OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Dipin Gupta

Predictive effect of determinants of health on prevalence of diseases in India.

An approach for measurement in a multi-factorial and multi-level setting.

Oslo and Akershus University College of Applied Sciences,
Faculty of Social Sciences

Abstract

The burden of acute and chronic diseases in India is higher than their respective global averages. To mitigate this in an effective and resource efficient way, the concerned health policy should undertake interventions against the determinants that can predict the prevalence of acute and chronic diseases in a multi-factorial setting. In addition, the policy should be customized at appropriate administrative levels to account for the variation attributed to the local context. To identify the appropriate determinants for intervention, and administrative levels for customization, data on 17 determinants from 274 district nested within 21 states was used, and analyzed using a combination of multiple regression analysis and multi-level analysis techniques. Consequently, 8 determinants were identified to have predictive ability on prevalence of acute diseases, while prevalence of chronic disease could be predicted by 10 determinants. State level was identified as the appropriate level for customization of the policies concerned with all the predictive determinants, while district level was identified as the appropriate level for customization for only half of the predictive determinants.

Key Words

Health inequity, Determinants of health, Analytical Sociology, Multiple regression analysis, Multi-level analyis

Oslo and Akershus University College
Oslo 2016

Preface

The relation between society and health has gained significant attention in the last few decades. This is a two-way relation, with manifestations of an individual's society on his/ her health, and manifestations of an individual's health on his/ her society. As both health and society are intangible factors, studying their relation is particularly complicated. Determinants of health can help us quantify this relation to an extent, but their measurement is often challenging. This study is an attempt to describe a context specific process to measure determinants of health.

My interest in this study originated from the identification of some aberrant cause effect relations in India. Obesity is considered a risk factor for cardio-vascular diseases. However, despite lower prevalence of obesity in India, prevalence of cardiovascular diseases is higher than the respective global averages. Similarly, smoking is considered a risk factor for respiratory diseases. However, despite lower prevalence of smoking in India, prevalence of respiratory diseases is higher than the respective global averages. These findings indicated towards the role of other factors in determining the prevalence of diseases, and hence, this study was developed.

I would like express my gratitude towards my supervisor Prof. Per Arne Tufte for his extremely helpful guidance and dedicated efforts to make this thesis better. If I get any credit for this thesis, I owe it to him. In addition, I would like to thank Prof. Sony Pellissery for introducing me to this course.

I would like to thank my wife Anshu for her constant encouragement and support, even on the days when I had to prioritize the thesis over her. Finally, I would like to thank my parents for their never-ending faith in me.

I hope you enjoy reading this thesis.

Dipin Gupta
Oslo, 14-Nov-2016

Table of Contents

List of Tables

List of Figures

List of Annexures

List of Abbreviations

AHS:	Annual Health Survey
CSDH:	Commission on Social Determinants of Health
DLHS:	District Level Household and Facility Survey
DoH:	Determinant of Health, or Health Determinant
DoHs:	Determinants of Health, Health Determinants
DV:	Dependent Variable
GPG:	Global Public Good
HI:	Health Inequity, or Inequity in Health
HIs:	Health Inequities, or Inequities in Health
IMR:	Infant Mortality Rate
IV:	Independent Variable
MLA:	Multi-level Analysis
MMR:	Maternal Mortality Rate
MoHFW:	Ministry of Health and Family Welfare
NHM:	National Health Mission
NHP:	National Health Policy
OGD:	Open Government Data
SDG:	Sustainable Development Goals
SHP:	State Health Policy
U5MR:	Under Five Mortality Rate
UHC:	Universal Health Coverage
WHO:	World Health Organization

Chapter one: INTRODUCTION

Health is often looked upon as an important enabling factor for an individual's ability to function is a society. Without the adequate level of health, an individual is likely to lose the opportunity and freedom to participate in functions that are deemed normal for his/ her age and gender. In a perfectly egalitarian society, no individual should be placed at a functionally disadvantageous position based on his/ her health. In fact, in a perfectly egalitarian society, there should not be any difference in health of individuals to begin with. However, we live in an imperfect world where differences in health are often observed between individuals and between their groups.

