CHAPTER 2

LITERATURE SURVEY

This chapter is organized as follows: Part I deals with need for and meaning of equity in health, part II deals with measurement of health, part III with measurement of health inequalities and part IV with determinants of health inequalities.

Part I: Need for and meaning of health equity

Need for equity

The treatise of welfare economics always attach a positive value to a higher mean value of a good and a negative value to the inequality in its distribution around the average. Reduction in inequalities is a matter of social justice. A society characterized by distributive injustice is unstable and contains within it the most potent seeds of social strife and revolution.

It may be argued that there is a stronger reason for being less tolerant of the inequalities in health than inequalities in the distribution of any other good, even income! The reason being that health is a special good that has both intrinsic and instrumental value. It directly affects a person’s well being and is an integral component of all the diverse notions of well being that exist in the literature. In this sense, inequalities in health directly bring about a skewed distribution of well being among the population.

Health is also a prerequisite to a person’s functioning as an agent. So inequalities in health would constitute inequalities in people’s capability to function. Isaiah Berlin would call this as an unequal distribution of the “positive freedom” of the people. According to Sen, impairments to health constrain peoples’ capability to “being & function”. This according to John Rawls is a denial of “fair equality of opportunity” while his conception of justice is anchored around ‘fairness’. So justice would require a fair distribution of health. Norman Daniels extended this notion of justice to fair access to health care.
While the ‘incentives’ argument could be used to take a somewhat lenient view of inequality in the distribution of income but no such leeway is permissible in the case of health. Health is a merit good, whose distribution should be divorced from the ability to pay for it. Nobel laureate James Tobin (1970) had labeled a phenomenon called “specific egalitarianism” to say that equity considerations feature prominently in the distribution of some commodities such as health.

On the demand side also, economists are of the view that the commodity health care has certain special characteristics because of which the term need and not demand is applied to health care unlike other consumption goods. Arrow (1963) and Le Grand (1991a) are of the opinions that demand for health care is stochastic variable because it is needed only when people fall sick and falling sick itself is a random variable. Therefore it is the term “need” Williams (1974) that has to be applied to health care and distributive justice has to be meted out to people.

Equity in the field of health / health care is a pressing need in the present times because a long history of medical care and preventive healthcare has not been able to address why some groups have a systemically higher rate of disease than others.

The need is further reinforced because this is a technology driven age when even health care is becoming technology intensive and high cost so that the vulnerable sections of the society face formidable barriers in obtaining the most basic existential necessity of life.

Under these circumstances, attaining egalitarianism becomes an important desideratum of health policy. However, unlike the case of income distribution, both the meaning and the need for equity in the case of health distribution are debatable issues. Equity in health has several interpretations and the need for equity in health is constantly contested by the need for efficiency in the provision of health.

Nonetheless, the governments of most countries and international agencies operating in the field of health have shown strong commitment to closing the gap in the health outcomes of different segments of society. The UK Department for International

---

1 The demand for health care is stochastic and conditional upon the probability of falling ill and its demand is characterized by the agency problem.
Development (1999) has committed itself to reducing health inequalities in UK and to focus its aid on improving the health of the world's poorest.

The National Health System in the UK was founded in 1941 with the objective of attaining distributional equity in the field of health care as a means to attaining equity in health. In OECD countries, equity is a prominent issue in the debate on health care financing and delivery reforms (van Doorslaer, Wagstaff et al 1993). Even policy makers in developing countries agree that equity should feature prominently in health policy decisions (Gilson 1998). The Medicaid and Medicare plans were initiated in the USA specifically to deal with health problems of the poor and the elderly respectively (the two most vulnerable groups in the society). The extent of inequalities that prevail in India and the cognizance taken in the official policy documents to bring about a more equal distribution of health have already been glanced at in the previous chapter.

From the foregoing discussion it appears that the need for health equity is incontestable. But there are several contradictions in the conceptualization of the notion of equity in health. The following discussion serves to highlight some of the key issues in the matter.

**Meaning of equity**

The notion of health equity is surrounded by some bit of mystic on account of the fact that the issue of equity in health can be decomposed into equity in health versus equity in health care. This is so because health by itself is intangible and like income cannot be transferred across individuals and therefore any change in the distribution of health is most often sought to be brought about by changing the distribution of health care. (Our study is precisely to inquire whether it is health care or are there some other more influential determinants of the distribution of health)

In the policy documents of many countries around the world, one finds that health equity objectives are stated in terms of equity in health care rather than equity of health. The issue becomes more complex by the absence of any singular notion of equity in health care.
Mooney and Le Grand made very significant beginnings in clearing the haze around the matter by identifying several definitions and objectives of equity in health. These have been listed in the following paragraphs\(^2\).

Mooney (1983, 1986) has proposed the following definitions and objectives for equality:

1. Equality of expenditure per capita
2. Equality of inputs per capita
3. Equality of inputs for equal need
4. Equality of access for equal need
5. Equality of utilization for equal need
6. Equality of marginal met need
7. Equality of health

Le Grand ([1982, 1987a, and 1991a]) has proposed the following types of equality as the guiding objective of public expenditure distribution:

1. Equality of public expenditure
2. Equality of final income
3. Equality of use
4. Equality of cost
5. Equality of outcome

Only the last items in both the schemes above identify the equity objective in terms of equality in health. All the other criteria have been laid down in terms of equity of health care and they essentially deal with "equality of public expenditure per capita" or "distribution of health care according to need" or "equality of access".

**Equality of public expenditure per capita** simply requires that the budget allocations the government should be such that the public expenditure on health should be equal for all persons and there should be no systemic discrimination against individuals or groups of people in this respect. This definition obviously falls short of meeting the desired objective because it does not make any allowance for 'need'.

\(^{2}\) Although they are not the only ones to have identified the equity objectives but they were the forerunners in subject. Since ten others have also done commendable work in identifying and clarifying these issues but we are omitting them to avoid some overlaps.
Distribution according to need This criterion means that health care should be distributed to people according to 'need'. This distribution would need to satisfy both horizontal equity and vertical equity. Horizontal equity would merit similar treatment of people who are in equal need and vertical equity would require that people with greater need should be treated more favorably.

But the term need itself has been a subject of frequent economic and philosophical polemics, and has been defined in the following alternative ways in the literature.

1) Need as initial health
   Need may be defined as initial health. This means that need is taken to mean ill health. This is the view taken by Williams (1962), Gillon (1985), Le Grand (1978), O’Donnell and Propper (1991) and Wagstaff et al(1991a). People with similar health statuses are said to have the same need for health care while people with dissimilar health statuses are said to have different needs for health care. This is the most frequently pursued objective of health care equity.

2) Need as capacity to benefit
   This has been proposed by Barry (1965). This is said to be a consequentialist or instrumental objective which says that we should have a goal that we wish to attain, and then need may be defined as whether that goal can be attained or not. In the context of health care it would mean two things (Culyer & Wagstaff 1993):
   i. The expected marginal productivity of health care should be positive. In other words, if health care is ineffective then it is not needed or an equivalent way of putting it is that there should be an expected capacity to benefit from health care (William 1974).
   ii. A positive marginal product is necessary but not a sufficient condition for a need to exist: there may be other less expensive or more productive technologies which fulfill the goal more efficiently (Culyer 1978, 1989). An individual may need health care but may not be ill (preventive measures) or be ill but not need health care (when no health care is effective though on ethical grounds may still need palliative health care) or

3 In the case of health, the application of these principles was first proposed by West (1981).
be ill and need health care of type X which may be more cost effective but not type Y which may be unnecessarily more expensive.

3 Need as expenditures a person ought to have

Need may be defined as expenditures a person ought to have. Culyer (1976). This is a purely normative assessment of the need for health care by an individual, which may be decided by the society in general.

4. Need as expenditures required to exhaust capacity to benefit

Need may be defined as the expenditure required to achieve the maximum possible health improvement that is it may be defined as the expenditure required to reduce the individual’s capacity to benefit to zero.

5. Equality of marginal met need

Steele (1981) says that the equity in health care may be defined as “equality of marginal met need”. Here need is defined as the capacity to benefit and high priority needs are those where the return to marginal additional expenditure is high. Needs are then ranked according to priority and equity is defined as an allocation of resources where marginal met need is equalized.

**Equality Of Access**

Equal access for all is also another objective of an egalitarian health policy and this is also anchored around the notion of vertical and horizontal equity. Like need, access has also been defined in several ways. Some broad notions are mentioned below.

1) Access as utilization:

Access as actual utilization of health care has been the most common notion of access. When access is treated as being synonymous with utilization, equality of access would simply mean equality of treatment or at best an equal treatment for those in equal need. Mooney (1983) has drawn attention to the fact that at times policymakers have made clear distinction between access and utilization.

2) Access as the cost incurred in receiving health care

LeGrand (1982) is of the view that access can be viewed as the money and time costs faced by individuals in seeking treatment. Mooney (1983) endorses this
interpretation of access to say that equality of access is all about equality of opportunity i.e. potential utilization irrespective of actual utilization of health care. So this is a supply side criterion in contrast to the concept of actual treatment, which depends on both supply and demand, that is not only on the cost facing the individual but also on his perception of the need for and benefit from health care. Equality in access need not result in equal utilization of health care.

3) Access as maximum attainable consumption

The above conception of access would inter alia imply that two individuals facing the same price of health care would enjoy the same access even if they had no income. This is absurd. This implication can be avoided by defining access as the maximum attainable consumption (Olsen & Rodgers 1991) of the good in question.

In the figure 1.1, individuals with budget constraints AB and CD face the same price of health care and hence according to LeGrand’s interpretation of access, have equality of access but the maximum health care that they can consume is different, so according to Olsen & Rodgers’ definition, access is unequal.
4) **Access as foregone utility**

Culyer & Wagstaff (1993) argue that the two approaches could be reconciled by reinterpreting access costs in terms of the foregone utility associated with health care consumption. So the cost of health care consumption is measured in terms of utility rather than money. So when the marginal utility of income is greater for the poor than for the rich, their access cost would be less even if they face the same money price. To equalize access, the price of health care paid by the poor needs to be reduced to the level where the product of the price of health care and the marginal utility of income becomes equal to that of the rich. But Culyer & Wagstaff (1993) are also quick to point out that even doing so will not automatically equalize the maximum attainable consumption of health care.

