UNIT 7  DEMAND FOR HEALTH SERVICES

Structure

7.0  Objectives

7.1  Introduction

7.2  Health Indicators
   7.2.1  Mortality Indicators
   7.2.2  Morbidity Indicators
   7.2.3  Burden of Disease Indicators

7.3  Health Indicators and Economic Development: Linkage

7.4  Role of Economics in Health Sector
   7.4.1  Efficiency Wage Hypothesis
   7.4.2  Health and Human Development

7.5  Externalities in Health
   7.5.1  Health: Public Good and Merit Good
   7.5.2  Distinction Between Private Demand and Social Demand for Health

7.6  Role of Health in Economic Development

7.7  Demand for Health Versus Traditional Demand Function
   7.7.1  Becker’s Theory of Behaviour
   7.7.2  Grossmann’s Human Capital Model of Demand for Health

7.8  Supply Factors Affecting Demand for Health
   7.8.1  Role of Physicians
   7.8.2  Role of Pharmaceutical Companies
   7.8.3  Role of Health Insurance

7.9  Let Us Sum Up

7.10  Key Words

7.11  Some useful books and references

7.12  Answers/Hint to CYP Exercises

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 lifetime. 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 quality 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.

\[
\text{CBR/CDR} = \frac{\text{Number of live births} + \text{deaths mid-year total population}}{} \times 1000
\]

**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:

\[
X_0 = 0, \ X_1 = 1, \ X_2 = 5, \ X_3 = 10, \ldots, \ X_{21} = 100
\]

Let \( M = (M_i)_{21 \times 1} \) be the vector of mortality probabilities denoting the probability of dying between age \( X_{i-1} \) and \( X_i \), for \( i = 1, \ldots, 21 \). 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’, \( \text{LE}_i \) is:

\[
\text{LE}_i = (X_i - X_{i-1})p_i + a_i d_i
\]

where, \( X_{i+1} \) 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:

\[
p_i = P_j=1 \{ 1 - M_j \} \quad \text{with} \quad p_0 = 1 \quad \text{and} \quad d_i = p_{i-1}M_i
\]

Now, the life expectancy at birth is: \( \text{LE} = \sum_{i=1}^{21} \text{LE}_i \)

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 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 quantify 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 often 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)** = \( \frac{LE - 20}{85 - 20} \)

2. **Education Index (EI)** = \( \frac{MYS + EYS}{2} \) where

   \( MYSI = \frac{MYS}{15} \) and \( EYSI = \frac{EYS}{18} \)

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

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

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

**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 school). 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 one’s 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:

\[
P = V(LE) - wL. \tag{7.1}
\]

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:

\[
dP = \frac{\partial V}{\partial (LE)} LE dE - L dw,
\]

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

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 \( \frac{dP}{dw} \geq 0 \). 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 Millennium 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](image)
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 resultant 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:

\[ U = U(Z_1, Z_2, ..., Z_n) , \]  

(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:

\[ Z_i = Z(x_i, t_i; E) , \]  

(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:

\[ T = t_w + \sum_{i=1}^{n} t_i , \]  

(7.6)

as well as the usual income constraint:

\[ I = \sum_{i=1}^{n} p_i x_i , \]  

(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:

\[ S = wT + V = \sum_i (w t_i + p_i x_i) \]  

(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:

\[ L = U(Z_1, Z_2, ..., Z_n) - \bar{\varepsilon} [ \sum_i (w t_i + p_i x_i) - S] \]  

(7.9)

The first order conditions for maximization with respect to the commodities imply:
The ratio of the marginal utilities of any two commodities $Z_i$ and $Z_j$, i.e. $\frac{MU_i}{MU_j}$, will be equal to the ratio of their marginal costs, $\frac{\pi_i}{\pi_j}$, 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:

$$
\frac{\delta U}{\delta z_i} = \frac{MU_i MP_{ik}}{MU_j MP_{jl}} \frac{pf_{ik}}{pf_{jl}},
$$

(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 $\delta / \hat{\delta}$, where $\hat{\delta}$ is an index of all commodity prices. It may also affect the general price level $\delta$ itself by raising or lowering the average $\delta$ 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. $S/\delta$.

