions; and (iv) ensure quality services of healthcare by implementing proper regulation, rules and laws.

The development of a successful PPP model requires attention to a number of issues. As PPP is a developing concept, the first stage must be to create a supporting institutional

structure to develop, guide and manage PPPs on behalf of the public sector. This will entail the development of supporting national and local legislation and regulations to enable PPPs and development of institutional capabilities with creation of effective management and oversight structures (European Commission, 2003). India started with public initiatives since independence but approaches towards private ventures in its healthcare system began in the 1990s. Lack of adequate support from public healthcare system, expensive private sector facilities, unbridled commercial behaviour of the private sector, etc. are some of the compelling reasons for which the government has been showing its interest in healthcare provisions for a viable third alternative including PPPs since the phase of second generation economic reforms in the 2000s. Most of the plan documents have highlighted the important role that NGOs can play in all aspects of healthcare, especially for the underprivileged sections of the society in remote areas. Among the initiatives in this direction, a good example in this area is of Voluntary Health Association of India (VHAI), a Delhi-based national network of more than 4000 NGOs with its operations spread across the country. This is one of the world's largest associations of voluntary agencies working in the areas of health and development. Other examples of this nature are in the areas of immunization, HIV/AIDS work and newly formulated government programs on reproduction and child health. This phase of experimentation in the Indian healthcare system can be marked to the years of post-2000 owing to its initiatives to implement the PPP models in healthcare services. The main policy goal of PPP in healthcare is to bring out the synergies in key stakeholder interests of the public and private sector institutions and simultaneously work towards establishing a strong regulatory mechanism for maintaining quality and equity in health services. PPP models vary in design and implementation across states in India because health sector is under the purview of state governments.

8.5 LET US SUM UP

A distinction is made in literature between 'healthcare' and 'health' services. The former does not require the intervention of a doctor or a physician, can start at a young age and broadly relates to taking preventive measures from falling ill by maintaining good life style habits. In the literature on production of health services, healthcare is regarded as an exogenous variable and health as endogenous. Explaining the difference between medical care and healthcare, the unit explains how different markets of medical care co-exist in an economy. The supply of health services is viewed largely from the physician-centric angle although the significant role of other functionaries is also duly taken into account in the discussion made under different situations/sections in the unit. The unit discusses many contexts in which conditions necessary for having the optimal services for healthcare in the society can be established. The 'efficiency frontier' for the production of health services being different for different production technologies, the unit discusses how the optimal use of healthcare inputs can be used for producing a fixed amount of medical care. The relationship between average and marginal costs of healthcare is established both in the short-run and long-run. The derivation of average and marginal cost of healthcare production is then explained as an illustration. A specific illustration of deriving the cost function from the production function of a producer is made towards the end (in section 8.3.5). Although the producer of healthcare service may be either from the public or the private sector, it is only in the public sector that 'zero profit' is kept as the objective. With this, revenue becomes equal to cost and in a competitive market, a hospital in the long run would have 'price = average revenue = marginal revenue = average cost = marginal cost'. Since the standard microeconomic theory considers LRAC (long run average cost) curve as the supply curve, the treatment presented in section 8.3.5 emphasises the case of providing healthcare services

established by the public sector for the achievement of 'social optimum' to be the fundamental objective of the government. These analytical presentations are helpful in understanding the efficiency frontiers or the required optimal conditions for establishment of efficient healthcare services. They thus provide a benchmark against which we can compare the efficiency standards and focus on improving upon them. They are in that way helpful in formulating policy prescriptions which in turn is useful in allocation of scare resources. However, keeping in mind the limitations of public healthcare and the importance of PPPs in this context, the role of PPPs (and the initiatives taken in this direction by the Indian government) is outlined towards the end of the unit.

8.6 KEY WORDS

Marginal Rate of Technical Substitution (MRTS) : For a production with two factors of production, MRTS is defined as: keeping the total output of health desired to be produced as constant, how much of input 1 (X_1) has to be decreased if input 2 (X_2) increases by one extra unit. In other words, it shows the relation between inputs, and the trade-offs amongst them, without changing the level of total output of health.

Isoquant : Refers to the efficiency frontier representing all possible efficient ways to achieve a certain production level.

Isocost : The term 'isocost' means equal cost. It is related to producer's budget. The '*isocost' line* illustrates all possible combinations of two factors that can be used at given costs and for a given producer's budget.

Elasticity of Input Substitution : This is defined as the proportionate change in factors employed to the proportionate change in MRTS in such a way that the output remains the same. It measures the strength of substitution effect.

Multi-Product : Health care service is not homogeneous like any other service bought and sold in a market. The hospital services are a product of multiple inputs from skilled and unskilled staff. They make use of durable buildings and equipment which have to be maintained and periodically replaced. The products are not a single product but a related bundle of goods and services directed towards the same objective of improving the health. In many cases, the goods and services are joint products with the supply and costs of one being not independent of the others. They are also in the nature of personal services available on 24 hours emergency basis. The services include both hotel and medical care with the larger hospitals having several stages of production (e.g. from

boiler houses to pathology lab, to in-patient and out-patient services). Such services have a quality dimension, even in case of public hospitals, which is difficult to measure objectively.

8.7 SUGGESTED REFERENCES FOR FURTHER READING

1) Barros, Pedro and Xavier Martinez-Giralt (2012): Health Economics: An Industrial Organization Perspectives, Routledge.

2) Buse, K. and G. Walt (2000): *Global Public-Private Partnerships: Part II – What are the Health Issues for Global Governance?* (Bulletin of the World Health Organization, 78(5), p.704.

3) Pauly, M. and M. Redisch (1973): 'Not–for–Proft Hospital as a Physician's Cooperative', American Economic Review, 63(1), 87-99.

4) Pauly, M. and Mark A. Satterthwaite (1981): 'The Pricing of Primary Care Physicians' Services: A Test of the Role of Consumer Information', Bell Journal of Economics, 12(2), 488-506.

5) Raman, A. V. and J.W.Bjorkman (2009): Public Private Partnership in Health Care in India: Lessons for Developing Countries, Routledge, London.

6) Sloan, Frank. A. and Chee-Ruey Hsieh (2012): Health Economics, The MIT Press, 55 Hayward Street, Cambridge, MA, 02142.

8.8 ANSWERS/HINTS TO CYP EXERCISES

Check Your Progress 1

1. See Section 8.2 and answer.
2. See Section 8.2 and answer.
3. See Section 8.2 and answer.
4. See Section 8.2.1 and answer.
5. See Section 8.2.1 and answer.
6. See Section 8.2.1 and answer.

Check Your Progress 2

1. See Section 8.2.2 and answer.
2. See Section 8.2.2 and answer.
3. See Section 8.2.3 and answer.
4. See Section 8.2.4 and answer.
5. See Section 8.2.5 and answer. The Pauly-Redisch model explores the possibility of physicians themselves organising to run the hospital as a co-operative on a no-profit basis. The physicians play a dual role – one of hospital administration and the other as supply of healthcare service providers to the consumers. The framework aims to optimise on the use of inputs (i.e. the hospital's medical staff) in such a manner that it maximises the income per physician.

Check Your Progress 3

1. See Section 8.3 and 8.3.1 answer.
2. See Section 8.3.2 and answer.
3. See Section 8.3.2 and answer.
4. See Section 8.3.2 and answer.
5. See Section 8.3.3 and answer.
6. See Section 8.3.4 and answer.

UNIT 9 DETERMINANTS OF HEALTH SERVICES

Structure

9.0 OBJECTIVES

After going through this unit, you will be able to:

- discuss the impact of healthcare services on consumer welfare;
- examine the two-way relationship between income and health;
- explain how poverty is not the sole determinant of malnutrition;
- outline the concept of 'social gradient of health';
- illustrate how wages and expected income can affect the supply of healthcare services;
- evaluate how the effects of organisational change and technical efficiency impacts the supply of healthcare services;
- state the implications of up-gradation of medical technology on healthcare costs;
- describe the methods and impact of pharmaceutical pricing on healthcare services; and
- analyse the impact of government policy on the demand for and supply of healthcare services.