5) **Access to health care and consumption of non health goods – the catastrophe impact**

In order to consume a given quantity of health care, the poorer person has to make do with less of other goods. For e.g., using the figure, Culyer & Wagstaff show that to consume $M_0$ units of health care, the poorer individual has to restrict consumption of $Y$ to $Y_1$. Whereas, the better off individual is able to consume $Y_2$ units of $Y$. This is realized by Olsen and Rodgers (1991) as well as Le Grand (1991).

This conflict may however, be tempered by what has been called ‘specific egalitarianism’ by Tobin (1970).

**Equality of health, per se** rather than health care implies a concern for substantive justice rather than procedural justice. According to Deaton (2003) the notion of health equality is analogous to the idea of income equality. Inequality aversion or preference for more equality in income distribution can be coded into a social welfare function according to which mean preserving but equalizing transfers of income lead to an increase in social welfare. In the context of health, the aggregate welfare can be thought of as the product of mean health and health equality (complement of inequality).

---

4 Wagstaff (2001) has used the term catastrophe impact to say that some times the expenditure on health care turns out to be so excessive so as to push the existing income below the poverty line after incurring health expenditure leading to a catastrophic decline in the consumption of some basic goods like food.
Here one can introduce a distinction between equity and equality in health\(^5\) which can be encoded into the degree of aversion to inequality.

Suppose there are two individuals or two groups of individuals, A and B, in the society. An equal distribution of health would require a similar treatment of A and B and so in the figure below, if the health of A is measured on the horizontal axis and health of B is measured on vertical

\[\text{Figure 2.2a Equal weight for A & B} \quad \text{Figure 2.2b Greater weight for B}\]

axis and the welfare contour FF represents the frontier of the health possibility set, representing the feasible combinations of health for A and B. The health possibility set is continuous, concave to the origin and monotonically decreasing from left to right. Then an equal distribution of health between A and B would require that FF is inclined at an angle of 45 degrees to both the axes. While if some normative theory of distribution

---

\(^5\) Le Grand (1987) has shown that there is a distinction between equity and equality, the term equality is essentially descriptive while equity is essentially a normative term. Equity statements are statements of value; equality statements are statements of facts. We can observe a particular distribution, of say, medical care and decide by observation whether it is equal or unequal. However, the presence or absence of equity cannot be established by reference solely to facts. It is necessary to couple the facts with value judgements.
asserts that a greater degree of aversion to health inequality would require a greater allocation of health to B then the contour FF would be inclined at an angle greater than 45 degree to the vertical axis.

Whether A and B are to be treated alike or differently would depend on the value judgments of the distributive policy. So, any discussion of equity in health and health care inevitably has to be grounded in an ethical theory. An ethical theory serves to identify a context and reasoning to determine what ought to be, in contrast to a mere positive analysis of what is. Ethical theories that serve to determine a just or a fair distribution of economic resources (health and health care in our case), are called theories of social justice. In the context of health, the different theories of distributive justice serve to either impose a constraint on the health possibility frontier or on the welfare contours. Some fundamental theories of justice are as below:

Utilitarianism: The idea was brought to the forefront by Bentham & Mill in the eighteenth and nineteenth century and still used most widely to determine distribution of resources among different members of the society. Here the social optimum is interpreted in as those choices that provide for the greatest good for the greatest numbers. The socially optimal choices, therefore, are those that maximize the sum of utilities of all persons in the society. This principle would require greatest health for the largest number. This would require that resources should be redeployed away from people who have a low capacity to benefit from treatment (on account of some rare disease or diseases requiring prolonged treatment). The principle of utility therefore, while seemingly is an impersonal principle of welfare, actually works contrary to the principle of equity in the case health by being prejudiced against those who require special treatment on account of some acute or rare disease. So no restriction on health possibility frontier and social welfare function may be linear if no diminishing marginal utility of health.

Rawls' theory of justice: John Rawls (1971) proposed that the principle of justice is one where social choices are fair. It would be unfair if social choices were to be dominated by individuals with economic or political power. To be fair choices are to be made behind a "veil of ignorance". A rational individual acting behind a veil of ignorance would choose to arrange social and economic inequalities in such a way that they are to the greatest benefit of the least disadvantaged. This is known as the maximin principle. This
would mean equal opportunity for equal need in the case of health. So the welfare contour would be L shaped and the health possibility frontier would also be reduced by ethical constraints.

**Nozick’s entitlement theory**: According to Nozick (1974) if people have acquired what they possess in a just fashion, then they are entitled to use those endowments in whatever way they like. This is a libertarian theory of justice that argues for minimal government intervention. This principle is not applicable in the case of health. This theory does not impose any constraint on the welfare contour and the health possibility frontier.

**Sen’s theory of capability**: According to Sen’s (1980) principle of justice there should be equality on the space of capabilities. A capability set is defined as a collection of functioning n-tuples. A functioning n-tuples is a list of various states of being and doing.

**Egalitarianism theory**: According to Veatech (1982) this would require equality of welfare among all individuals. In the case of health it can be interpreted as equal health. Here the welfare contour would be L shaped.

**Part II: Meaning and measurement of health**

**Meaning of health**

Health is an abstract notion. It may mean different things to different people. It is generally defined as the absence of illness. It may also be defined as having strength and robustness and high quality of life or it may also be defined in terms of quantity i.e. length of life or it could be defined as a combination of both quantity or quality of life. It could also be defined as absence of physical and physiological disability and pain. Health has many dimensions – on the one hand it has medical dimension, on the other hand it has social, economic and spiritual dimension. Larson (1992) has put forward four stylized models of health. These are the medical model, the WHO model, the wellness model and the environmental model.

According to the medical model health is the absence of disease and disability, both mental and physical. According to the wellness model, health is defined as the ability to
overcome illness and progressing towards a higher level of functioning and integration of
body, mind and spirit. According to the WHO model, health is a state of complete
physical, mental and social well being and not merely the absence of disease. Finally,
according to the environmental model, health is the ability of the individual to adapt to
physical and social environments and to achieve a state of balance with the environment,
which is free from undue pain, discomfort or disability.

There is also a functional conceptualization of health, which looks upon health as the
ability of a person to perform his tasks of daily living (Bowling1991) or “usual
activities” (Patrick, Bush & Chen 1973). Functional status of health is defined as the
degree to which an individual is able to perform socially allocated roles free of physically
or mentally related limitations. Disability measures are more meaningful to people’s lives
than some objective clinical measures.

Health may also be looked upon as a person’s subjective well-being. Health measures in
this domain are called Quality of Life indices (QoL). This is a much broader view of
health that goes beyond the clinical outlook on health and the ability to fulfill role
performance by an individual. Health related QoL is a multidimensional concept that
encompasses the physical, emotional and social components associated with an illness or
treatment (Revicki 1989). This measure of health outcome is used for obtaining patients’
own preferences for health states. The QoL indices also enable economic evaluation of
health care programmes.

The foregoing discussion serves to underscore the fact that health has a myriad of
meanings that have been evolving overtime. This evolution consists of a shift away from
viewing health in terms of survival, through a phase of defining it in terms of freedom
from disease, onward to an emphasis on the individual’s ability to perform daily activities
and further on to an emphasis on positive themes of happiness, social and emotional well
being and quality of life. Techniques of health measurement have also been consequently
evolving in synchronization with the evolution in the meaning of health.

In this study, a comprehensive and holistic approach to health is taken which goes well
with the WHO model as a state of complete physical, mental and social well being. This

6 Usual activities are defined as a set of social expectations determined by what persons are able to
experience and perform during the course of a day for their given age, gender and social role.
also means that health care would consist of preventive, promotive and curative care. But as far as the measurement of health status is concerned, we do not have any readily available measure of health. So we need to delve into the theory of measuring health status to adopt any one of the available techniques to construct a suitable measure of health status.

**Measurement of health status**

Health is an unobserved multi attribute variable that does not have any unique or natural measure. Measures of health may be objective or subjective depending upon whether the information on health has been elicited by objective measurement or through self reporting by the respondent. Health measures may be unidimensional or multidimensional depending on whether the measure pertains to a single attribute of health or multiple attributes of health. Measures of health could also be individual measures or population measures. The former measure the health of individuals while the latter measure health of groups of people.

The measures of health can broadly be classified into the following three broad groups:

1. Anthropometric measures
2. Mortality indicators
3. Morbidity indicator

**Anthropometric measures**

These are objective individual level measures of certain aspects of physical features of people, which are found to reflect the status of health of the people. For e.g., weight for height, height for age. Alderman (2000), Fogel (1991), Behrman (1993), Waaler (1984) have found that height for age, weight for height and body mass index (BMI: weight for height squared) are all found to be effective measures not only of the nutritional status but also of the state of health of a person (both morbidity and mortality).

**Height for age:** of a person is a summary statistic of chronic malnutrition and morbidity.

**Weight for height:** of a person is a summary statistic of his current and acute nutritional status.
Weight for age: is a summary statistic of a person’s short-term deprivation.

Waterlow (1992) found that low birth weight is a significant contributor to infant mortality. Fogel (1991) found that adult height is inversely related to chronic middle age health problems. Fogel (1994) found that height, weight and BMI are predictors of mortality. Extreme values of BMI less than 20 or more than 30 have high mortality risk.

One can obtain these measures of health either by conducting biomedical tests: i.e. by measuring the concentration of nutrients in a person’s blood or urine or by observing the behavior of a person i.e. by noting person’s appetite, extent of bodily movement, extent of activity level etc. to draw inferences about a person’s health status.

Mortality Indicators

These indicators tell us about the risk of death. These are also objective indicators and are normally used to report about the health status of populations as a whole or sub groups of population. They have a dual advantage of being objectively measurable and being easily and widely available. There is a wide variety of indicators that are used to represent this aspect of health; some of these are as follows:

- Death rate
- Birth rate
- Life expectancy at birth
- Total fertility rate
- Infant mortality rate
- Perinatal mortality rate
- Neonatal mortality rate
- Under 5 mortality rate

Death rate: This is defined as the annual number of deaths per thousand population.
Birth rate: This is defined as the number of births per thousand population.
Life expectancy at birth: This can be defined as the number of years a random new born baby can expect to live at the current age specific mortality rates.
**Total fertility rate:** This is defined as the total number of children a woman would have by the end of the reproductive period if she experienced current age specific fertility rate throughout her child bearing age.