The difference in health create a pseudo-division in the society, wherein individuals with a better status of health are likely to have better opportunities, and therefore, they are likely to hold better social and economic positions. This is similar to the circle of poverty, and I would like to call it the *circle of health*. The circle of health enables an individual with good health to get better health, but it disables a person with poor health to improve it. This is extremely unfortunate, especially because the factors that govern an individual's health are often beyond his/ her control.

The health of an individual is an outcome of multiple factors acting together, and these factors are commonly referred to as *determinants of health*. There is no finite number of determinants of health. In fact, they can vary from something as small as the breakfast that we eat, to something as big as the status of world politics. It will not be an exaggeration to say that *anything under the sun* has the potential to influence human health, either positively or negatively. So when health can be influenced by anything and everything, why do we bother about the determinants of health? This is because health is an intangible resource. We cannot buy or sell health *per se*. However, we can trade in determinants of health, but even this trade can just add to the probability of better or worse health. For example, an individual can buy healthier food, better medical facilities and comfortable assets, but that does not guarantee better health, it merely increases its probability.

It is relatively straightforward to explain how an individual can regulate his/ her health by trading in determinants of health. However, this takes a completely new meaning in context of community health. The health of a community is the aggregate sum of the health of its

individuals, and this is often measured in terms of various *indicators of health*, like morbidity rates and mortality rates. From the perspective of a Government, in order to regulate an indicator of health, first is to identify the determinants that it can be intervened upon, and second is to carry out the interventions in an effective and resource efficient way. However, this is not simple, and following are the reasons for it:

Determinants of health do not act in isolation. Multiple determinants act together to define the status of an individual's health, and all individuals in a community, irrespective of how much their health differ from each other, aggregate together to define community health indicators. Thus, it is very challenging to identify the determinants that hold the maximum probability to influence the community health indicators while acting at the level of individuals. In addition, individuals can be segregated based on their belongingness to groups, and such groups can be geographical or socio-economic categories. Each group can have sub-groups, which can again be based on geographical or socio-economic categories. In such a nested structure, it is likely that the variation among the individuals within a group will be less than the variation among the individuals from different groups. Therefore, we can say that an individual's belongingness to a group is in itself a determinant of his/ her health. Let us take the following example:

If Europe is one geographical group, all countries in Europe are its geographical sub-groups. It is likely that the variation in health among individuals within Norway will be less, when compared to the variation in health among individuals from whole Europe. Similarly, if we divide the population of a country as per income levels, it is likely that the variation in health among individuals from the top quartile is less, when compared to the variation in health among the entire population of the country. Comparisons like this can be described as per the context of a study.

In this study, I have tried to describe an approach to identify the determinants of health that can predict community health indicators in India, while encompassing the dual challenge of (i) multiple factors acting together, and (ii) the effect of belongingness to a group.

Chapter two: THEORETICAL BACKGROUND

"When health is absent, wisdom cannot reveal itself, art cannot manifest, strength cannot fight, wealth becomes useless, and intelligence cannot be applied."

- *Herophilus (335-280 BC, Greek Physician)*

In this chapter, I have reviewed the causal relationship between determinants of health (cause) and health inequities (effect). The chapter starts with a description of theories behind health inequities, with emphasis on examples from India. Then, the concept of determinants of health is discussed, with an emphasis on the challenges in their measurement. Towards the end of this chapter, I have provided a description of the mechanisms underlying the cause-effect relationship between determinants of health and health inequity.

1. Health Inequities

Health is defined as *'a complete state of physical, mental and social well-being and not merely the absence of disease or infirmity'* (WHO 1946). This definition of health was adopted by the representatives of 61 member states of the World Health Organization (WHO) at the International Health Conference in 1946, and it is widely used and accepted even today. Although this definition seems fairly simple and straightforward, it has deeper meanings and wide applications. If we decompose this definition, we reach two basic premises: first, health is beyond mere absence of disease or infirmity, and second, health is a combination of physical, mental and social well-being. Both these premises acknowledge of importance of social phenomena in describing individual health. Thus, health of an individual can be simply divided into two different but related aspects: *medical* and *social*.

The medical aspect of an individual's health (or his/ her physical and mental well-being as per the definition) is described by the absence or presence of diseases, and their consequent physiological, anatomical and psychological manifestations on the human body and mind. The social aspect of an individual's health (or his/ her social well-being as per the definition) reflects

on an individual's relative ability to participate in the functions that are deemed normal for his/ her constitutional factors like age and gender. For example, an individual has good medical health if he/ she do not have a disease, and has a good social health if he/ she are equally capable to a function as compared to other folks from his/ her constitutional group. Social aspect of an individual's health can also be called as the *functiona*l aspect of health as it refers to the functionally relevant manifestation of its medical aspect.