9.1 INTRODUCTION

Initially, like any other market, income was thought as the prime determinant of demand for healthcare services. The fact that wealthy and rich countries possess far better

health status, and also demand higher levels of healthcare services, was the main reason for this belief. However, it has since become evident that certain non-income factors are equally crucial in shaping the demand for healthcare services in a country. Two broad type of such determinants classified in literature are: ***Structural*** and ***Proximal*** determinants. The former are distal characteristics that generate social and economic stratification such as economic, political, social and education systems. Proximal determinants are the circumstances of daily life like the quality of the family environment and peer relationships, availability of food, housing, recreation, access to education, etc. Proximal determinants are thus generated by the social stratification resulting from structural determinants as also the cultural, religious and community factors. In other words, while the structural determinants capture the nature of macro development issues, the proximal determinants concern themselves with the micro issues across and within the household framework. Likewise, the supply of healthcare services is controlled by factor inputs (e.g. wages, price, etc.) and organizational structure of pharmaceutical industry which is a major component of the healthcare sector. In this, we shall particularly see how the pharmaceutical industry suffers from strong non-price pseudo-competitive factors resulting in very high drug prices across countries. Against this background, the present unit deals with the structural and proximal factors that control the demand for and supply of healthcare services. Both of them, taken together and addressed duly, results in better health status of a community.

9.2 DETERMINANTS OF DEMAND FOR HEALTHCARE SERVICES

Healthcare is generally defined to encompass those goods and services whose primary purpose is to improve, or prevent deterioration in, health. Availability of healthcare services is one of the many determinants of the health status of a community. From the perspective of economic theory, healthcare services are an essential input into the production of health. As a result, unlike most of the other consumer goods, which are consumed for their direct utility generating properties, healthcare services are used to produce health, which again has high utility. Had healthcare goods and services followed the same economic principles as other goods, aiming for the optimization of utility of the consumers by striving to establish perfectly competitive markets would have led to a Pareto optimal situation. However, the fact that consistent state intervention in the healthcare market is found necessary, both in developed and developing countries, provides evidence to the fact that leaving everything to the invisible hands of market cannot solve the problems faced by the healthcare market.

Let us consider the concept with the help of a simple consumer framework in which the consumer's utility is construed to depend on general goods and services (X) and healthcare services (HC). The health status (HS) is then determined by a joint result of both the exclusive as well as the combination of the two factors Z and HC. That is:

$$U = f[X, HC, HS(Z, HC)] \quad \text{.....} \quad (9.1)$$

The effect of healthcare on welfare is, therefore, determined by:

$$\partial U/\partial HC \quad \text{.....} \quad (9.2)$$

which is the direct effect of healthcare on the utility of the consumer in which

$$(\partial U/\partial HS)(\partial HS/\partial HC) \quad \text{.....} \quad (9.3)$$

is the contribution of health status to welfare combined with the contribution of healthcare to health status. The first effect $\partial U/\partial HC$ in expression (2) is the normal direct effect of

healthcare on the person's welfare, which is common to any other consumer good X. The indirect effect of HC comes via HS in expression (3) which involves two factors: one, the marginal contribution of improvements in HS to utility ($\partial U/\partial HS$) [a subjective concept not identifiable easily] and, two, the marginal productivity of HC in producing HS ($\partial HS/\partial HC$) [a *technical relationship* guided by scientific research]. Both these marginal effects should have positive values for the expression of (3) to be positive. The total effect, which is a summation of expressions (2) and (3), would partially offset each other to finally reach a value which is normally positive. However, if a healthcare service is found to be ineffective in improving health (i.e. $\partial HS/\partial HC<0$), it may not necessarily improve the utility of the consumer.

Since the net effect on welfare depends on the production relationship between HS and HC, a simple demand analysis as in case of any other consumer good is inadequate. This brings in the concept of '*need'* in the health sector as a necessary condition for a good healthcare service to exist in the sense of effectively influencing the desired objective of improving the health status of the society. The 'need' concept is important as unnecessary administration of costly medical procedures (due to inefficiencies in the system like the medical profession's vested interest) could not only reduce welfare but also economically drain a patient's resources who might otherwise be better off than undergoing the unnecessary procedure. In other words, the medical care is a good 'good' when a person is really in need of the treatment and a 'bad' good when the treatment is administered for ulterior motives. Hence, the *technical relationship* can serve as a basis of the concept of 'need' as opposed to 'demand' which also depends on preferences backed by price conditions. *Demand for healthcare is thus derived from the 'need' for health itself* (Grossman 1972). This concept of the 'derived demand' in assessing the need for healthcare services involves the interaction between both the supply side and the demand side concepts and cannot be judged in isolation.

The relationship between medical care and health ($\partial HS/\partial HC$) is, however, far from exact. This is in the sense that there is not only a considerable lack of medical knowledge concerning the efficacy of certain types of medical interventions but there is also extreme uncertainty about how far a particular care can effect the health status. As a result, healthcare providers disagree about the treatment of some types of medical problems where the demand for medical services becoming ***fuzzy***. For instance, there is debate on when surgery is necessary for elderly males with prostate cancer. Thus, it is extremely difficult to accurately delineate the relationship between the price and the quantity demanded for medical care. The relationship between price and quantity demanded can therefore be better represented by not a well-defined line but a gray band like the one depicted in Figure 9.1. Two implications are associated with the fuzzy demand

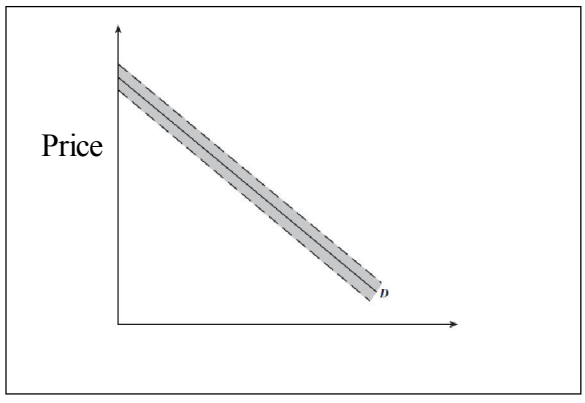


Figure 9.1: Fuzzy Demand Curve for Medical Care (Grey Band)

curve. One, for a given price, we may observe variations in the quantity or types of medical services rendered. Two, for a given quantity or type of medical service, we are likely to witness price differences. This is illustrated by the substantial variation in physician fees for similar procedures in the same geographical area.

Apart from questioning the basic premises of utility maximisation in healthcare, some economists even question the three very basic assumptions of an optimisation exercise. These are: *(i) whose utility are we maximising – Consumers' (i.e. patients') or his family members*?; (ii) violation of the assumption of sovereignty of the consumer in choosing the level and quality of healthcare service since most of the time it is the physicians who choose this on his behalf; and (iii) if we assume the consequences alone, and hence only the end-results (paying no importance to the process as we do in normal utility optimisation), we actually cannot capture the change in the health status or the stock due to healthcare.

With regard to the market structure of most of the healthcare goods, monopolistic competition describes the market pretty well. Healthcare providers sell differentiated products, while patients idiosyncratically choose one provider based on real or perceived differences in their abilities to cater to their match. Most healthcare providers compete with each other in the market, whose services though may be good are not perfect substitute to each other. Further, the service providers take their competitor's strategy as given and choose to respond with product differentiation. Because of this, coupled with the consumer's asymmetric information, a downward sloping demand curve emerges. In most of the cases, this demand curve becomes quite steep mainly due to:

- healthcare, to many, is a necessity and hence its purchase is not responsive to prices that they need to pay. This is particularly true in the presence of strong supplier-induced demand.
- many non-price factors play considerable impact. Apart from their socio-economic conditions, personal preference and trust assume critical importance.
- for high-end patients availing of employer paid or insured check-ups, consumers do not bear the full burden of price differentials and hence their demand for healthcare does not relate to price changes.