**Infant mortality rate:** This is defined as the number of live births out of every thousand that die every year.

**Perinatal mortality rate:** This is defined as the annual number of deaths per thousand children in the first year of their birth.

**Under 5 mortality rate:** This is defined as the annual number of deaths of children under 5 years of age per thousand live births.

**Morbidity Indicators**

In the West, in the earlier times, mortality indicators were significant because life expectancy was low and a large number of pre mature deaths were recognized as the most important health problems at that time. Also, the prevalence rates of morbidity in the community must have been highly correlated with mortality rates of the community. So that mortality may have been a good indicator of total spectrum of health.

Later during the twentieth century, two developments took place in the field of health. One was that mortality was reduced considerably. Secondly, morbidity aspects of health loomed larger. Thirdly, medical advances prolonged life such that chronically ill people also continued to live life. As a result while on the one hand, mortality became a less important problem. On the other hand, correlation between mortality and morbidity must have declined substantially. Therefore, morbidity indicators became more important.

Morbidity rates reflect experiential state of health. Morbidity is a subjective measure of health. It has a direct bearing on the quality of life rather than quantity of life. This also makes it more complicated to measure than mortality rates because it has several complicating dimensions like duration, intensity, severity, stigmatization etc. Unlike mortality as an indicator of health for an individual, which can only take two values – life or death, morbidity rates reflect the fact that a person’s health can take a continuum of values.
Morbidity has cultural, social, emotional and biophysical antecedents. In fact, one interesting rider to morbidity rates is that they are subject to “substitution effect” and “income effect” in the same way as the demand for a commodity (Dasgupta 1993). As income or wage rate increases, income foregone on account illness increases, this makes work more valuable, making people less prone to report sickness. This is the substitution effect. On the other hand, as income increases, it can lead to shift towards greater health concerns and to increased awareness and reporting of morbidity. This is the income effect.

Measurement of morbidity

Morbidity rates can pertain to grouped data or individual data. They can be measured through subjectively or objectively. Subjective measurement is carried out through self reporting by the respondent himself using any one of the several standard techniques for eliciting responses from the people like the standard gamble technique, time trade off method, visual analog scale etc.; and objective measures are obtained by clinical examination of the people by a physician.

Health index

Health index is a special class of health measures that are constructed to subserve special needs by combining the chosen qualitative and quantitative dimensions of health. A health status index can be viewed as a general measure of health independent of any disease or programme. It can serve as a social indicator for comparing health status of populations of different socio economic groups, of different geographic regions; it can provide criteria for evaluating efforts to improve the health status of a community or target populations and help assess the projected benefits of new programmes competing for limited health resources.

Creation of health index

Two steps are involved in developing a health index

1. Creation of a multi attribute health state, classification system
2. Mapping of the system into a single metric scale through the use of a suitable rating scale.

There are several well established health indices that exist in the literature that are being used for different purposes and that have been constructed by using different dimensions of health and different rating scales. Some significant health indices are listed below:

A. EuroQol
It is a health status measurement instrument that aims to allow comparison of health status between people of different countries with the same disease or between people with different diseases. The aim is to produce an internationally standardized, valid, reliable and feasible measurement that is not disease specific. EuroQol produces an indication of a person's actual health status (e.g. by asking a patient) or hypothetical valuation of health status (e.g. by asking people how they rate certain possible or hypothetical health states).

B. HRQoL
Health related quality of life: this seeks to capture the impact of health care interventions on the condition of the lives of the people. This is distinct from the clinically focused measures such as the extent of tumour size etc. this distinction is necessary because Jachuk et al (1982) show that doctors and patients have very different ideas about the improvement in health status. The doctors may consider a treatment to be successful, if for e.g., the BP goes down, but patients may not feel any change in health status.

C. QALY
Quality adjusted life years: It is a measure of outcome that combines changes in quality of life and change in life expectancy. Weights are assigned to each period of time, from 0 to 1, corresponding to health related quality of life during that period. Weight of 1 corresponds to optimal health and weight of 0 corresponds to a health status judged to be equivalent to death. QALYs relating a health outcome are expressed as the value given to a particular health state multiplied by the length of the time spent in that state.
D. **HYE**
Healthy Years Equivalent: reflects the number of years in good health that is equivalent to a longer lifetime spent in poor health.

E. **DALY**
Disability adjusted life years: are calculated as the loss associated with premature mortality and morbidity, in terms of years of life in full health.

*Continuous self rated health measure*

Wagstaff & Doorslaer (1994) have constructed a continuous measure of health from multiple category self rated health measure by drawing a distinction between the observed health indicator and an underlying latent health. By assuming that corresponding to the observed categorical self rated health there exists a latent continuous health and by further assuming that this latent health follows a standard lognormal distribution if it is an increasing measure of ill health. They have then rescaled the self assessed health by scoring the value of the self rated health at the mid point of the interval corresponding to the value of the standard lognormal distribution.

Further refinements have been made in this methodology to get over the unduly stringent assumption about the latent health having a standard normal distribution and categorizing every individual with a given self rated health at the mid point of the interval in the log normal distribution.

Ordered probit regressions using self rated health categories as the dependent variables has been used by Cutler & Richardson (1998) & Groot (2000). Interval regression method which uses external information on the means per category of Self rated health contained in a more generic measure of health from another survey has been used by van Doorslaer & Koolman (2000), Doorslaer & Jones (2003), Retana & Jones (2003).

These methods have been specifically devised for the analysis of inequalities in health, to overcome the unreliability of the inequality measure that occurs by an arbitrary dichotomization of a multiple category health indicator.
Choice of health indicators

The health indicators should fulfill the following requirements Hunt & McEwen (1980)

1. The methodological criteria: validity and reliability.

2. The administration of field work on a large scale: the indicator should be short, simple, easily understood, elicit unambiguous response.

3. Aim of the study:
   a. to establish whether there are differences in health between different social groups.
   b. to determine the dimensions of health in which these differences exist.
   c. to provide possible explanations for these differences.

For these aims, the health indicators have the following desiderata: which dimension of health should be measured, should the indicator represent health or illness should the indicator be capable of measuring small/recent changes in health, should the indicator be based upon health complaints, functional capabilities or illness behaviour.

Part III: Measurement of inequalities in health

Health inequalities may be defined as systemic inequalities in the incidence of morbidity or mortality or both, between classes of people in a given population or between individuals of a given cohort.

The measures of inequality in health may be classified into two categories. There may be class-based measures of inequality or individual measures of inequality (Illsley & Le Grand 1987, Gakidou, Murray & Frenk 2000, Wagstaff 2002).

When the classes are identified on the basis of socio economic status such that the relative position of individuals can be ordered hierarchically, then the inequalities in health status across these social groups are known as socio economic inequalities in health.

When the social ordering is done on the basis of income then the health inequalities are called income related inequalities in health.
When the measurement of inequalities is carried out without any reference to the income class or socio economic status of the people then it is known as the pure measure of inequality.

It follows from the above that we may have two classes of inequality measures for health as under:

1. Class based measures of inequality.
2. Individual measures of inequality.

These individual measures of inequality may be classified as:

a. Pure measures of inequality
b. Socio economic measures of inequality

One factor common to all measures of inequality is that they need to be standardized for age and gender so as to reflect systemic, non-biological, extrinsic differences in health between people or classes.

**Class based measures of inequality**

Class based measures as the name suggests is the measurement of inequalities in health among different groups of population, where the groups may be identified on the basis of income, social status, occupation, educational achievements or any other desired criterion.

Several measures of inequality have been used in the literature to analyse class-based inequalities in health. When these classes are identified on the basis of their socio economic status these inequalities are called socio economic inequalities in health. The following measures of inequality have been used in the literature:

1. **The Range measure**
   
   This is the most frequently used measure of inequality in health. It involves comparing the experiences of the top and bottom groups of the population. It has been used by Black Report (1980), Kagamimori, Tibuchi, Fox(1983), Lecrec, Lert, Goldberg(1984), Valkonen (1990), Wilkins, Adams, Brackner (1989) and Carr-Hill(1990).
There are two variants to this measure:

(i) **Absolute measure**: The absolute measure of the difference between the health experience of the top and bottom classes. Here any one indicator of health may be chosen and the difference in the value of that indicator across the extreme classes may be measured. Lahlema and Valkonen used it in 1964 for Finland.

(ii) **Ratio measure**: The difference in the health experience of the two extreme classes may be measured as the ratio of one extreme value to another. This has been used in the Black Report (1980) for analyzing health disparities in England.

There are two defects in the range measure of inequalities. First, it overlooks what is going on in the intermediate groups. The gap between top and bottom group may remain unchanged, but the extent of inequality between the intermediate groups may diminish or increase. Second, it takes no account of the sizes of different groups being compared. This can lead to misleading results when comparisons are performed over time or across countries.

2. **Regression-based index**

The drawback of the first kind in the range measure can be overcome by applying a regression analysis in which the health indicator being considered is used as the dependent variable and is regressed upon the classes which are taken as independent variable. Here the class requires to be measured on an interval scale. There may be absolute and relative variants of this measure. These measures have been used by Winkleby, Jatulis, Frank & Fortman (1992), Volkenon (1989), Kunst & Mackenbach (1994).

(i) **Regression based absolute index**

A regression analysis of the absolute indicator of health as the dependent variable is conducted. The regression coefficient represents an absolute index.

(ii) **Regression based relative index effect**

When the dependent variable is logarithmically or logistically transformed then the regression coefficient represents a relative index.
3. **Population – Attributable Risk (%)**

This is also called the Etiological Fraction. This measure is interpreted as the proportional reduction in overall morbidity and mortality rates that would occur in the hypothetical case that everyone experiences the rates of the highest socio economic group.