The importance of social (or functional) aspect of health has gained significant focus in recent decades. Nobel Laureate James Tobin introduced the concept of *specific egalitarianism* and applied it to health. He defined specific egalitarianism as *'non-market egalitarian distributions of commodities essential to life and citizenship'*, and elaborated that "the social conscience is more offended by severe inequality in nutrition and basic shelter, or in access to medical care or to legal assistance, than by inequality in automobile, books, clothes, furniture, boats" (Tobin 1970). In other words, specific egalitarianism advocates that certain specific goods like health should be less unequally distributed than people's ability to pay for them. Thus, distribution of health should not be governed by people's income.

S. Anand used Tobin's concept of specific egalitarianism to describe health as a *special good*. Anand states "health is not just an ordinary good such as income, wealth or prestige, because health is so deeply involved in our capability to function as an *agent* in the society. Inequalities in health mean a denial of *equality of opportunity*. If we want equal opportunity, then we must abolish all unnecessary and unjust differences in health" (Anand 2000).

Nobel Laureate Prof. Amartya Sen emphasized upon Anand's description of health as a special good, by stressing that the lack of health results in an individual's *lack of capability* and thus *lack of freedom.* He elaborates that human wellbeing involves capabilities, *i.e.* the capacity to act, to function, and to do things. Lacking capabilities mean lacking the opportunity to lead one's life as one wishes. Thus, ill health and disease means that the individual's capabilities are restricted, reduced and limited (Sen 2002).

The status of health as a special good means that we should be more averse to, or less tolerant of, inequalities in health than inequalities in income, as health has both *intrinsic* (within itself) and *instrumental* (for others) value. Income, on the other hand, only has instrumental value (Anand

2000). This means that we need good health not just because it helps us function better and achieve other goods (instrumental value), but also because we do not want poor health (for example pain) for ourselves anyways (intrinsic value). Contrary to this, we need money only to get other things (instrumental value) but not in a setting where money cannot buy anything. Anand further adds that there are sometimes reasons to tolerate income inequalities.

> Economists often assert, with some justification, that income incentives are needed to elicit effort, skill, enterprise, and so on. These incentives, and the resulting income inequalities, have the effect of increasing the size of total income (or the cake) from which, in principle, the society as a whole can gain (through taxation and possibly trickle down). Thus, the increase in the size of the cake has to be balanced against the income inequalities that must be tolerated to provide the appropriate incentives for efficiency. Furthermore, effort, skill, enterprise, and so on are regarded as legitimate and fair reasons for some people to earn, perhaps even to deserve, more than others. However, this incentive argument would not seem to apply in the case of health. Inequalities in health do not directly provide people with similar incentives to improve their health from which society as a whole benefits. There thus seem to be no incentive reasons for accepting inequalities in health, other than those that might be derivative on tolerating income inequalities *(ibid).*

However, I am inclined to differ from Anand's view slightly and argue here that similar to income inequalities, health inequalities also result in improving overall global health. Bettcher argues that health is a *global public good* (GPG), as it represents a positive sum, that is one person's good health has a positive effect for the rest of the population and that it is a key element for social and economic growth. The requirements of a GPG is that it covers countries all over the world, that the benefits of it are accessible for all without discriminating, and that it does not cause harm to future generations. Furthermore, a public good is non-excludable and non-rival, unlike private goods in a market (Bettcher 2009).

To elaborate on the above view of health as a GPG, let us imagine a world where every individual has absolutely the same level of health, thereby allowing everyone an equal opportunity to participate in the social and economic activities. In such a perfect world setting,

will it be relevant for anyone to strive towards better health? I would say no, and that is because there will be no functional advantage to have a better health. In other words, if everyone enjoys the same level of health, then no one will be deprived of participation in social and economic activities based on the status of their health. Thus, there will be no need to strive for better health when there is no realizable benefit of it. Nobody will know what better health is. This is a utopian view. However, we live in an imperfect world where there are wide differences in health of individuals.