Given the above, the equilibrium price remains at a point greater than the marginal cost of production primarily due to the intrinsic differentiation of products and services with significant amount of price inelasticity.

Check Your Progress 1 [answer the questions in about 100 words in the space given]

1) Distinguish between the structural and proximal determinants of healthcare services.

.....

.....

.....

.....

2) Are healthcare services different from other goods and services? In light of your answer, what would you say about the need for government intervention in the healthcare market?

.....

.....

.....

.....

3) Using the standard consumer utility framework, state the two expressions whose marginal contributions should both be positive in order that the 'net indirect effect' of healthcare on the health status is positive.

.....

.....

.....

.....

4) Why is 'need' a necessary condition for a good healthcare service to improve the health status of a society?

.....

.....

.....

.....

5) Give arguments to bring out why the demand for healthcare services cannot be viewed in isolation of its supply side?

.....

.....

.....

.....

6) What is meant by 'fuzzy demand'? In the context of healthcare services, what are the two implications that follow from the presence of this type of demand?

.....

.....

.....

.....

7) State three reasons why in the case of healthcare services/market, the demand curve is quite steep?

.....

.....

.....

.....

8) Why does the equilibrium price in the healthcare market normally remain at a point above the marginal cost of production?

.....

.....

.....

.....

9.2.1 Income and Health

Income normally enters the demand for healthcare through the constraint in the optimisation exercise of utility function in equation (9.1). It is believed that increase in income, ceteris paribus, would automatically enhance the purchasing power and hence would increase the demand for healthcare. However, a more complex relationship emerges in this respect when one considers the effect of $\partial HS/\partial HC$. Once consumption of healthcare improves the health status, the productivity and hence income of the consumer is expected to increase, thus relaxing the assumption of linear budget constraint based on fixed income. Therefore, to start with, two things can be identified:

a) The impact of income on the demand for healthcare cannot bypass the issue of health status.

b) The relationship between the demand for healthcare and income would not be based on the assumption of fixed income but would be inter-twined with other cofactors.

Although the correlation between health status and income (across individuals and countries over time) is undeniable, economists and public health professionals have looked at the issue differently. Generally, economists focus on health as a source of human capital and hence economic output, while the public health exponents argue that the causal link between income and health is grounded in the redistribution of healthcare resources. Therefore, although the economists propose state intervention in healthcare and good health as a means of providing an avenue out of poverty, the public health group question such spending arguing for the fostering of economic growth to improve the lives and health of the global poor.

That the link between health and income is not unidirectional becomes clear once we incorporate the effect of income on health through $\partial HS/\partial HC$. The first obvious route from health to income is a direct link by productivity improvement. Higher levels of nutrition and the resulting better health, increases the productivity of workers in general. Further, lower incidence of morbidity also similarly impacts the worker's time spent on the job positively (vis-a-vis length of time devoted to job and absenteeism). Healthier children can attend school more regularly and learn more efficiently performing better in studies. Good health of a person within the family can free others from entering the labour market prematurely (via child labour) and continue skill development and education. This is well exemplified from the fact that the orphan children from HIV/AIDS epidemic in Africa and Latin America are forced to dropout from school very early. Also, due to the presence of strong externality in healthcare, improvement in one person's health can lead to improvement in health of others through better awareness (e.g. use of bed-nets in malaria-prone areas), sharing public health infrastructure (e.g. improved drinking water facilities and sanitation) and behavioural change (e.g. using condoms to prevent sexually transmitted diseases). Grossman's celebrated model (1972) by considering the long run life cycle approach, succinctly points out that

improvements in health add to the stock of human capital and hence can incentivise individuals to invest in health and education. Essentially, this brings into focus the term in equation (9.2) i.e. $\partial U/\partial HC$.

Cross-country correlation between population health (measured by life expectancy) and income per-capita, Pritchett and Summers (1996) found strong evidence of the fact that the relationship between income and health status is not just associative, but also causal and structural (Figure 9.2). Though technical advances and its diffusion in medical care have shifted this relationship, correlation between income and health has continued to remain robust. Other studies have also showed that better health is correlated with larger subsequent changes in income, suggesting that health not only

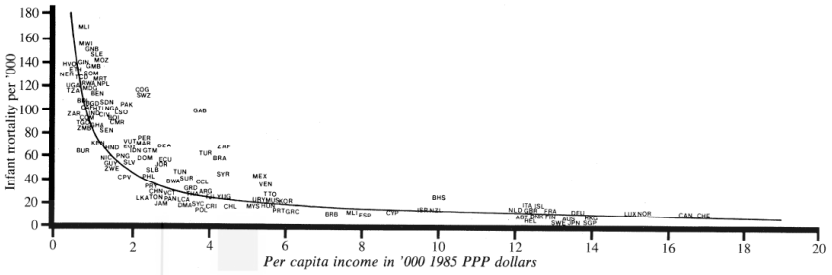


Figure 9.2: Association Between Income and Health (1990)

Source: Pritchett & Summers 1996

affects the level, but also the growth rate (Gallup and Sachs, 2001). At the macro level, cross-country data analysis on per-capita health expenditure has shown strong correlation with per-capita GDP with the elasticity found to be higher than one (i.e. suggesting healthcare to be a merit good). Thus, the structural effect of GDP on health status is undeniable.

However, there are also studies which have revealed low elasticity of income for utilisation of healthcare across households. This might be due to the strong presence of insurance market in developed countries due to which households pay only a minor fraction of the healthcare costs as direct out-of-pocket expenditure. In developing countries, healthcare utilisation like institutional delivery or immunisation is strongly cornered by the wealthy households, even though these services are offered free of cost by the state. This is mainly because of associated transaction costs (which are far from zero), suggesting a strong income effect. For instance, the report of National Family Health Survey (NFHS)-3: 2005-06 reveals that only 12.7 percent women coming from households with poorest wealth index have delivered their children in any institution, while the corresponding figure for the well-to-do group is as high as 83.7 percent. Similar figures for vaccination coverage posit that while 24.4 percent of children from the poorest group of households received all dosages of vaccination recommended by WHO, 71 percent children from the well-to-do group received all such dosages. On the other extreme, some studies have also come out with the finding that the hypothesis of 'income protecting health' is not supported by their analysis. These studies posit that recessions in developed countries have improved the adult health arguably because individuals engage in better preventive lifestyles during downturn by lowering drinking and smoking, doing more exercises and taking the meals regularly. This is, however, a condition that cannot work in a typical developing country set-up where lack of income is strongly cofounded with lack of health insurance, education, awareness, social rank and productivity.

9.2.2 Poverty and Malnutrition

The magnitude of malnutrition is generally assessed by comparing the food energy intake of persons with the proposed norms. Thus, at the onset, the poor who lack enough availability and affordability of good food, are believed to be prone to malnourishment. The nutritional status is measured by anthropometric indices viz. height for age (HAZ) representing *stunting* or chronic deficiency, weight for age (WAZ) representing *wasting* or acute malnourishment and weight for height representing *underweight* as a combination of HAZ & WAZ. Children whose anthropometric measure is less than 'minus 3 standard deviation (-3 SD) from the median' are labelled as severely malnourished (stunted/ wasted/underweight) and between 'minus 2 and 3 standard deviation from median' as moderately malnourished. Children, on this basis, are divided into three distinct categories viz. (i) properly nourished (Z-score above -2 SD); (ii) not properly nourished but not functionally impaired (Z-score between -2 SD and -3 SD); and (iii) malnourished (Z-score less than -3 SD). Any of this condition can occur not only due to lack of enough food, but also from over-consumption or imbalanced consumption of food. Essentially, therefore, it represents an imbalance between food intake and its assimilation within the body. Again, under-nutrition can be of two types: Protein Energy Malnutrition (PEM) and Micro Nutrients Deficiency (MND). While PEM is associated with lack of energy in food consumption, MND represents lack of vitamins and minerals which could cause diseases like anaemia, goitre, rickets, etc. It is, however, argued that 'food energy intake' is a poor measure of nutritional status as it does not depend *only* on nutrient intake but also on non-nutrient food attributes like drinking water facilities, sanitation habits and hygiene behaviour. For instance, in Kerala, low per capita food consumption coexists with low level of malnutrition while in Rajasthan, high per capita food consumption has not reduced the high prevalence of malnutrition. Thus, the relationship between poverty and malnutrition is not a linear one but rather involves complex understanding of the standard of living and awareness.