**Absolute version**

It is calculated as the difference between the overall rate and the rate of the highest socio economic group, expressed as the percentage of the overall rate.

**Regression based version**

This has not yet been applied but Mackenbach & Kunst (1997) suggest that if the reference rate is taken not as the rate of the highest socio economic group but the predicted rate estimated for the some high SES value. The corresponding implication is the proportional reduction that would occur in the hypothetical case that everyone experiences the predicted rate.

4. **Relative Index of Inequality**

The morbidity or mortality rate of a socio economic group is regressed upon the relative position of that group in the population. The resulting figure can be interpreted as the ratio of the morbidity and the mortality rates of those at the bottom of the social hierarchy as compared to those at the top, estimated on the basis of the systematic association between morbidity or mortality and SES of all groups. A large score of RII implies greater inequality.

5. **Slope Index of Inequality**

This is the absolute equivalent to the RII, which expresses the health inequality between the top and the bottom grouping terms of rate differences instead of rate ratios.

6. **Psuedo Gini coefficients**

These measure are classified as pseudo Gini coefficients because while like Lorenz curve they deal with cumulative proportions of populations under study but unlike Lorenz curve they deal with grouped data rather than individual data. This method has been employed by Preston, Haines & Pamuk(1981) and Lecrec,
Lert & Fabien (1990). The classes between whom the health disparity has to be studied are classified and then ranked by their health status from lowest to the highest. Then cumulative percentages of population, grouped into classes, ranked by health status, are graphed against cumulative percentages of health status. This measure orders the classes according to their health and not their social status this measure therefore, is a kind of a pure measure of inequality.

7. **Index of Dissimilarity**
Suppose there are \( j = 1, \ldots, J \) classes in the population. Then the index of dissimilarity is

\[
ID = \frac{1}{2} \sum (s_{jh} - s_{jp})
\]

Where \( s_{jh} \) is the jth group’s share of the population health and \( s_{jp} \) is the jth group’s population share. The greater is the difference between \( s_{jh} \) and \( s_{jp} \) the greater is the degree of inequality.

This is also shown to be a pure measure of inequality in contrast to a socio economic measure of inequality.

**Individual measures of inequality**

A. **Pure measures of inequality**
Health inequality measures can also be measured for individual or ungrouped data. All such measures can be grouped into two broad classes: individual mean differences and inter individual differences.

1. **Individual mean differences**
This measure compares the health of each individual to the mean health of the population. The measure is calculated as:

\[
\text{IMD}(\alpha, \beta) = \frac{\sum_{i=1}^{n} |y_i - \mu|^\alpha}{n \mu^\beta}
\]

where, \( y_i \) is the health of individual \( i \), \( \mu \) is the mean health of population, and \( n \) is the number of individuals in the population. The parameter \( \alpha \) changes the
significance attached to differences in health observed near the mean of the distribution. The parameter $\beta$ controls the extent to which the measure is purely relative to the mean or absolute.

1(i) Variance
when $\alpha = 2$ and $\beta = 0$ then the above measure is the variance of the distribution.

1(ii) Coefficient of Variation
when $\alpha = 2$ and $\beta = 1$ then the above measure gives the coefficient of variation of the distribution.

1(iii) Relative measure
When $\beta = 1$, the measure is relative.

1(iv) Absolute measure
When $\beta = 0$, the measure is absolute.

2. Inter – Individual differences
Here, each individual’s health is compared to every other individual’s health. The general form of measurement is:

$$\text{IID (}\alpha\beta) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|^\alpha}{2n^2 \mu^\beta}$$

Where, $y_i$ is the health of the health of individual $i$ and $y_j$ is the health of individual $j$, $\mu$ is the mean of the population and $n$ is the number of individuals in the population.

2(i) Gini Coefficient
When $\alpha = 1$ and $\beta = 1$ then the above measure gives the Gini coefficient. This is obtained from the Lorenz Curve, when the cumulative proportions of population (starting with the sickest person and ending with the healthiest) against cumulative proportions of health. If health is distributed equally, the
Lorenz curve coincides with the diagonal. Otherwise it lies beneath the diagonal. The area between the diagonal and the Lorenz curve is equal to the Gini coefficient and is measured as twice the area between the Lorenz curve and the diagonal. The greater is the value of the coefficient, the greater is the degree of inequality.

When \( \alpha = 2 \), the IID is equivalent to IMD.

The choice of health inequality index is based upon the three normative issues—(i) whether to measure inter individual difference or inter mean difference, (ii) what value of \( \alpha \) to choose, and (iii) what value of \( \beta \) to choose.

The value of \( \alpha \) helps to determine whether the gains or loss of health that occur at the ends of the distribution should be treated differently from the gains or loss that occur near the mean. When \( \alpha = 1 \), it implies that the evaluator is indifferent between the gains to health accruing near the mean or away from it. But a value of \( \alpha > 1 \) implies greater weight being attached to gains in health at the end of the distribution.

\( \beta \) can take values between 0 and 1. The choice of \( \beta \) will depend on whether we are concerned with absolute differences in health or relative differences in health or a mix of both with suitable weighting.

Choice between IID & IMD

Suppose two populations have the same mean value of health, and the same amount of health is transferred in both cases. The initial distribution of health is different in the two populations. Which of the two represent a greater decline in inequality? When we use IID, it means not only where the individual starts, but where the rest of the population is that matters in the quantification of inequality. Using IMD implies that what matters, is the absolute change achieved, irrespective of where people are in the distribution.
The normative choice about the intensity of transfer and about the family of measures is not completely separable. IID measures, even when $\alpha = 1$, are more sensitive to equivalent transfers farther from the mean.

**Socio economic measure of inequality**

**Concentration index**

This involves examining inequality in health across different dimensions of social and economic stratification like race, income, gender and location. The measure used for quantifying socio economic inequalities in health is called the Concentration Index. This index has been suggested by Wagstaff, Paci & Doorslaer (1991). In contrast to the Lorenz curve which has been shown to be a pure measure, Wagstaff et al have suggested that a socio economic measure may be obtained if people are ranked by their socio economic status, beginning with the most disadvantaged and this is plotted against the cumulative proportions of health. This gives the concentration curve for health. If health is distributed equally across socio economic groups, the concentration curve will coincide with the diagonal. If poor health is concentrated in the lower socio economic groups, the health concentration curve lies below the diagonal. The farther the actual curve lies below the diagonal, the greater is the degree of inequality in health.

The health concentration index is twice the area between the diagonal and the concentration curve.

The Concentration index can be computed as

$$C = \frac{2}{\mu}(\sum_{i=1}^{n} x_i R_i - 1)$$

$x_i$, for all $i=1,\ldots,n$, is the health score of the $i^{th}$ individual

$\mu = \text{mean individual health}$

$R_i = \text{relative rank of the } i^{th} \text{ person in the socio economic gradient, beginning with the worst off person.}$
Demographic Factors and unavoidable inequalities

The influence of demographic factors, namely, age and sex on health status, is biologically and genetically determined and therefore such differences in health are unavoidable. Hence a standardization of age and gender on the rates of a health measure is needed to be carried out to arrive at an inequality measure, which truly reflects only those inequalities which are avoidable, systemic and amenable to policy influences. Three kinds of standardization methods have been suggested in the literature. They are as below:

1. Direct Standardization
   In this method, grouped data is used and age – sex specific average illness rates of each socio economic group to the age and gender structure of the population (Rothman, 1986).

2. Indirect Standardization
   This method can be used on individual level data. It involves replacing person i's degree of illness by the degree of illness suffered on average by persons of the same age and gender as person i (Rothman 1986).

Part IV: Determinants of health inequity

The economic approach to the analysis of health inequalities is based upon modelling the determinants of demand for health and health care (Maynard 1983, Muurinen & LeGrand 1985). There is a wide spectrum of factors that influence health and its distribution (Evans et al 2001, Gibbons 2005). These determinants may exert their influence at individual level, household level or community level.

While it is impossible to draw an exhaustive list of the factors determine health prospects, they can be broadly grouped into five domains, which are as follows:

Biological Factors The predisposition to health or disease of an individual is influenced by his biological factors such as genetic endowments (Strohman 1992, Evans et al 1994), age (Deaton & Paxson 1986) and gender (Kawachi et al 1999, Hunt & Annadale 1999).


Behavioral choices Behavior patterns represent a prominent domain of influence over health prospects. The choices with respect to diet, physical activity and substance abuse all affect health. Dietary factors have been associated with coronary heart disease. Physical inactivity has been associated with increased risk for heart disease, colon cancer, diabetes, dementia, and osteoporosis. Similarly, use of tobacco and other intoxicants affects health and is considered to be a prominent cause of preventable illness.

Medical care There is a good deal of evidence to show that health services also affect health outcomes and its distribution. Utilization, availability, accessibility, prices and quality of health services also determine health outcomes (Lavy V et al 1996, Panis et al 1994, Rosenzweig 1982). One important dimension of accessibility is the travel time entailed in using a health care facility. This travel time may depend upon distance to the facility, transportation system, road infrastructure. Distance has a significant impact on health utilization and health status (Beneo et al 1996, Mwabu et al 1993, Thomas et al 1996). Price also influences utilization behaviour and health outcomes. A higher money price tends to reduce utilization, especially among the poor (Alderman et al 1996). Similarly, insurance tends to raise the usage of health services (Gertler et al 1997). Measures of perceived quality — e.g., availability of drugs, opening hours, training of staff

The relative contribution of medical care to life expectancy rose during the latter part of the century and is likely to continue growing as technology is better able to address the health care needs of the population (DHHS, Public Health Service 1994).

We find evidence of the influence of both medical and non medical factors on health. The triumph of scientific models of inquiry led to an eclipse of the non-medical determinants of health in the twentieth century (Evans & Stoddart 1990). But in the recent times, because health services research has failed to find consistent evidence of a systematic connection between provision of health care on the one hand and patients' needs and health outcomes on the other hand (Feeny D, et al 1986, Anderson TF & Mooney G 1988). Undue reliance on health care has also precipitated a "health care cost crisis" because no matter how much resources are spent on health care, there is always some unmet need and this is creating a tension between ever increasing needs and increasing resource constraints, [Schieber & Poullier 1989]. This has led to a resurgence of interest in the non medical determinants of health.