As the status of their health places individuals (and the groups they belong to) in a functionally advantageous or disadvantageous position compared to other individuals (and groups), there is a constant endeavor by everyone to achieve a healthier life. Individuals with poorer health strive to minimize their functional disadvantage, while individuals with better health strive to maintain their functional advantage. As a net result, the overall health of global population improves significantly, but as everyone is getting better health, the differences in health still exist.

For example[1], the global average *life expectancy at birth* has more than doubled in the last century, from 31 years in 1900 to 66.4 years in 2000. In recent past, between 2000 and 2015, global average life expectancy increased by 5 years (7.5%), and currently stands at 71.4 years. However, the 2015 gap between regional life expectancies is still very high, and ranges from 60 years in Africa to 76.9 years in Americas. With increasing focus from the global community, health differences have been narrowing down in the recent past, but considerable differences still exists and they need to be addressed. For example, between 2000 and 2015, the difference between Global and African life expectancy has reduced by 27.8%, but a gap of 11.4 years still exists and needs to be mitigate.

Above was an example based on life expectancy as a *health indicator*[2]. Similar trends of past improvement and current differences can be noticed for various other indicators of health. For example[3], between 2000 and 2015, Infant Mortality Rate (IMR) reduced from 53.1 to 31.7 deaths

[1] Data source: World Health Statistics Report (WHO 2016d).

[2] A health indicator is a measure that is reported on regularly and that provides relevant and actionable information about population health and/or health system performance and characteristics. An indicator can provide comparable information, as well as track progress and performance over time (Canadian Institute for Health Information 2016).

[3] Data source: World Bank Open Data (World Bank 2016a).

per 1,000 live births, and Maternal Mortality Rate (MMR) reduced from and 341 to 216 deaths per 100,000 live births. However, wide differences still exist today, with IMR ranging from two in Norway to 104 deaths in Angola, and MMR ranging from three in Iceland to 1,510 deaths in Sierra Leone.

Table 1: Global variation in important public health indicators[4]

Health Indicator	***Average***	***Range (countries)***
Life Expectancy (years)	71	49-86
IMR (infant deaths per 1,000 live births)	35	2-104
U5MR (child deaths per 1,000 live births)	47	2-172
MMR (maternal deaths per 1,000 live births)	232	3-1,510

For any indicator of health, differences in health of individuals can be aggregated based on the individual's belongingness to a group. These groups can be *geographical locations* (like regions and countries, as already stated in the examples above) or *socio-economic parameters* (like income, education, housing, etc.). In addition to the differences observed between groups, variation in health is also observed between the sub-groups of a group, and between the individuals within a group or a sub-group. Various examples can be stated to elaborate this. However, in this study, I have focused on examples from India, and the same are discussed on page 10.

Differences in health are commonly referred in the literature as *health inequalities*. However, in the recent decades, the concept of *health inequity* has taken precedence over the concept of *health inequality*. The concept of health inequity differs from the concept of health inequality as the former encompasses health differences that are unnecessary, avoidable, unjust and unfair (Whitehead 1990). *For the rest of this document, health inequities are referred to as HI when referring to a single inequity in health, and as HIs when referring to multiple inequities in health.*

While defining health differences as HIs, it is a challenge to define which particular type of health difference is avoidable, unjust, and unfair. If a difference in health is purely random, or

[4] Data source: World Health Statistics (WHO 2016d).

entirely determined by natural, inborn, traits, or a result of conscious, well-informed, voluntary actions, then the health difference cannot be assessed as *unjust*. Alternatively, HIs are health inequalities caused by how societies are organized, how resources are distributed in society, and how health care is offered. Thus, "HIs exist when people fail to achieve good health because of inadequate social arrangements" (Elstad 2014). In other words, HIs exist when health differences are caused by *factors* that would disappear if societies have fair distributions of resources. The WHO further adds a rights based approach to HIs, and states that "HIs entail a failure to avoid or overcome inequalities that infringe on fairness and human rights norms" (WHO 2016b).

It should be noted that health is a *fundamental human right*. The article 25 of the United Nations' Universal Declaration of Human Rights states that "everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services" (United Nations 1948). The preamble of the WHO Constitution defines the right to health "as the enjoyment of the highest attainable standard of health" (WHO 1946).