Empirical evidence, however, also show some cofounding relationship between poverty and malnourishment across Indian states, though factors other than poverty also play a significant role. Prevalence of malnutrition among non-poor household too is not only significant but is also revealing an increasing trend. Apart from the household level characteristics and mother's education status, type of food intake too is important. The NFHS-3 has, in fact, revealed that 8.2 percent of children in the top quintile class suffer from severe stunting and 25.3 percent from moderate stunting. The situation is far worse in case of urban rich group. Recent trends on consumption pattern among urban middle and rich income classes reveal that they are moving away from food grains and are consuming oily and fatty foods besides leading a sedentary life style. These habits have made the problem of controlling malnourishment way beyond the issue of poverty.

An yet another vicious cycle of malnutrition operates through the chain mother-child-mother. Children born to malnourished mothers or who are themselves malnourished during childhood can suffer cognitive losses that are associated with lower productivity in adulthood. WHO has, therefore, recognised the importance of initial 1000 days (i.e. from a woman's pregnancy to the second birthday of the child) to be extremely crucial for a child right after the mother conceives. The right nutrition during this 1,000 day window can have a profound impact on a child's ability to grow, learn, and rise out of poverty as well as shape a society's long-term health, stability and prosperity. Children malnourished before birth have higher risk of dying in infancy and are more likely to face lifelong cognitive and physical deficits and chronic health problems. Again, after birth, under-nutrition in early days can be life-threatening. It can weaken a child's

immune system and make the child more susceptible to death from common illnesses such as pneumonia, diarrhoea and malaria. Mothers with high anaemia, tender age and lack of empowerment, especially in the patriarchal societies of developing countries, tend to deliver infants more prone to malnourishment in the long run. It is, therefore, extremely crucial to break this chain in order to reduce the incidence of child malnutrition, with or without poverty.

9.2.3 Socio-economic Determinants of Health

Economic growth is undoubtedly important, particularly for poor countries, as it provides the wherewithal to invest in social sector development. But growth by itself, without appropriate social policies to ensure fairness in the way its benefits are distributed, brings little change to health equity. In fact, mal-distribution of healthcare services (i.e. not delivering care to those who most need it), has always been recognised as the main reason behind poor health and disease. Much of the burden of illness leading to appalling premature loss of life, both in developing as well as developed countries, are identified to both the distal and proximate conditions in which people are born, grow, live, work, and age. In all societies, both rich and poor, those with greater privilege tend to enjoy better health. Such privilege can emanate from better economic status, better education, better overall awareness, etc. While the difference in economic status would affect the incidence and prevalence of common communicable diseases, huge difference in the incidence of chronic illness, mental depression and healthcare utilisation also suggest a far deeper cause of inequality, reaching the '*causes of causes'*. Social scientists generally agree that a broader underlying dimension of social stratification (or *social ordering*) is the prime factor behind this. Thus, policies to reduce health inequity would have to involve a number of non-health areas of concern. For those involved in designing policies to address these gradients, therefore, a deeper understanding of the diversity and mechanisms of various factors is crucially needed. Conceptual framework of this theory of 'social gradient of health' draws on the *capability approach* of development (propounded by Amartya Sen), where development is taken to mean 'freedom to lead the life'. These issues have been discussed in detail in the WHO report on Social Determinants of Health (2008).

The social ladder in which an individual stands, therefore, becomes extremely crucial in determining his health status. Health differs across individuals not only due to the income earning capacity and employment but also due to one's position in social hierarchy. Lack of proper employment not only affects one's capacity to buy healthcare, but it also creates psychological stress due to feelings of insecurity. Studies have found that in a rich economic setting like Whitehall in Great Britain, ***occupational hierarchy*** determines the mortality figures of people in a major way. The social ladder theory also goes across generations. It assumes that employed mothers can spend on their child's healthcare services readily. It also assumes that participation in labour market offers the mother a sense of empowerment within the household and hence she can decide favourably for her child's healthcare. However, employment also reduces the time spent on childcare. In developing countries, where much of the women employment is associated with informal sector (without any provision for maternity and childcare), mothers often cannot take their children to healthcare centres even for routine immunisation. Recent studies have found out that child immunisation status is negatively associated with mother's employment (Shivakami, 1997; Jatarana, 2003). This kind of forced employment in poor households, therefore, rarely offer the women any benefit in terms of empowerment. Thus, not merely the employment, but the nature of employment is a crucial determinant of the health status and healthcare utilisation. A study on highly educated lady IT professionals in Kolkata identifies that women suffer

from untreated morbidity as they hardly get time to see a doctor when they are unwell. The general 'bargain model' within the dyad of husband and wife also challenges the general myth that an employed woman enjoys better empowerment and hence enjoys decision making power within a household.

Ethnicity is another factor having a very strong health gradient. It is often argued that this social determinant works through the economic conditions. Ethnic people often stay in remote places and in poverty. However, economic conditions cannot always account for all the difference in health status. The child-care practices, social beliefs, cultural norms, language barrier, difficulties in patient-provider communication, residential segregation and legacy of history often differ across ethnic and religious groups, thus resulting in health inequity. In the United States, black adults are more likely to die earlier than white adults (due to cardiovascular diseases, cancer, diabetes, etc.) and also face higher homicide hazards and incidence of drug addiction. Egregious examples of ethnic gradient on health come from comparisons of life expectancy at birth of indigenous people in developed countries like Canada, New Zealand, and Australia with that of the total population (Table 9.1) where in each case, the gap between the indigenous group and the total population is substantial. Australia, a developed country by any standard, finds her indigenous tribal men (living in Northern Territory and Saint Torres Islands) have life expectancy 17 years less than the average men.

In developing countries like India, ethnicity plays a crucial role in determining the healthcare status of the people. NFHS-3 (2005-06) finds that more than 20 percent women from Scheduled Tribe do not use any maternal care services (ante-natal checkups, consumption of iron tablets, anti-tetanus immunisation and institutional delivery), while the corresponding share for the general caste is 6.03 percent. Likewise, the infant mortality rate (IMR) and the child mortality rate (CMR) of ST children are far higher compared to the general caste population in India. As per Census 2011 figures, IMR for Scheduled Tribe population was 62.1 per 1000 live births in 2005-06, while for general castes the figure was 48.9 percent. The corresponding figures for CMR were 95.7 and 59.2 percents respectively.

Table 9.1: Life Expectancy of Indigenous People in Developed Countries (Years)

Country	Indigenous	All men	Gap
Australia (1996-2001)	59.4	76.6	17.2
Canada (2000)	68.9	76.3	7.4
New Zealand (2000-2002)	69.0	76.3	7.3

***Source**: Marmot 2006*

Another crucial source of inequality is ***education***. The focus on education has two rationale. One, education is expected to improve productivity and hence the earning capacity of an adult, thereby improving his health status. Two, it offers intrinsic benefits normally associated with better awareness level like healthcare practices for preventive healthcare. Such intrinsic benefit also flows inter-generationally. Better educated parents, especially mothers, have healthier children who grow up to become healthy adults. Better educated parents prefer to give their children better healthcare facilities, including immunisation coverage, safe drinking water, improved sanitation and proper nutrition.