Interest in the socioeconomic and environmental determinants of population health emerged in the 19th century with the work of public health pioneers such as Virchow, Villermé, Chadwick and Shattuck. It took until the 20th century however, for researchers to establish a body of knowledge on the broader determinants of health (Amick et al., 1995). Many studies have documented the importance of socioeconomic determinants of health.

McKeown and others looked at early death rates in England & Wales from 1841 to the middle of the twentieth century for each of the major infectious diseases. This has been the time when effective modern medical prevention and treatment became available. It was observed that most of the decline in death rates for almost all the major infectious diseases occurred well before the discovery and general availability of modern medical techniques such as antibiotics, immunization, intravenous rehydration and vitamin supplements. His findings in nutshell could be represented as:
modern medicine could account for only a very small part of the increase in English life expectancy to 70 years by the middle of the twentieth century.

(i) clean water supplies and sewage systems played an important role in the final decades of the twentieth century in reducing exposure to air, water and food borne diseases such as diarrhea, typhoid, fever etc. (McKeown 1976).

One can see the realization about the importance of non medical determinants of health in the American literature also following the early work of Fuchs (1966). In the American studies health is viewed as an outcome of a production process involving health inputs such as behavioural variables, environmental factors, education etc. (Wagstaff 1989).

The recent historical anthropometric analysis for a number of European countries has shown strong correlation between nutrition and decline in mortality (Fogel 1992).

When we turn to South Asian history, most demographers have argued for a case for modern medical science in ameliorating the poor health and low life expectancies of people. However, Zubrigg (2001) has shown that while tracing the history of decline in malaria mortality in Punjab (India) since 1868 till the post independence period nutrition played an important role.

Dreze & Sen (1989) have shown that British life expectancy increased during the decades between the two World Wars despite (i) increase in urban crowding, (ii) decline in medical services and (iii) occurrence of the influenza epidemic in 1918. They say that this happened because it was during this time that access to basic food was ensured through many forms of government support including public employment, food price control and food rationing.

Income and health

The relationship between income and health has been a widely debated topic in the research on health equity. Evidence on the nature of this relationship is critical to formulation of public health policy. There are different dimensions of the relationship between income and health. While there is no denying that there is a positive association between income and health; there are many conflicting opinions about the causal
relationships and pathways between income and health (& its gradient). It can be said that there is a dual relationship between income and health (Smith 1999). While, on the one hand there is evidence to show that income is a determinant of health. On the other hand, there is strong evidence that shows that it is the health status of the people, which determines their income through their working capacity and productivity. It could also be the case there might be endogeneity of income in health determination that is, both health and income are determined simultaneously (Ettner S, 1996).

In the case, that income has a causal effect on health; there are two hypotheses about pathways through which income affects health. One is the “absolute income hypothesis” and the other is the “relative income hypothesis”. The absolute income hypothesis states that an individual’s health depends on his own level of income, irrespective of the income of everybody else in the society. (Preston SH. 1975, Pritchett L, Summers LH. 1996, Adler NE, Boyce W, Chesney MH, Folkman S, Fymes L. 1993, Gravelle 1998). According to this hypothesis, the pathway to health from income is through absolute material deprivation, which affects health directly through lifestyle, access to nutrition, health care etc. According to the “Relative income hypothesis”, it is the distribution of income in a society that affects the individual’s health (Wilkinson RG. 1996, Ben-Shlomo Y, White IR, Marmot M. 1996, Kaplan GA, Parnuk ER, Lynch JW, Cohen RD, Balfour JG. 1996, Kennedy BP, Kawachi I, Prothrow-Smith D. 1996). In other words, it means that it is not only an individual’s own level of income but his relative position in the income distribution that affects his income. Here one of the pathways to health is argued to be through psychosocial stress that leads to erosion of social esteem triggered by relative deprivation of material factors (Wilkinson 1996, 2000). The other pathway has been through the pattern of social investment (for example, lower state effort on education and welfare spending) that often accompanies a growing distance between the rich and the poor (Kawachi & Kennedy 1997). Erosion of social capital (defined as the resources available to individuals and society through social relationships) has been cited as another mechanism that affects health through relative income (Kennedy, Kawachi & Lochner 1997).

Gravelle, H (1998) is of the view that the relative deprivation theory is a statistical artefact that has arisen on account of use of aggregate rather than individual data.

Adams P et al (2003) have conducted a study using panel data and they do not find any causal link from wealth to health.
Diez-Roux et al (2000) have eloquently stated that, “The analytical separation of these two mechanisms (i.e., the effects of absolute and relative differences) may be theoretically interesting but is also artificial, because both are inextricably linked.

Sometimes the influence of income on health may be felt only after a lag. There is a growing recognition of the importance of examining people’s current health in terms of their life-course experience (Kuh & Ben-Shlomo 1997). It is possible that income measured at one point of time may be a poor marker for an individual’s access to material resources across their lifetime (Blundell & Preston 1995). Investigations have been therefore carried out analyzing the relationship between long-term income and health (Benzeval & Judge 2001). Such studies help to distinguish between the transient and permanent impact of poverty on health.

Distribution of non medical determinants of health and their impact on the distribution of health

It has been found by Kunst (1997) that the higher levels of alcohol consumption amongst the lower socio economic groups in northern Europe and US. In India, the NSS data also shows a greater tendency for smoking among poorer and uneducated people using NSS data 52nd round.

Distribution of medical determinants of health and their impact on the distribution of health

It could be possible that health care facilities may be distributed unequally in the society. How much of this inequality would be responsible for the inequality in health would also depend upon its impact on health.

According to Wagstaff (Poverty and Health 2001) relatively little work has been undertaken to assess the relative contribution of the inequalities in the proximate and underlying determinants of health to the inequalities in health itself. North et al (1993) sought to explain the strong inverse relation between grade of employment and sickness absence. Several risk factors were identified including health related behaviour, work characteristics, low levels of job satisfaction and adverse social circumstances outside
work. One third grade differences in sickness absence was accounted for by inequalities in these factors. Marmot (1978) carried a similar exercise for coronary heart disease and found that smoking, lack of exercise and high blood pressure are more common among the lower civil service grades. They found that inequalities in these risk factors account for about 40% inequalities in coronary heart disease.

Analysis of health inequities

The modern literature on the analysis of health inequalities can be traced to the publication of the Black Report in UK, in 1980. The Report brought to the forefront the fact that health inequalities among people bear a systematic association with their position in the socio economic hierarchy of the society. The Report also served to highlight the multidimensional determinants of health inequality. Ever since, most analyses of inequalities in health has been carried out on the premise that just as health is determined by a multitude of factors, similarly its distribution may be governed by a myriad elements and the identity and significance of these elements may vary in different contexts.

The Whitehall study (Marmot, Shipley & Rose 1978) compared the inequalities in the mortality rates of the civil servants according to their employment grade and found a steep inverse relation between mortality rates and their employment grade.

The Black Report (1980), Black et al 1992, Marmot & Wilkinson 1999 also found health inequalities associated with social classes primarily identified by occupation in UK at different time periods and over time.

The drawback of these studies however is that they use range measure to capture health inequality in the society. This measure of inequality is a class based measure of inequality consequently variations in health are known only across groups and not among individuals within groups. Secondly, the range measure captures only the average value of the health indicator for the two extreme groups, as a result (i) any changes taking place in the intermediate groups are ignored and, (ii) the relative sizes of the different classes might be different.
To overcome these defects, LeGrand & Rabin (1986) & Illsley & LeGrand (1987) have used the Gini coefficient to compute the inequality in age at death for all causes of death and age at death for specified diseases. LeGrand (1987) has also used this methodology for undertaking international comparisons of inequality in health.

The limitation of these studies is that they have used mortality indicators as measures of health so nothing is known about the distribution of morbidity. Secondly, it is not possible to identify and quantify the underlying determinants of health inequality from these studies.

**Regression based decompositions of concentration index**

The above shortcoming has been overcome by using the concentration index for measuring health inequality and using a regression based method for decomposition of health inequality.

After the development of concentration index by Wagstaff, Paci & Joshi (1989) that became a popular measure of income related inequalities in health, Wagstaff, Paci & Joshi (2001) have developed a regression based decomposition of concentration index of health to analyze the socio economic causes of inequality. This method has the several interesting features namely, (i) it analyzes inequalities in self assessed health as opposed to mortality; (ii) it looks at the distribution of health among individuals rather than groups of people; (iii) it uses a summary statistic to capture the socio economic inequalities in the health of the people; (iv) it can decompose this measure of inequality into its constituent contributions by different factors.

This methodology has also been used to decompose the inequalities in child survival in Cebu (Wagstaff 2000) and malnutrition inequalities among children in Vietnam (Wagstaff, , van Doorslaer, Watanabe, 2003).

Wagstaff, Doorslaer & Watanbe (2003) have decomposed the Concentration index of health within a regression framework, following Rao’s (1969) theorem in the income inequality literature and making use of the formula of elasticity of the Gini coefficient.
with respect to specific income sources to disaggregate the Gini coefficient by factor components (Podder 1993).

If \( G \) is the Gini coefficient of total income and \( C_k \) is the concentration ratio of the \( k \)th factor and \( E(Y) = \mu \) and \( E(Y^k) = \mu_k \), then, Rao in 1967 derived that

\[
G = \sum_{k=1}^{K} C_k \left( \frac{\mu_k}{\mu} \right) \tag{1}
\]

Wagstaff et al have shown that if we have a linear regression model linking health to a set of \( k \) determinants in the following manner

\[
y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i \tag{2}
\]

where \( \beta_k \) are coefficients and \( \varepsilon_i \) is the error term. We assume that everyone in the selected sample faces the same coefficient vector \( \beta_k \). So interpersonal variations in health are assumed to be derived from systematic variations across income groups in the determinants of health i.e., \( x_k \). Then using the formula in equation (1) for elasticity of Gini coefficient and because mean of fractional rank, \( R_i \) of individuals = \( \frac{1}{2} \) and 
\[
\mu = \alpha + \sum_k \beta_k x_k^- \tag{3}
\]

We can write the concentration index for health \( C \), as

\[
C = \sum_k \left( \frac{\beta_k x_k^-}{\mu} \right) C_k + GC_e \left/ \mu \right.
\]

where, \( \mu \) is the mean of health

\( x_k^- \) is the mean of \( x_k \)

\( C \) is defined as \( 2/n \sum_{i=1}^{n} y_i R_i - 1 \)

\( C_k \) is the concentration index of \( x_k \)

\( GC_e = 2/n \sum_{i=1}^{n} \varepsilon_i R_i \) is the generalized Lorenz curve (Shorrocks 1983).