Global health policies are now paying increasing attention to mitigate differences in health. In 2015, the United Nations adopted the 2030 Sustainable Development Goals (SDGs) with significant focus on equity in development. SDGs are a set of 17 goals covering a broad range of development issues. For example, SDG 3 calls for healthy lives for all at all ages, positioning equity as a core-crosscutting theme, while SDG 10 calls for the reduction of inequality within and between countries. "Equity is also a key consideration with regard to Universal Health Coverage (UHC), which is both central to the health goal and founded on the principle of equal access to health services without risk of financial hardship" (WHO 2016d).

1.1. Administrative hierarchy in India

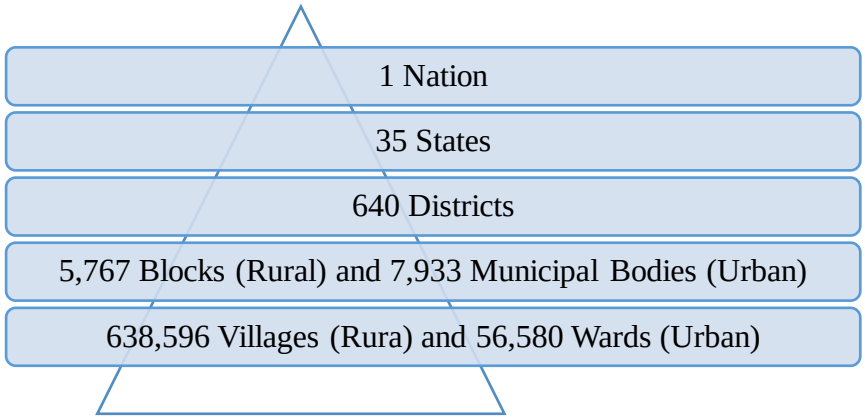
The Indian public administration structure is divided into five hierarchical levels. The top three levels are common to all areas, while the bottom two levels differ between rural and urban areas. From top to bottom, the three common administrative levels are *national*[5], *state*[6] and *district*. The

[5] The national level is also called as central level, while the district level is also called as peripheral or local level.

[6] Throughout this study, the term state, in reference to Indian administrative units, also includes union territories. A union territory differs from a state as it is considered economically and administratively weak to form its own

bottom two administrative levels are *block*[7] and *village* in rural areas, and *municipal body*[8] and *ward* in urban areas. In rural parts of India, 638,596 villages are clustered into 5,767 blocks. In urban areas, 56,580 wards are clustered into 7,933 municipal bodies. The blocks and municipal bodies are nested within 640 districts, which are further nested within 35 states[9].

Figure 1: Hierarchy of Indian public administration system



The Constitution of India delineates the roles and responsibilities of the Government at national and state levels only. Therefore, public policies in India are formed at the national and the state levels, and consequently, the associated commands and resources flow in a top-down manner. In this study, I have concentrated on the state and district levels of administration. This is done because of the following three reasons. (i) The nature of health administration in India is such that the objectives and implementation strategy of the health policies are drafted at the national level, customized at the state level and implemented at district level (this is explained more in the following paragraphs). (ii) Because of their very large number, drafting a policy or strategy for each administrative unit under the district level is not feasible. (iii) State and district are the common tiers to both rural and urban areas, and they thus maintain uniformity in hierarchy throughout the nation.

Government, and is thus ruled directly by the national Government through Lieutenant Governor as its administrator. There are 7 union territories in India.

[7] Blocks are also commonly known as tehsils, mandals, talukas or sub-districts.