On obesity, particularly in women, there is an inverse association with socioeconomic status. The relationship of obesity with education as a measure of socioeconomic position is depicted in Figure 9.3 where countries are classified on the horizontal axis by the degree of economic development. Above a GNP per-capita of above $2995 (i.e. upper middle-income economies), the higher the education the lower the obesity rate

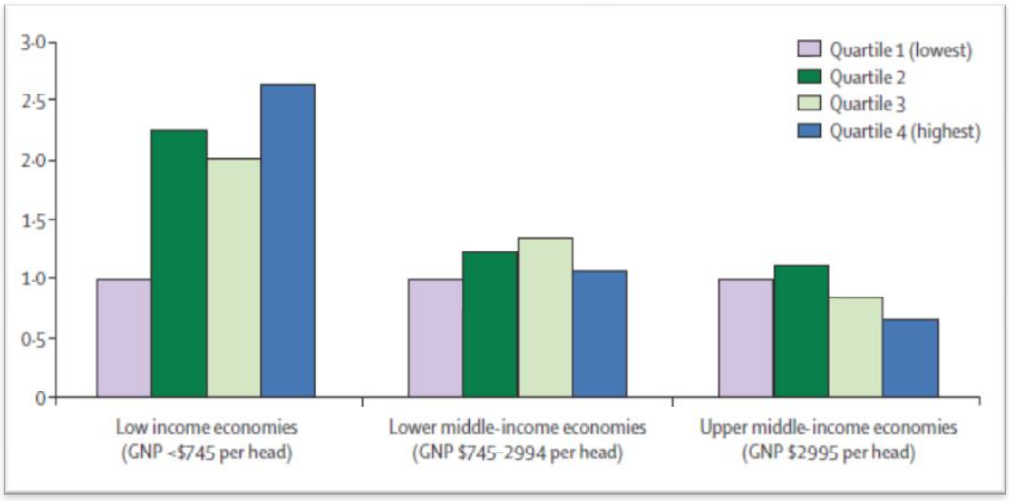


Figure 9.3: Women's Education by Education Quartiles and Income

Source: Monteiro, et al. (2004)

while the corresponding trend is opposite in low income countries. Similar differences are observed for risk factors like smoking, addiction and mental health. Differences in health behaviour thus can be attributed to the autonomy and social engagements.

Recent studies also indicate some inept impacts of education on healthcare benefits. Two reasons are suggested for this. One, there may be a possibility of reverse causality from good health to good education. Evidence suggests that healthy students regularly attend schools and perform better. Two, the quality of education imparted is of crucial importance particularly in developing countries like India, where the main thrust of education policy is on improvement in gross enrolment with the result that people receiving education are able to sign but have poor comprehensive and analytical skills. Apparently, education imparted fails to offer any significant intrinsic benefits by healthcare in these countries.

9.2.4 Healthcare Finance

Healthcare is essentially financed by three pre-dominant modes viz. public finance, third party payment through private insurance and out of pocket expenditure (OOPE). With healthcare reform spreading its wings across countries along with tight fisted fiscal policies, countries have adopted policies to restrict the first option only for the poorest and the most vulnerable sections of the population. Commercialisation of healthcare market has resulted in more dependence on private insurance market in developed countries, while sky-rocketing the OOPE in developing and poorer countries. A ready indicator to capture the process of commercialisation has been the share of a country's healthcare expenditure financed directly by private individuals (i.e. OOPE and private insurance). Among rich countries only the United States and Singapore have private healthcare spending share above 50 percent, while in an overwhelming majority of poor countries the private healthcare share is greater than 70 percent. Additionally, the poorer a country, the more likely the population is to face the most regressive form of health finance. Also, *'socialized medicine'* (i.e. healthcare financed through tax-based public expenditure and social insurance), appears a luxury good in economic terms since more of public financed healthcare is cornered by the people in the higher income segment relative to those in the lower income segments.

Given this pattern of healthcare finance, a few observations linking the type of finance and the health status across the countries is in order. *First,* countries with better health

outcomes are associated with lower commercialisation of health expenditure. As per WHO data in 2000, Healthy Life Expectancy (HALE) was significantly higher and child mortality lower in countries with lower ratios of private to total health expenditure. *Second,* countries spending more of private funding do not report to have better health status. *Third*, countries that spend more of their GDP on healthcare through public expenditure have significantly better health outcomes. More recent evidence indicates that higher spending via universal health coverage (UHC) impacts health outcomes positively, with larger benefits for poorer countries. Figure 9.4 plots the per capita public spending on health against IMR for 182 countries for 2011. It indicates that there is an inverse relationship between the two. An analysis of the relationship between health status and public spending across Indian states reveals that: (i) per-capita public health expenditure positively influences health status; (ii) poverty declines with better health; and (iii) growth and health have a positive two-way relationship.

In the healthcare market, information asymmetry primarily occurs due to the fact that consumers have information on marginal effect of health on his utility (∂U/∂HS) which the provider does not have. Again, the provider has information on effectiveness of healthcare service on health status (∂HS/∂HC) which the patient does not have. This typically leads to a market failure as due to lack of information on ∂HS/∂HC, patients demand unnecessary and irrational healthcare options. Since, Arrow's uncertainty principles make ∂HS/∂HC variable and unpredictable, both of these factors induce

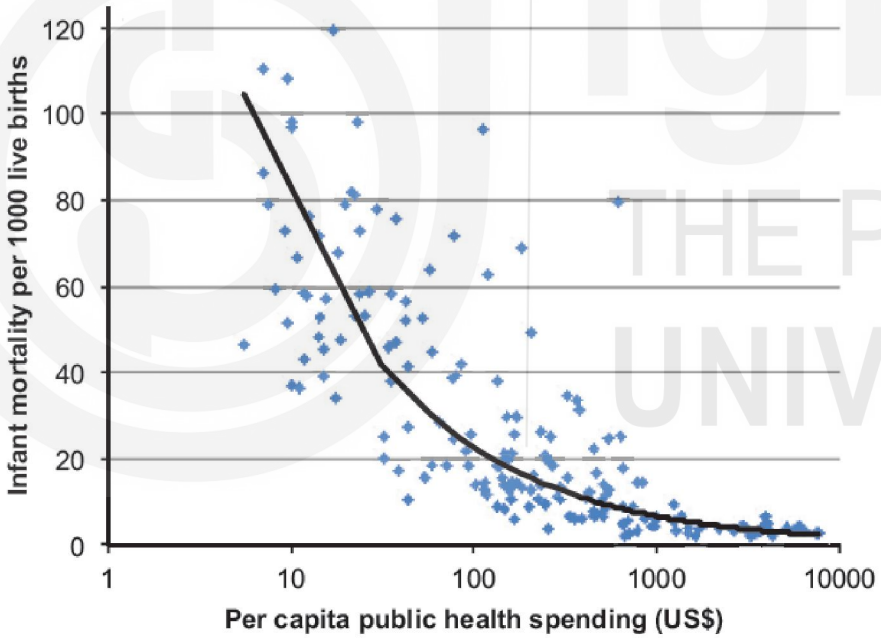


Figure 9.4: Relationship Between Per Capita Health Spending and IMR (2011)

Source: Gupta & Mondal 2014

the patients to be risk averse and they take private insurance by choice. The decision to buy no insurance rather than full coverage is more likely to be optimal with greater cost of insurance. However, with greater probability of individual's illness, even a risk-neutral individual will buy health insurance. The latter is also the case when some health services are indivisible and necessary for survival (like renal dialysis and organ transplants). Economic theory predicts that with an insurance policy, individuals become a bit more reckless and reduce the use of preventive healthcare practices (called Moral Hazard). The scope for this gives rise to asymmetric information between the insurance provider and patient thereby making the relationship between 'moral hazard' and

'asymmetry of information' two-sided. As a consequence, the demand for curative healthcare services eventually increase due to the presence of the third party insurance to pay for the costs.