RHS in equation (3) is made up of two terms. The first term is interpreted as the deterministic component of the concentration index of health, which is equal to the weighted sum of the concentration indices of \( k \) regressors. The second term is a residual component, which reflects the inequality in health that cannot be explained by systematic variations across income groups in \( x_k \).
The model has also been extended to incorporate individual heterogeneity in the response of health to the explanatory variables by using quantile regression (Jones & Nicolas 2002) and by exploiting the information contained in the panel data to specify heterogenous parameters in the health regression model such that all the parameters become individual specific in the model (Jones & Nicolas 2004).

A fundamental limiting factor in the foregoing analysis is that these methodologies exclusively deal with the decomposition of the concentration index of health. This means that they are actually looking only at that subset of the total inequalities in health, which is correlated with income or some socioeconomic factors, rather than looking at the overall inequalities in health.

Pradhan, Sahn & Younger (2003) have computed the overall inequality in health. Wherein, the height of children is used as the indicator of health and the inequality in health is computed by the Gini coefficient. They have used this inequality measure for international comparisons of health and for decomposing this inequality between countries and within countries.

Sahn (2003) has also computed overall inequality in health using the height of pre school children for India and which has been used for interstate and intertemporal decomposition of inequality.

The analysis done by Sahn et al (2003) & Sahn (2003) are constrained by the fact that the decomposition of inequality can only be carried out to decompose inequality within groups and between groups. Determination of the causal factors of inequality is not possible from this kind of analysis. To determine the causal factors of health inequalities, we can use another regression based method for inequality decomposition. This method is discussed in the next section. Before we move on to a discussion of that method, It is necessary to mention that there exist two other approaches to analysis of health inequalities namely, multilevel modeling and health production function approach.
**Multilevel Modeling**

Multilevel modeling *approach* provides a useful way of analyzing health inequalities. This approach anticipates that determinants of health inequalities occur simultaneously at several levels, from the individual, to neighborhoods, regions, and states. Consequently, multilevel regression techniques are essentially about modeling heterogeneity at each of the desired levels of the conceptual model through a range of variables that tell us something about each of the levels. These methodological and substantive perspectives (Leyland AH, Goldstein H 2001) are supported by a robust technical estimation process. Multilevel methods consider most data structures within a nested framework and such nesting could be hierarchic and/or non-hierarchic (Pickett KE, Pearl M 2001). In fact, seen this way, repeated/ longitudinal analysis (whether it is people who are repeatedly measured or places), multivariate analysis (when there are more than one inter-related outcomes) or a cross classified analysis (when we do not have neat hierarchic nesting) are simply special cases of a multilevel regression framework.

**Production Function Approach**

Contoyannis & Forster 1999 & 1991 have used health production functions to analyse the distribution of health in the society. Their analysis rests on the assumption that the rich and the poor groups in the society may have different health behaviors, (what they call the healthy and unhealthy behaviour). So while it is the same set of health inputs, which affects the health status of the two groups, their health production functions are different. The differential shape of the health production functions is used to explain the given distribution of health status in the population.

For given population of $i = 1, \ldots, n$ individuals, an individual’s health ($H \geq 0$) is a function of individual’s income ($I \geq 0$) conditional upon his health related behavior. For the sake of exposition, it is assumed that an individual’s health related behaviour can be classified as either healthy($h$) or unhealthy($u$). The probability ($p$) of behaving healthily is an increasing function of the individual’s income, $l$; So, $p = p(l), \ 0 \leq p \leq 1, \ \delta p/\delta l \geq 0$. 

52
It is assumed that the health production function $H = H(I)$ is linear for both ‘$h$’ and ‘$u$’ and, $H_h(I) \geq H_u(I) \quad \forall I$, where $H_h(I)$ and $H_u(I)$ denote the levels of health attained by behaving healthily and unhealthily respectively.

The expected health of an individual with a given income level $I$ is

$$E(H/I) = p(I)H(I) + (1-p(I))H(I).$$

**Distribution of population health**

To obtain the average health of the population, the weighted average of the expected health function for income level $I$, over the $n$ individuals that comprise the population is taken.

$$E(H) = E_i [E(H/I_i, Z)]$$

The expected population health depends upon the characteristics of the individual health production functions and their distribution within the population.

From the expected population health function we can also derive the extent of health inequality in the population using any of the inequality measures that exits in the literature. Variations in the inequality in community health in its average value can also be derived in response to alternative health promotion policies such as improvement in education levels. Greater investment in the vector of public health services, improved nutrition level, greater equality in access to health care and greater equity in the burden of healthcare expenditure. This can be done by computing the elasticity of individual’s expected health with respect to income at each level of $I$.

**Regression based decomposition of inequality**

However, as has been discussed earlier, the concentration index of health captures only the income (or socio economic) related inequalities in health, which is only a part of the total inequalities in health. If we wish to quantify total inequalities in health and carry out the decomposition analysis of the inequalities in health in similar manner then Doorslaer
and Jones (2003) have shown that Gini coefficient of health, \( G \) (Le Grand 1989), can be defined analogous to the health concentration index in equation 7.

\[
G = 2/ \mu \sum_{i=1}^{n} y_i R'_i - 1
\]  

(5)

Where, all other variables are the same as before but \( R'_i \) now denotes the relative fractional rank in the health distribution with individuals being ranked from lowest to highest health. Decomposition of \( G \) in equation 8 can either be carried out exactly as the decomposition of \( C \) in equation 6 or, we can make use of the relationship between Gini coefficient and concentration index (Kakwani 1980)

\[
\rho(y, R'_i) \quad G = \frac{\rho(y, R'_i)}{\rho(y, R_i)}
\]

(6)

where, \( \rho(.) \) is the correlation coefficient.

But this approach has the following limitations:

1. It is applicable only to Gini coefficient as a measure of inequality,
2. it ignores the contribution of the constant term to inequality,
3. it measures the contribution of factor \( k \) to inequality at the mean value of the \( k^{th} \) income source,
4. it is applicable only to linear regression model.

**Limitations of the conventional decomposition method**

Traditional approaches to inequality decomposition are mostly descriptive rather than prescriptive. They involve measuring the extent of inequality and speculating on its determinants. Inequality decomposition by population subgroups is an unconditional approach to decomposing inequality based on the natural decomposition rule suggested by Shorrocks (1982, 1999) wherein, quantification of the impact of different factors like education, age etc. on inequality is based upon dividing the sample into discrete categories and then calculating the level of inequality within each sub sample and between the means of the sub samples. This provides rather limited information on the fundamental determinants of inequality. This also means that even though some factors
like age etc. are essentially continuous in nature they have to be divided into discrete categories. In such methods, handling of multiple factors also becomes unwieldy. One very significant limitation of the decomposition by population sub groups is that it fails to take into account the endogeneity of the variable under question. This lack of control for endogeneity limits the decomposition to being a purely descriptive analysis (Murdoch & Sicular 1998).

The work of Shorrocks (1980 and 1982) on income inequality decomposition by income sources requires an identity to express income as the sum of several components. This does not help in identifying the sources of income. All that this approach does is to quantify the role of different sources of income in generating inequality.

A straight jacket transplantation of this decomposition technique in the case of health is obviously ruled out because it is not possible to express the health of an individual as a sum total of health from different sources.

We therefore use regression based methods of decomposition to decompose health inequality and analyse the determinants of health inequality. The regression framework provides a method for assigning explanatory power to the several independent variables in explaining the inequality in the dependent variable by holding other things equal and by decomposing the overall inequality in the sense that the contributions of the several independent variables sum to the contribution of the overall model.

**Regression Based Method of Decomposition**

Since the early 1970s, economists have developed a regression-based approach to inequality decomposition. Unlike its traditional counterparts, this approach enables identification as well as quantification of the determinants of inequality. It also allows the inclusion of any number or type of variables even including social, economic, demographic and policy factors (Zhou & Wan 2003). This method provides an efficient and flexible way to quantify the conditional role of different variables in a multivariate context (Heshmati 2004) and it also allows for individual heterogeneity in responses (Jones & Lopez 2004). It also enables to take into account the endogeneity of the
dependent variable and random errors while identifying and quantifying the determinants of inequality.

Oaxaca (1973) and Blinder (1973) are the pioneers of the regression based approach. They focused on the determinants of the difference in mean wage earnings between males and females. Juhn, Murphy and Pierce (1993) extended this approach so that the decomposition depends on the difference in the entire income distribution between two groups rather than on the difference in mean income only. Bourguignon, Fournier and Gurgand (2001) relaxed the requirement of a linear income-generating function imposed by Juhn, Murphy and Pierce (1993). The common theme running through the above analysis is the focus on explaining differences in income distribution between distinct groups of income recipients. No attempt has been made to quantify the contributions of specific factors to total inequality in the above studies.

In a different strand of literature using semiparametric and nonparametric techniques, DiNardo, Fortin and Lemieux (1996) and Deaton (1997) have described and compared the entire distribution of the target variable in terms of the density function, rather than attempting to decompose a summary measure of inequality. Although they impose few structural assumptions, the findings are less conclusive and so the use of their techniques is not very widespread.

**Recent Advances in the Regression Based Approach**

The analyses by Oaxaca and Blinder have been carried forward within the regression framework by Fields (2000) and Murdoch & Sicular (2002) to decompose inequality in the full distribution of income. Wan (2002) has carried forward the regression-based approach to decomposition to present the entire analysis in a more integrated framework using Shapley value approach (Shorrocks 1999).