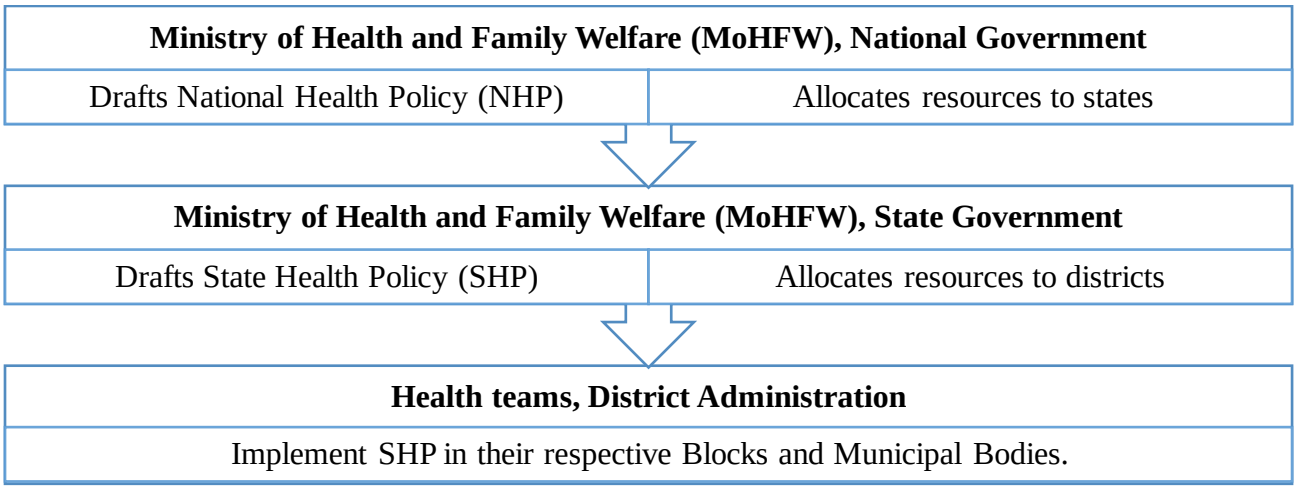
[8] Municipal Bodies include municipal corporations (for population of more than 1 million), municipalities (for population between 300,000 and 1 million) and municipal councils (for population below 300,000).

[9] Data source: Census of India (Ministry of Home Affairs, Government of India 2011a)

1.2. Indian health administrative system

Under the ambit of the general public administration structure described above, the *public health* care in India is administered at three levels: national, state and district. The Ministry of Health and Family Welfare (MoHFW) under the national Government sets the overall health objectives for the country, drafts National Health Policy (NHP), and allocates resources to states. The state level MoHFW customizes the NHP into State Health Policies (SHP), and allocate resources to their respective districts[10]. The health teams at the district level implement the SHP in their respective administrative areas. Thus, the Indian public health policies are drafted at national level, customized at state level and implemented at district level.

Figure 2: Indian health administration structure



1.3. Health inequity in India

Inequity in health can be discussed based on individuals' belongingness to a geographical or socio-economic group. Geographical groups are more commonly studies due to their relevance in the local and global administrative framework. Among the Indian administrative units discussed above, significant differences in key public health indicators can be observed, and the same are illustrated in the table below:

[10] The states have the liberty to customize policies and missions as the Constitution of India describes health as a state level subject with the national level Government having an overall regulatory role. However, health care in union territories is controlled directly by the national Government.

Table 2: Health inequity among different administrative units in India[11,12]

Public Health Indicator		*Range (States)*	*Range (Districts)*
Measures of Mortality*	IMR (infant deaths per 1,000 live births)	36- 68	20- 100
	U5MR (child deaths per 1,000 live births)	43- 92	23- 142
	MMR (maternal deaths per 1,000 live births)	167- 339	155- 437
Measures of Morbidity**	Prevalence of acute diseases	1.8- 20.8	0.4- 31.3
	Prevalence of chronic diseases	2.1- 16.0	0.4- 24.3

**based on a sample of 284 districts belonging to 9 states*
***based on a sample of 274 districts belonging to 21 states*

It should be noted that in all of the above-mentioned public health indicators, the range of variation at district level is larger than the range of variation at the state level. This is expected as districts are nested within states, and some districts in a state are likely to have a better (or a worse) health indicator than the state average. In addition, it can be established from the above data that in India, variation in key health indicators is observed both between states and within states (between districts).

The above is an example of inequity in health between the Indian health administration units, and is based on some of the most important health indicators. This example is important as Indian health policies are formulated and implemented at these levels, and the concerned health indicators are used globally. However, various other examples of HI can be established between other geographical and socio-economic groups, and based on other health indicators in a similar way. Following is an example of inequity in health based on socio-economic groups:

In its World Health Statistics Report, the WHO describes six *indicators* of health, segregated for three different socio-economic *parameters*. The six indicators are contraceptive prevalence, antenatal care coverage, births attended by skilled health personnel, immunization coverage, under-5 stunting, and U5MR. The three parameters are divided into two *categories* each, and they are place of residence (rural, urban), wealth quintile (highest, lowest) and education level of

[11] Data Source: Annual Health Statistics Report (MoHFW, Government of India 2016)

[12] Data Source: District Level Household and Facility Survey (MoHFW, Government of India 2014)

a mother (none, secondary or higher). Individuals belonging to a socio-economic category tend to portray a similar pattern of health, which can differ significantly from the pattern observed among individuals from a different category.