For voluntary health insurance (as is the case in most developing countries), the consumer will choose to use a level of medical care and its associated levels of spending conditional to the hazards of facing a grave illness in order that the marginal health and welfare benefits of healthcare equals the marginal cost of that care in the market. That level of spending is exactly the amount the consumer would want to pay as premium for insurance. To put it differently, let us assume that

- the illness with uniform severity might strike the patient with a probability 'p';
- without insurance, the person is ready to spend S* on the cost of treatment for a disease; and
- he would compare the marginal benefits from additional spending for a package of care X*.

Given this, in absence of any moral hazard, the person will buy insurance for a premium (irrespective of whether the illness strikes him or not) if the cost of his actuarial premium is equal to the expected value of claims for covering the expenses. This will take place irrespective of the degree of risk aversion for all patients. The value of that additional cost would be $P = p\ S^*$.

But the situation would differ when the patient does not know what is the exact severity of the illness. At some severity level 'L', the person is ready to spend S*, and at a higher severity level L′ he is ready to pay S′. However, with the insurance coverage in the absence of any additional OOPE, the person may prefer S′ even in the state of L, because additional spending provides positive marginal benefits. Thus, the consumer will claim to be on higher severity level L′ (where the expected value of cost of healthcare with insurance is higher than in the case of without insurance) which amounts to moral hazard. Thus, moral hazard leads to consumption of healthcare that is less worth than its benefit. With severity L, the person consumes S-S* additional care, which will not offer him benefits less than that value. The premium will rise from $P' = p(L)\ S^* + p(L')S'$ to $P'' = \{p(L) + p(L')\}\ S'$ and hence additional consumption of healthcare would lead to additional premium not worth it.

Check Your Progress 2 [answer the questions in about 100 words in the space given]

1) Identify the correlates between income and health to bring out that it is a two-way relationship.

.....

.....

.....

.....

2) Do you agree that the difference in view points between economists and public health exponents on public healthcare expenditure is similar to the debate on 'growth' versus 'development'?

.....

.....

.....

.....

3) What is the empirical evidence that has established the relationship between income and health as a merit good?

.....

.....

.....

.....

4) What is the empirical evidence that is available to contest the hypothesis of 'income protecting health'?

.....

.....

.....

.....

5) Do you agree with the view that 'food energy intake' is a good measure to reveal the relationship between 'poverty and malnutrition'? Illustrate your answer.

.....

.....

.....

.....

6) Do you agree that in the prevailing employment conditions in developing countries, characterised by large informal sector employment, employment of women makes a good deal of difference as a determinant of health status and healthcare?

.....

.....

.....

.....

7) What is an empirical indicator generally used to measure the commercialisation of healthcare in countries? What is meant by 'socialised medicine' and why is it regarded as a luxury good?

.....

.....

.....

.....

8) What major findings flow from the studies on Indian states in respect of public spending on healthcare ?

.....

.....

.....

.....

9) How is the relationship between 'moral hazard' and 'asymmetry of information' two-sided? What is an unhealthy consequence of this feature?

.....

.....

.....

.....

10) Illustrate how additional consumption of healthcare due to moral hazard leads to additional premium cost incurred unworthy of it.

.....

.....

.....

.....

9.3 DETERMINANTS OF SUPPLY OF HEALTHCARE SERVICES

The supply of health workforce, engaged in delivering different types of healthcare services, is supposed to have direct impact on the health status of the population. Health workforce individually and taken together collectively can be seen as a profit-maximising economic actor. We can see this from the perspectives of price, wage, efficiency, etc. as follows.

9.3.1 Price, Wage and Supply of Health Workers

Individuals are in the first place tempted to take to higher education more out of the attraction to the earning potential that such a professional career commands. Thus, higher demand for healthcare services and higher expected income play crucial roles in embracing the career in medical education. Historically, many developing and developed countries in the world have experienced severe shortages in the supply of health workers, particularly doctors and nurses. As a result, the market for doctors and nurses have suffered from the pressure of excess demand. This situation of excess demand puts upward pressure on their prices i.e. their wages or salaries. However, being far from a perfect market, the price mechanism in the market is often characterised by one or more of the following situation.

- Since the medical education market remains highly controlled, any automatic increase in the supply of medical workforce cannot be ensured. As a result, imperfection creeps into the market of medical education with hefty charges paid for a seat in the medical colleges.
- Doctors then attempt to increase their income by resorting to factors like *demand inducement*. A positive association between the density of health workforce (captured by physician density per thousand population) and their price is often

interpreted as evidence of monopolistic market and/or physician's inducement of demand for medical care.

- In expectation of further income enhancement, workforce often migrate from developing to developed countries. This phenomenon of export of health workers from developing to developed countries has historically seen out-migration of trained persons from countries like India, Philippines and Vietnam in huge numbers to countries in Europe, Middle East and African countries.

- The above imperfection resulting from a state controlled healthcare system, does not allow the price for doctors' services to be adjusted by the market. As a result, the price of health services itself reflects an average quality of service in the market. The workers with above-average quality would then opt for high salary positions in the private sector or migrate to other countries. This would further lower the average quality of the workforce in the market. This process would repeat itself till the market is characterised by low-quality doctors and nurses. Consumers too shift their demand to unregulated medical care, especially in developing countries. As a result, studies indicate that the poor patients often prefer untrained quacks or Rural Medical Practitioners (RMP) instead of doctors in the public sector. In India, the predominance of demand for services from RMPs is a combined result of market imperfections and information asymmetry.

In short, due to the presence of limited price mechanism in the market of medical manpower especially in developing countries, doctors and nurses either make use of non-price mechanisms like demand inducement to increase their income or exit from a market in pursuit of higher wages in other places. Quality too tends to suffer in the absence of perfect market forces to establish the required adjustment in the healthcare market. In Arrow's words, '*the general uncertainty about the prospects of medical treatment is socially handled by rigid entry requirement*' (1963).

9.3.2 Organizational Change and Technical Efficiency

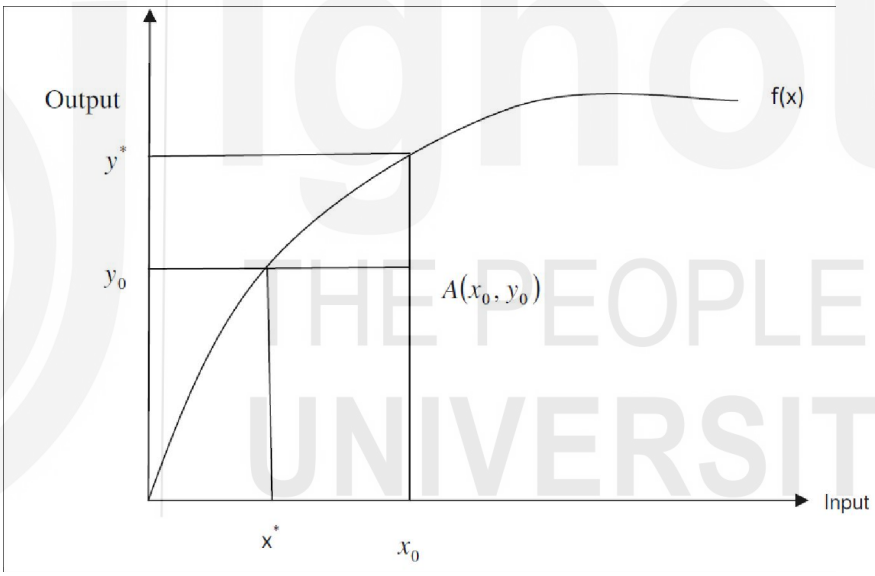
Variations in the efficiency of healthcare workers could also be due to the differences in the efficiency levels of healthcare infrastructure as well as the presence of other adverse correlates of health outcomes like social determinants (like poverty and lack of education). The significance of organisational factors is underscored in this context in terms of higher budgetary provision which often help countries and specific hospital units to succeed in having the required allocations made in this regard. The measurement of technical efficiency is helpful to assess such organisational factors which in the case of a healthcare system is often calculated by using non-parametric methods like Data Envelopment Analysis and Stochastic Frontier Approach. These methodologies do not make any assumptions on the distribution of outputs nor do they need any *a-priori* information related to price. Hence they are currently being used in sectors like health and education, where profitability is not considered as the prime motive in the industry.