Fields and Yoo (2000) and Morduch and Sicular (2002) have employed conventional techniques to specify and estimate parametric income-generating functions and derive inequality decompositions in the full distribution of income rather than decomposing the gap in the mean value of income between different groups as in Oaxaca (1973).
Fields (2000) paper shows how regression models can be supplemented by decomposition analyses to learn the relative importance of different explanatory factors in the distribution of income. He devised a regression-based procedure for assigning weights to various independent variables for explaining the variation in a dependent variable. The variation in the dependent variable is decomposed into as many components as there are factors. He shows how the explained portion of the regression (R-squared) can be decomposed into weights for each of the regressors. Moreover, if we accept certain decomposition rules, then the weights given hold for a broad class of dispersion measures. The methods devised in that literature are equally applicable to studies of health inequality. The essential results are the following.

**Decomposing the Dependent Variable into Contributions Made by Various Explanatory Factors**

**The Method**

Consider a standard regression equation of the form

$$Y = \beta^0 + \sum_{k=1}^{K} x_k \beta^k + \varepsilon$$

(7)

When the equation is estimated in data, the resultant parameter estimates are

$$(\hat{\beta}^0, \hat{\beta}^1, \ldots, \hat{\beta}^K)$$

(8)

and the calculated residuals are

$$\hat{\varepsilon}^i = y^i - \hat{\beta}^0 - \sum_{k=1}^{K} x_k \hat{\beta}^k$$

(9)

The decomposition analysis is intended to supplement regression analysis by using the estimates in (8) and (9) to assign explanatory power to the several independent variables in a way that a) holds other things equal, b) decomposes in the sense that the contributions of the several independent variables sum to the contribution of the overall model, and c) allows for variation in the dependent variable to be gauged by the variance and by other indexes as well.

Criteria a) and b) can be fulfilled in the following manner. Given the regression equation (7) and estimates (8) and (9), let $s(X^k)$ denote the share of the variance of $Y$ that is attributable to the k'th explanatory factor and let $R^2$ be the fraction of the variance that is
explained by all of the X’s taken together. Then, Fields (2003) shows that the variance of Y can be decomposed as
\[
\text{Var}(Y) = \sum_{k=1}^{k} \text{Cov} [X^k \hat{\beta}^k, Y] + \text{Cov} [\hat{\varepsilon}, Y] \tag{10.a}
\]
or, upon dividing through by \(\text{var}(Y)\),
\[
\text{Var}^2(Y) = \sum_{k=1}^{k} \text{Cov} [X^k \hat{\beta}^k, Y] + \text{Cov} [\hat{\varepsilon}, Y] \tag{10.b}
\]
where each “s-weight” \(s(X)\) is given by
\[
s^{k}(X) = \sum_{i=1}^{2} \text{Cov} [X^k \hat{\beta}^k, Y] \tag{10.c}
\]
and the weight associated with the residual is given by
\[
s^{k}(\text{residual}) = \text{Cov} [\hat{\varepsilon}, Y] \tag{10.d}
\]
It may be noted that when the last term in (4.b) is excluded, the remaining s-weights sum exactly to \(R^2\). Finally, expressing the \(s(X)\)’s in terms of their percentage contribution to \(R^2\), we obtain the “p-weights”
\[
p^{k}(X) = \frac{s^{k}(X)}{R^2} \tag{10.e}
\]
such that the \(p(X)\)’s sum to 100%. The results given in (10.a)-(10.e) provide a full and exact decomposition of the variance.

The weights derived take the variance as the measure of dispersion of Y. In some contexts, it may be desirable to allow the dispersion of Y to be measured by something other than the variance. Fields (2000) applies a result developed by Shorrocks (1982) for decomposition of income inequality by sources of income to show that under six specified decomposition conditions (described in the appendix), the s-weights and p-weights given in (10.a)-(10.e) hold for any measure of dispersion that is continuous, symmetric and takes the value zero when all Y’s are identical. These three properties are satisfied by virtually all dispersion measures including the Gini coefficient, Theil index, Atkinson index, coefficient of variation, income share of the richest x%, income share of the poorest y%, and the M%/N% ratio (e.g., 90/10).

This is an extremely powerful result. It says that as long as we agree on the conditions for carrying out the decomposition, we get the same s-weights and p-weights. The decomposition here is “exact” in the sense that the variance of Y is decomposed exactly into the sum of components attributable to each regressor and the residual. By contrast, in standard analysis of variance, the sum of squares explained by the model is not the exact sum of the sums of squares attributable to each component.
Limitation of Fields methodology

The method developed by Fields is useful in non-linear regressions. But, it has the disadvantage that the variance of logarithm of income violates the transfer axiom (Morduch and Sicular 2002). Another major limitation of his work is that only the Coefficient of Variation can be used as a measure of inequality. FY's method can be shown to be equivalent to decomposing the R-squared statistic ($R^2$) from the standard regression model. Consequently, negative or equalizing contributions are ruled out by the method (Wan 2001). Apart from violating the transfer axiom Field's technique requires inequality to be measured using the log of income. Although this scaling does not affect income rankings, it may lead to distortions in the decomposition results.

Morduch and Sicular (2002), have also developed a regression based specification for decomposition of inequality. This technique unlike Fields methodology does not require the regression model to be formulated in terms of the logarithm of income and the inequality measure need not be restricted to variance of logarithm of income.

According to their specification, if we have a model of income determination as
\[ Y = \beta X + \varepsilon \]  
Then the predictions of the per capita income may be obtained as
\[ \hat{Y} = X\hat{\beta} \]  
The predictions in equation 12 may be obtained by using any flexible modeling such as weighted least squares regression, quantile regressions, dividing the sample into sub samples or correcting for endogeneity. The predictions of the regression model produce estimates of the income flows that may be attributed to different explanatory variables.

This enables decomposition of income inequality by factor sources. This means that inequality in income is apportioned to different components of income where the total income is equal to the some of these components, i.e.
\[ Y_i = \sum_{k=1}^{K} Y_{ik} \]  
When income is replaced by predicted income, the analogues become $x^m = x^m$. The income contributed by each of the explanatory factors is given by the regression results. By construction, total income is the sum of these flows plus the regression residual.
\[ Y_i = \sum_{m=1}^{M} \hat{Y}_i^m, \text{ for all } i \]
\[ \text{Where, } \hat{Y}_i^m = \hat{\beta}^m X^m, \text{ for } m = 1, \ldots, M \quad (14) \]
\[ \hat{Y}_i^m = \varepsilon_i, \text{ for } m = M+1 \quad (15) \]

These income flows can then be used directly to calculate decomposition components for all regression variables. The shares take the form:
\[ s(X^m, Y) = \hat{\beta}^m (\sum_{i=1}^{n} a_i(Y) X^m) Y(Y) \text{ for } m = 1, \ldots, M \quad (16) \]

Similarly, we may formulate a model for determination of health and predictions of the model may be used to estimate health flows attributable to different explanatory variables.

**Limitations of the regression based decomposition as proposed by Murdoch & Sicular as pointed out by Wan (2002)**

Wan (2002) however, has shown that procedure of Murdoch & Sicular suffers from the following pitfalls:

1. Only additively decomposable measures of inequality can be used in Murdoch & Sicular. They have shown that in their framework, Gini violates the property of uniform additions and only Theil's first measure is satisfactory for decomposition.

2. It ignores the contribution of constant to inequality. Though a constant as a source of income is not frequent in empirical analysis of income distribution, the presence of a constant is almost a rule rather than exception in a regression equation. Such a source of income in reduces total inequality if it is positive and vice versa.

3. Similarly though the existence of residual in the regression model has been taken into account, its contribution to inequality has not been properly accounted for. Although the error term is a white noise by definition, so it does not affect the mean of the dependent variable nor does it affect the shape of the empirical Lorenz curve, but its presence or absence does result in different income density function and thus affects income distribution or measured inequality. So determining the contribution of residual term is important.
In both Murdoch & Sicular and Fields, the contributions of the constant term and the residual term are not derived from the natural rule of decomposition of Shorrocks (1999) or equivalently the before-after approach of Cancian and Reed (1998). Consequently, the residual term and the constant term, which is a uniform addition to or deduction from income from all recipients, may or may not contribute to total inequality depending on the particular inequality measure used. Podder and Chatterjee (2002) show that a positive (negative) constant will lower (raise) inequality if relative inequality measures such as CV$^2$ or the Gini index are used.

To avoid these problems and restrictions, Wan (2002) has developed a general regression-based procedure for decomposing any inequality measure, including the Gini coefficient and Theil’s measure. This approach is consistent with the natural rule of decomposition. It is also flexible enough to be applied to any type of income transformation, e.g., the logarithm and Box-Cox transformations.

**Decomposition Procedure Developed By Wan**

The contribution of the residual term to inequality

$$Y = \hat{Y} + \varepsilon$$  \hspace{1cm} (17)

Where, \( \hat{Y} = a + Y^*(X) \) \hspace{1cm} (18)

Now, suppose we use Gini coefficient as a measure of inequality then, \( G(Y) \neq G(\hat{Y}) \) unless all \( \varepsilon = 0 \). So the presence or absence of the residual term alters the measured inequality in any regression framework.

Although, the residual terms are not explainable by the structural income generating function still it is not advisable to discard them altogether. This is because the residual term to some extent is sometimes viewed as representing factors other than those included in the regression model. Its contribution can inform policymakers as to how much included factors can explain overall inequality. So ignoring the residual term means throwing away useful information on non-included determinants of income or its distribution.
Interpretation of the contribution of residual term

From Shorrocks (1999, equation 2.4) by removing \( e \) from (17) we obtain

\[
I(Y|e = 0) = I(\hat{Y})
\]

where \( I \) stands for an inequality measure.

The contribution of \( e \) to \( I(Y) \) is then given by \( COe \):

\[
COe = I(Y) - I(\hat{Y})
\]

The difference between \( I(Y) \) and \( I(\hat{Y}) \) is subtle and important. For example, if the Gini index is used, \( I(Y) = G(Y) = C(Y) \) must be calculated with \( Y \) as the ranking variable, while \( I(\hat{Y}) = G(\hat{Y}) \) must be calculated with \( \hat{Y} \) as the ranking variable. This is the case despite the fact that the expected values of \( Y \) and \( \hat{Y} \) are identical. The rankings by \( Y \) and \( \hat{Y} \) would be equivalent if and only if there is a good enough fit of the income-generating function. Viewing from this perspective, decomposition (19) makes intuitive as well as theoretical sense. \( COe \to 0 \) as \( e \to 0 \) from all directions.