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 lifetime. 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. 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:

$$U = U \left[ r(H_0), X_0 \right] + \beta \ U \left[ r(H_1), X_1 \right],$$

(7.12)

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

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 $\delta$, 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 $t$ units of time on preventive effort. We thus have:

$$H_1 = H_0 (1 - \delta) + I (M_0, t_I),$$

(7.13)

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

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:

$$A_0 + w_0 (1 - t_r^1 (H_0) - t_I^0) + \frac{w_1 + (1 - r^1 (H_1))}{R} = pM + cX_0 + \frac{cX_1}{R}$$

(7.14)

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

$$L \left( H_1, t^1, M, X_0, X_1 \right) = U \left[ r^1 (H_0), X_0 \right] + \beta U \left[ r^2 (H_1), X_1 \right] +$$

$$\mu \left[ H_0 (1 - \delta) + I (M, t^1) - H_1 \right] +$$
\[ \lambda \left[ A_0 + w_0 \left( 1 - t' \left( H_0 \right) - t' \right) + \frac{w_1 \left( 1 - t'_1 \left( H_1 \right) \right)}{R} - pM - cX_0 - \frac{cX_1}{R} \right] \]

The Lagrangian multipliers \( \mu, \lambda > 0 \) 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 \( \mu \) and \( \lambda \) simply ensure that the constraints (7.13) and (7.14) are satisfied. Treating \( H_0 \) as given, the first order conditions are:

\[ \frac{\partial L}{\partial H_1} = \beta \frac{\partial U}{\partial t'} \frac{\partial t'}{\partial H_1} - \frac{\lambda}{R} w_1 \frac{\partial t'}{\partial H_1} - \mu = 0 \quad (7.15) \]

\[ \frac{\partial L}{\partial t'} = \mu \frac{\partial I}{\partial t'} - \lambda w_0 = 0 \quad (7.16) \]

\[ \frac{\partial L}{\partial M} = \mu \frac{\partial I}{\partial t'} - \lambda p = 0 \quad (7.17) \]

\[ \frac{\partial L}{\partial X_0} = \beta \frac{\partial U}{\partial X_0} - \lambda c = 0 \quad (7.18) \]

\[ \frac{\partial L}{\partial X_1} = \beta \frac{\partial U}{\partial X_1} - \lambda \frac{c}{R} = 0 \quad (7.19) \]

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

\[ \frac{\partial I / \partial t'}{\partial I / \partial M} = \frac{w_0}{p} \quad (7.20) \]

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

\[ \frac{\partial U / \partial X_0}{\partial U / \partial X_1} = \beta R \quad (7.21) \]

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

\[ -\beta \frac{\partial t'}{\partial H_1} \left[ \frac{w_1 \frac{\partial U}{\partial X_1} - \frac{\partial U}{\partial t'}}{c} \right] = \mu \quad (7.22) \]

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

\[ \mu = \frac{\partial U / \partial X_0}{\partial I / \partial M} \frac{p}{c} \quad (7.23) \]

Substituting (7.22) in (7.23) yields:

\[ -\beta \frac{\partial t'}{\partial H_1} \left[ \frac{w_1 \frac{\partial U}{\partial X_1} - \frac{\partial U}{\partial t'}}{c} \right] = \frac{\partial U / \partial X_0}{\partial I / \partial M} \frac{p}{c} \quad (7.24) \]
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 \( \text{will actually be} \) ‘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](image)

### 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, \( L_i \), with the probabilities, \( p_i \), serving as the weights. The expected loss, therefore, equals:

\[
\mu = p_1L_1 + p_2L_2
\]  

(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 \( \mu \), 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


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.