Measurement of technical efficiency can be of two types: (i) input oriented technical efficiency measure; and (ii) output oriented technical efficiency measure. The former deals with the minimum amount of input quantities (which can be proportionately reduced) without changing the quantities of output produced, while the latter deals with the maximisation of quantities of output (that can be proportionately increased) without altering the quantities of inputs. In case of output oriented measure, the TE of a Decision Making Unit (DMU) is computed by comparing its actual output with the maximum producible quantity on the frontier. In input oriented measure, the TE of a DMU is measured by comparing its actual input in use with the minimum input that

would produce the targeted output level i.e. by identifying how much can input quantities be proportionately reduced for each DMU without changing the actual output bundle.

The idea of input and output oriented technical efficiency can be presented by a graphical presentation for the case of a single input and single output (Figure 9.5). The horizontal axis represents the input used by the firm and the vertical axis shows the output produced with the point $A(x_o,y_o)$ representing the actual input-output bundle from which $y^*= f(x_0)$, the maximum output producible from input $x_{0,}$ can be potentially realised. The *output-oriented* measure of technical efficiency is given by $TE_o = (\frac{y_o}{y^*})$, which is the comparison of actual output with the maximum producible quantity from the observed input. Assuming that for the output level y_0 the quantity of input can be proportionately reduced till the frontier y_0 is reached with an input x*, the *input-oriented* technical efficiency measure for the concerned DMU is given as $TE_I = \frac{x^*}{x_o}$. Since the value of TE, calculated by this method, lies between 0 and 1, a score of 1 will signify that the firm is fully efficient and a score below 1 signifies inefficiency.

Figure 9.5: Input and Output Oriented Technical Efficiency



The healthcare finance, along with other inputs, is often not fully utilised to reach the maximum output (health status) and therefore there remains a large scope for improving the effect of health finance on the health status. Since improving the efficiency of spending would remain a critical strategy in the financing for 'universal health coverage', funding would remain an important correlate in improving the health status of people.

9.3.3 Pharmaceutical Pricing

The global pharmaceutical industry selling vaccines, drugs, consumables, etc. represents classical monopolistic competitive conditions with patents and product differentiation leading to prices that are far higher than the production and marketing costs. With firms attempting to move to partial monopoly by introducing newer products, they undergo huge R&D costs, as well as promotional expenditures, sometimes sacrificing the short run profit. Patents, in the form of guarantee document for the intellectual property of the firm, offers virtual monopoly to the patent holder to charge high price for the product for the entire term of the patent. After the expiry of patent term, the companies face the competition from generic producers forcing them to reduce the prices. However, under the existing market structure, many firms go for '*ever-greening*

of patents' by applying for new patent term, often foregoing their earlier patent term by one or two years, with little value added to the existing product so that they can enjoy the monopoly rights for a longer period. The prices at which the drugs and vaccines are supplied in the market, therefore, seldom represent the pure market conditions. Further, the political economy behind the pharmaceutical lobby often controls the prices of medicines. Added to this is the induced demand created by brand-sensitive physicians, who continue to prescribe high priced originator brands even when low price generics are available in the market. A partial market which is price-inelastic, therefore, co-exists with another part of highly elastic market segment. These characteristics often result in the co-occurrence of a plethora of generics for each molecule in any market.

With the pharmaceutical industry always appearing to be at or near the top of industry rankings in terms of after-tax profit returns, most of the countries, apart from patenting system, offer a host of institutional arrangements for controlling the prices and profits of the pharmaceutical companies. To curtail the burden of the citizens by the rising healthcare costs, most of the developing as well as developed countries, impose pharmaceutical price controls. Though there are countless variations in the ways governments control prices, mainly they take the form of any one or combination of following five forms: (i) reference pricing; (ii) case-to-case negotiation; (iii) formula pricing; (iv) profit regulations; and (v) capping constraint controls. Under the first, comparable drugs are put in a reference category and prices are allowed only at the lowest of the group. This, however, creates non-price pseudo competition particularly in developing countries. The pharma giants refuse to reduce their price in these countries in accordance with the prevailing market conditions. General over-the-board price regulation system in developing countries are often used as a safeguard to control the healthcare costs. Although it is argued that these controls, in effect, distort the pharmaceutical and healthcare market, it must be recognized that leaving everything to the markets would lead to extremely high prices inducing more inequity and vulnerability particularly for the large number of poor in developing countries.

9.3.4 Technology and Healthcare

New technologies have revolutionised the ways in which the healthcare services are practiced. Among the best selling prescription drugs, nearly 10 percent are new every year. Advances in medical technology, involving both diagnostics and treatment, have been a driving force behind the rapid growth of healthcare expenditures. Research indicates that the growth of medical technology is a prime cause for quadrupling the per-capita healthcare costs over the period 1970 to 1986. Due to the availability of new technology, previously untreatable conditions become treatable and hence one encounters a larger and unexpected medical care costs than before. This essentially means a higher mean and variance of expenditure. Thus, a rational individual would like to demand not only higher medical services, but also higher insurance coverage to protect himself from major expenditure like the organ transplant resulting in higher average health expenditure over the life time of a person. On the other hand, the discoveries of newer vaccines of dreaded diseases have resulted in a reduction of individual's expected level of expenditure for treating the disease in the long run. Thus, the nature of new technology actually determines the degree and direction of change in the demand for healthcare. Three types of technology can be typified as follows:

- ***No-technology***: It offers better provider-patient relationship by pain management and psychological support (for diseases like advanced stages of cancer, severe arthritis, stroke, etc.).

- ***Half-way technology***: It adjusts to disease and postpones death (e.g. organ transplant, chemotherapy, radiation therapy, etc.).
- ***High technology:*** It comes from the genuine understanding of disease for preventing the incidence of disease (vaccines, prevention of nutritional disorders, etc.).

The effect of technology on medical care would depend on how far the high technology is being replaced by halfway technology or no technology. The halfway technology is encouraged by cost-reimbursement insurance system that has dominated hospital and medical care so far.

The issue of technology and supply of pharmaceutical products is thus linked to the patent system. Patents are of two types: product patent and process patent. The former stops anybody else other than the patent holder from producing the product itself, while the latter stops the competitors from producing the product by the same process as the patent holder. Surely, the former is far more stricter patent framework than the latter. The global pharmaceutical giants invest a large share of their turnover in R & D trying to hold patent on the newer drugs. Thus, they charge very high monopoly price on the originator brands. After the expiry of patents, competition starts coming from generic producers, when under competition, the price of the drugs tend to fall. However, studies have found out that even under generic competition, originator patent holder does not reduce the price as he enjoys a substantial market share by promoting their brands through the physicians. The doctors, who do not pay for the drugs, choose the originator brands for the patients, thus undermining the impact of patent expiry on the healthcare costs.

9.3.5 Government Policy

Earlier in section 9.3.3 we outlined the government's efforts to control prices of medicines and vaccines to ensure the supply of drugs to the wider population. However, evidence suggests that even with strict control in prices the governments have not sufficiently succeeded in providing access to medicines in majority of countries. According to WHO estimates, nearly a third of the world's population lacks access to the most basic essential medicines, while in the poorest parts of Africa and Asia this figure is close to half (WHO, 2003). In contrast, only 0.3 percent of population in high-income countries lack access, thus putting a highly disproportionate share of burden of non-access on the developing and poorer countries.