Contribution of the constant term to income inequality

If a constant is present in the income generating function, we can write

\[
\hat{Y} = \alpha + Y^*
\]

If we use the Gini coefficient as an example measure of inequality and apply it to both sides of (21) we obtain

\[
G(\hat{Y}) = \alpha/E(\hat{Y}) \cdot C(\alpha) + E(Y^*)/E(\hat{Y}) \cdot C(Y^*) = 0 + COy^*
\]

If the contributions to inequality are allocated according to (22) then the constant term plays no role at all. This problem always occurs in the framework of FY no matter what measures of inequality are used. It also occurs in MS, if the Gini index or CV/variance is used as measures of inequality. But this is not a correct way to obtain the contribution of the constant term to total inequality.

Interpretation of the contribution of the constant term

The addition of a positive (or negative) constant causes a reduction (increase) in the contribution of included factors, denoted by \( COy^* \) in (22), as \( E(\hat{Y}) \) becomes smaller (larger) than otherwise. So the impact of \( \alpha \) did not disappear; it is being distributed over or absorbed by other terms in \( \hat{Y} \). If one can ‘squeeze out’ the impact entangled in the other source(s), it is natural to attribute the impact to the constant term. Applying the most natural rule of Shorrocks again, we have
Thus, the contribution of the constant term \( CO \alpha \) is

\[
CO \alpha = I(\hat{Y} | \alpha = 0) - I(Y^*). \tag{24}
\]

\( CO \alpha < 0 \) if \( \alpha > 0 \), and vice versa.

So, \( I(Y) \) can be decomposed into \( COe, CO \alpha \) and \( COY^* \).

\( COY^* \) represents contributions by various non-constant \( Xs \).

The percentage contributions are

\[
P Ce = \frac{I(Y) - I(\hat{Y} | \alpha = 0)}{I(Y)} \times 100 \tag{25}
\]

\[
P C \alpha = \frac{I(Y) - I(Y^*)}{I(Y)} \times 100 \tag{26}
\]

\[
P CY^* = \frac{I(Y^*)}{I(Y)} \times 100 \tag{27}
\]

The decompositions given by (25–27) always add to 100 percent.

Further decomposition of \( I(Y^*) \) into contributions of individual determinants \( Xs \) can be undertaken using conventional methods or the Shapley value approach of Shorrocks (1999).

Wan points out that although these results are obtained by using the Gini index as an inequality measure and a linear income-generating function, this approach can be applied to any measure of inequality and is not limited to linear regression models. However, the interpretation of the decomposition results may not be straightforward. In general, a contribution is interpreted to measure the effect on total inequality if a particular variable or income source is eliminated. By elimination, it could mean removal or assuming an even distribution of the relevant variable among all income recipients. If inequality increases (decreases), this source or variable makes a negative (positive) contribution and it is inequality reducing (inequality-increasing).

Using a linear income-generating function and taking the Gini index, the sign of \( CO \alpha \) depends on \( E(Y_i) \) and \( C(Y_i) \). Since \( \hat{Y} \) is generated from all income determinants, it should be positive. However, \( E(Y_i) = \beta_i \ E(X_i) \) and \( C(Y_i) = C(X_i) \). Since \( E(X_i) > 0 \) for most economic variables, the sign of \( CO \alpha \) depends on \( \beta_i \) and \( C(X_i) \). From (2), \( \beta_i \) is the marginal impact of \( X_i \) on income not income inequality; however, \( C(X_i) \) takes values between \(-1\) and \( 1 \), representing the association between total income \( Y \) and the component \( Y_i \) or the variable \( X_i \). Put differently, \( \beta_i \) is the partial regression coefficient that measures the impact
of $X_i$ on $Y$ controlling for all other factors, whereas $C(X_i)$ captures the correlation between $Y$ and $Y_i$ or $X_i$ without controlling for any factors. For example, a variable such as education level is expected to produce a positive impact on income so that $\beta_i > 0$. However, highly educated people are likely to be in higher income groups so that the correlation between total income and income flows from education should be positive, i.e., $C(X_i) > 0$. As a consequence, the contribution of education to total inequality is positive and thus inequality increasing. This is equivalent to saying that overall inequality would drop if everyone were to receive the same income flow from education regardless of the education level attained.

Alternatively, a welfare or poverty-relief payment goes to the poor so that it is negatively correlated with total income, i.e., $C(X_i) < 0$. Conversely, holding all other variables constant, an increase in the welfare or relief payment to even one recipient will increase the mean of the dependent variable $Y$. Hence, the marginal impact of this variable on income is positive, i.e., $\beta_i > 0$. Consequently, the overall impact is negative and thus inequality-reducing. As another example, the variable of dependency ratio is expected to have a negative marginal impact on household income, i.e., $\beta_i < 0$. However, this variable is likely to be inversely correlated with total income so that $C(X_i) < 0$. Therefore, the dependency ratio is expected to be inequality increasing. In other words, if every household had the same dependency ratio, total inequality would be lower.

The above analysis has the following properties. It is
1. independent of inequality measures
2. independent of functional specifications of $F(X)$.
3. any arbitrary transformation of the target variable is possible and subsequently inequality may be measured over the transformed values. Inequality may still be measured over the original target variable by solving the estimated equation for the original variable.

So the framework that we have adopted for analyzing the determinants of inequality is as follows:

We know from the computation of the income related concentration index of health in the previous section that, in our data, when health status which is increasing in the value of ill
health, exhibits a positive value of $C$ instead of negative value of $C$ implying that it is the rich who are at a disadvantage with respect to distribution of ill health.

When we build a regression model of ill health seeking to explore the determinants of ill health, the model shows a positive coefficient of income for ill health. This positive relationship between ill health and income is contrary to the intuitive relationship that is expected.

Reasons for these have been explored in another chapter. But as of now to get over this confounding relationship and to control for income we have divided our sample into four groups on the basis of per capita consumption expenditure (which is a proxy for income). Division of the data into subsamples is permissible for decomposition analysis as has been shown by Murdoch & Sicular.

Thereafter, we build a regression model that links health (measured by ill health in our case) and compute the determinants of the inequality in ill health. We compute the contribution of constant, residual and each of the explanatory factors in the total inequality of health.

The contribution of the error term indicates the amount of inequality in health that is attributable to unobserved factors. Contribution of the constant term in the case of health is a significant element in the overall health inequality because by undertaking a given amount of expenditure on public health, we can have a higher constant amount of health for all individuals. Moreover, the marginal impact of health seeking behaviour on health would also be higher for those individuals who have a greater access to better public health.

The constant term also includes the effect of genetic endowments on health and the impact of income on health.

**Decomposition of level of inequality V/S differences (changes) in Inequality**

Wagstaff et al (2000) have pointed out that decomposition can address two kinds of problems:
1. Causes of level of inequality - which originate from the inequalities in determinants.

2. Causes of differences in inequality - they may originate from
   (a) differences in inequality of determinants.
   (b) differences in mean of determinants.
   (c) differences in their impact on the dependent variable.

The point mentioned at 2(c) is noteworthy because it represents a potential tradeoff between improving inequality and improving the mean of the explanatory variable (Contoyiannis & Foster 1999). It is possible to carry out the second kind of analysis under the framework initially devised by Oaxaca (1973). The differences in the health between different groups can be obtained by fitting separate regression models for the two groups under question. (The two groups could be poor non poor; rural urban; educated uneducated; manual labor non manual labor; male female etc.). If $h_i$ is the health outcome of the $i^{th}$ individual, $x$ is the vector of explanatory variables and if for exposition we assume that there are two groups poor and non poor in the model then, the determinants of the differential health outcome between the two groups can be represented as follows:

$$ h_i = \begin{cases} 
\beta_{\text{poor}} x_i + e_i & \text{if poor} \\
\beta_{\text{non poor}} x_i + e_i & \text{if non poor}
\end{cases} $$

If it is assumed that the non poor have a more advantageous regression line than the non poor; that is to say that, at each value of $x$, the health outcome $h_i$ is better for the non poor than the poor and if it is also assumed that the non poor have a higher mean value of $x$, then the outcome would be that the poor have a lower mean value of health the non poor.

The gap between the mean outcomes $h_{\text{poor}}$ and $h_{\text{non poor}}$ is given by

$$ h_{\text{non poor}} - h_{\text{poor}} = \beta_{\text{non poor}} x_{\text{non poor}} - \beta_{\text{poor}} x_{\text{poor}} $$

$x_{\text{non poor}}$ and $x_{\text{poor}}$ are the explanatory vectors that are evaluated at the means for the non poor and poor respectively and the conditional expectations of the error terms are assumed to be zero. This gap in health outcome between health outcome of the two
groups can be thought of as being partly due to (a) differences in mean value of $x$ between poor and non poor (b) partly due to difference in the value of $\beta$ between poor and non poor. This may be expressed in two ways

$$h_{\text{non poor}} - h_{\text{poor}} = \Delta x \beta_{\text{poor}} + \Delta \beta x_{\text{non poor}}$$

or, $$h_{\text{non poor}} - h_{\text{poor}} = \Delta x \beta_{\text{non poor}} + \Delta \beta x_{\text{poor}}$$

where, $\Delta x = x_{\text{non poor}} - x_{\text{poor}}$

and, $\Delta \beta = \beta_{\text{non poor}} - \beta_{\text{poor}}$

In the first case the differences in $x$'s are weighted by the coefficients of the poor group and the differences in the coefficients are weighted by the $x$'s of the non poor group, in the second case the differences in the $x$'s are weighted by the coefficients of the non poor group and the differences in the coefficients are weighted by the $x$'s of the poor group. In both cases we are partitioning the gap in the health outcomes between the poor and the non poor into a part that is attributable to the fact that the poor have worse $x$'s than the non poor and a part attributable to the fact that ex hypothesi they have worse $\beta$'s than the non poor.