India is a paradox in this context, where more than 65 percent of population has absolutely no access to essential medicines, in spite of enough industrial capabilities to become the third highest producer of drug volume in the world and one of the largest exporters of generic drugs. India, known as the *'global pharmacy of South'* suffers mainly from a lack of proper public distribution facilities. Recently, many states have taken up specific policies to improve access to medicines in the country. In the Tamil Nadu model, considered the benchmark model to provide free medicines to every patient visiting public health facilities, an autonomous public sector body [named Tamil Nadu Medical Services Corporation (TNMCS)] procures high quality generic medicines through a transparent bidding system and then supplies to the public health facilities through a demand sensitive coordination. Bihar, Delhi, Madhya Pradesh, Kerala and Rajasthan have also come up with similar, though differentiated models. West Bengal, for the first time, introduced a Public Private Partnership model to achieve huge cost cutting of existing marketed products by introducing nearly one hundred Fair Price Medicine shops across the state.

Under the aegis of healthcare reform, many Asian and African countries have introduced Performance Based Finance (PBF) in order to bypass the typical *principal-agent* problem of the public sector health facilities. The idea behind PBF was, essentially, to incentivise the manpower at the supply side to improve the delivery of services. PBF is defined as a mechanism by which the health providers are, at least, partially funded on the basis of their performance. It is in contrast to the *line-time approach* where a health provider offers the inputs, including manpower, equipment and drugs. By increasing the income of health providers, PBF aims to bring together greater motivation and better technical efficiency, and even encourage staff to work in remote areas. While increasing the overall service utilisation, it can also lead to greater accountability of the health workers. Critics, however, point out that expansion of service utilisation might result in severe inequity with the potential to adversely affect the related non-incentivised services.

India has also introduced the PBF in the form of Accredited Social Health Activist (ASHA) under the National Rural Health Mission (NRHM). ASHAs are expected to be paid according to incentive schemes based on PBF to encourage Institutional Delivery (ID), Antenatal Check Up (ANC) and Complete Immunization (CI). They are paid in piecemeal for each ANC taken, ID in government hospital, accompanying mothers for ID, etc. Though several studies have shown significant improvement in healthcare utilisation after the initiation of NRHM, some counterproductive evidence of neglecting non-incentivised services and unequal utilisation of maternal care services have also come up.

Check Your Progress 3 [answer the questions in about 100 words in the space given]

1) In a government controlled education market for medical personnel, how does the price mechanism fail to establish an efficient market for healthcare manpower?

.....

.....

.....

.....

2) Distinguish between the two major types of technical efficiency measurement techniques that are usually applied in industries like health where profitability is not the main priority.

.....

.....

.....

.....

3) Do you agree that 'finance' would remain an important correlate in improving the health status of the population? Why?

.....

.....

.....

.....

4) What is meant by 'ever greening of patents'? How does this, along with other factors, influence the prices in the healthcare market to remain high?

.....

.....

.....

.....

5) What mechanisms are usually adopted by both the developed and developing countries to control the monopolistic behaviour of pharmaceutical industry? How does it work?

.....

.....

.....

.....

6) Do you agree that the issue of technology and the supply of pharmaceutical products is linked to the system of patents? How?

.....

.....

.....

.....

7) How is the 'performance based finance' (PBF) system defined? What has been the feedback on its working in India?

.....

.....

.....

.....

9.4 LET US SUM UP

The demand and supply of healthcare ideally do not follow similar economic logic as for other goods. Though income, wage and prices interact in some way to control the market for health services, the unique nature of healthcare affecting the consumer's utility via the indirect route of controlling health status needs to be recognised. Also, healthcare utilisation is profoundly affected and determined by several non-economic social factors which result in serious inequity and vulnerability. Education, ethnicity and occupation impose strong impact not only on healthcare demand, but also on the final health outcome. Likewise, several non-market relations and government interventions often distort the supply of healthcare goods, including medicines. Existence of patents, price and profit controls, barriers to migrate, etc. create strong effects on the supply of health goods and services. Additionally, the impact of technology on healthcare costs and hence demand would depend on the nature of technology per se, rather than the cost involved.

9.5 KEY WORDS

Structural Determinants : Are characteristics that generate social and economic stratification, such as economic, political, social and education systems. The structural determinants thus capture the nature of macro development issues.

Proximal Determinants : Proximal determinants are the circumstances of daily life like the quality of the family environment and peer relationships, availability of food, housing, recreation, access to education, etc. The proximal determinants concern themselves with the micro issues across and within the household framework.

Fuzzy Demand : A situation where it is extremely difficult to accurately delineate the relationship between the price and the quantity demanded for medical care.

HAZ/WAZ : Variables used for representing chronic nutritional deficiency and acute malnourishment respectively. HAZ relates to stunted height while WAZ refers to excess weight.

Performance Based Finance (PBF): PBF is defined as a mechanism by which the health providers are, at least, partially funded on the basis of their performance. It is in contrast to the *line-time approach* where a health provider offers the inputs, including manpower, equipment and drugs.

Arrow's impossibility theory : The Arrow's impossibility theorem is a social-choice paradox illustrating the impossibility of having an ideal voting structure that is reflective of specific fairness criteria, such as Pareto efficiency. Arrow's impossibility theorem states that a clear order of preferences cannot be determined while adhering to mandatory principles of fair voting procedures.

9.6 SUGGESTED REFERENCES FOR FURTHER READING

1. Arrow, K (1963), Uncertainty and the Welfare Economics of Medical Care, *American Economic Review* 53: 941–973.

2. Gallup, John Luke, and Jeffrey D. Sachs. 'The Economic Burden of Malaria.' *The American Journal of Tropical Medicine and Hygiene,* 64.1 suppl (2001): 85-96.

3. Grossman, M (1972), On the Concept of Health Capital and the Demand for Health, *Journal of Political Economy,* 80: 223–250.

4. Gupta Indrani & Swadhin Mondal (2014), Health Spending, Macroeconomics and Fiscal Space in Countries of the World Health Organization South-East Asia Region. WHO South-East Asia Journal of Public Health, July–December 2014, 3 (3–4).

5. Hurley Jeremiah (2000), An overview of Normative Economics of the Health Sector *in* A. Culyer and J. Newhouse (eds.), *Handbook of Health Economics*, Vol. 1A, Elsevier, Amsterdam, 347–408.

6. Jatrana, S (2003), 'Explaining Gender Disparity in Child Health in Haryana State of India', *Asian Meta Centre for Population and Sustainable Development Analysis*, Paper Series No. 16.

7. Marmot, M (2006), Health in an Unequal World. 368: 2081–94.

8. Pritchett Lant & Lawrence Summers (1996), Wealthier is Healthier, *The Journal of Human Resources,* Volume 31, No 4: 841—868.

9. Sivakami, M (1997), 'Female Work Participation and Child Health: An Investigation in Rural Tamil Nadu, India', *Health Transition Review*, 7: 21-32.

10. World Health Organization (2003), 'The World Health Report: Shaping the Future', Geneva.

9.7 ANSWERS/HINTS TO CYP EXERCISES

Check Your Progress 1

1) See 9.1 and answer.
2) See 9.2 and answer.
3) See 9.2 and answer.
4) See 9.2 and answer.
5) See 9.2 and answer.
6) See 9.2 and answer.
7) See 9.2 and answer.
8) See 9.2 and answer.

Check Your Progress 2

1) See 9.2.1 and answer.
2) See 9.2.1 and answer.
3) See 9.2.1 and answer.
4) See 9.2.1 and answer.
5) See 9.2.2 and answer.
6) See 9.2.3 and answer.
7) See 9.2.4 and answer.
8) See 9.2.4 and answer.
9) See 9.2.4 and answer.
10) See 9.2.4 and answer.

Check Your Progress 3

1) See 9.3.1 and answer.
2) See 9.3.2 and answer.
3) See 9.3.2 and answer.
4) See 9.3.3 and answer.
5) See 9.3.3 and answer.
6) See 9.3.4 and answer.
7) See 9.3.5 and answer.