In India, significant difference can be observed in all the above-mentioned six health indicators, while comparing between categories of the three parameters. For example, prevalence of contraceptives is 66% higher and antenatal coverage is 5.5 times higher among people belonging to highest income quintile when compared with people belonging to lowest income quintile. Similarly, proportion of births attended by skilled health personnel is nearly three times higher, and immunization coverage for children is nearly two times higher among mothers who have at least secondary level education when compared with mothers who have no education. Among residential categories, the under-5 stunting and mortality rates are 22% and 35% higher (respectively) in rural areas when compared with urban areas. Refer to the table below for details.

Table 3: Health inequity among different socio-economic groups in India[13]

Public Health Indicator	***Place of Residence***		***Wealth Quintile***		***Educational Level of Mother***	
	Rural	*Urban*	*Lowest*	*Highest*	*None*	*Secondary*
Contraceptive prevalence (%)	45	56	35	58	46	50
Antenatal care coverage (%)	28	63	12	78	16	64
Births attended by skilled health (%)	37	73	19	89	26	75
DTP3 immunization coverage (%)	51	69	34	82	37	77
Children aged < 5 years who are stunted (%)	51	40	60	26	57	36
Under-5 mortality rate (per 1000 live births)	93	60	116	39	106	49

So far, in this chapter, I have tried to elaborate that the health of an individual not only includes his/ her medical fitness, but also his/ her social (functional) fitness. In addition, despite the importance of having good and egalitarian distribution of health, individuals differs in status of their health and such differences can be aggregated based on belongingness of the individual to a

[13] Data Source: World Health Statistics Report (WHO 2012)

certain geographical or socio-economic group. These differences in health are primarily a matter of organization of, and distribution of resources in, a local or global social setting (referred to as HIs). In India, public health policies are drafted at national level, customized at state level and implemented at district level, and inequities in health are observed both between states and within states (between districts). Organization of a society is the generic reason behind HIs and it must be broken down into tangible factors that can be evaluated and intervened to bring about the desired egalitarian distribution of health. Such tangible factors are commonly known as *Determinants of health*, and the same are discussed next.

2. Determinants of health (DoHs)

Poor health is described by a set of three terms: *sickness, illness and disease*. In 1973, Susser, an epidemiologist, proposed some definitions that remain useful. He used the term illness to refer to the subjective sense of feeling unwell; illness does not define a specific pathology, but refers to a person's subjective experience of it, such as discomfort, tiredness, or general malaise. The way a patient reports symptoms is influenced by his/ her cultural background, and Susser applied the term sickness to refer to socially and culturally held conceptions of health conditions, which in turn influences how the patient reacts (AFMC 2009). The term disease implies a focus on pathological processes that may or may not produce symptoms and that result in a patient's illness. For example, a patient complains of tiredness and malaise, his illness as he experiences it. He consults a doctor about it, because he believes that he might have a sickness. The doctor might attribute the patient's symptoms to a thyroid condition, a disease. Patients suffer illness, doctors diagnose and treat disease (Eisenberg 1977).

In simpler words, disease leads to illness, which is interpreted as sickness by the individual and society. This is the general understanding, but not a rule. For example, chronic diseases like cancer and hypertension may not manifest any symptom for a long time (disease without illness). Stress due to an existing disease can become precursor for another disease like diabetes (illness leading to disease). Placebo cures sickness in absence of disease. Some countries accept homosexuality, while some classifies it as an illegal and punitive activity, a sin, or a disease. Similarly, it is difficult to describe psychiatric disorders, somatic disorders, disabilities,

congenital disorders, obesity, substance abuse, and similar culturally sensitive issues, in the conventional disease-illness-sickness format. "So, how do we distinguish properly between real diseases, and human behaviors or characteristics that we just happen to find disturbing?" (Scully 2004). A major part of such interpretations is governed by the local cultural and socio-economic factors.

I used the above example of diseases-illness-sickness format to illustrate how cultural, socio-economic and other relevant factors play an important role in defining what is good or bad health. However, a more practical and significant role of these factors is in determining how an individual reaches a state of good or bad health, i.e. how do they actually determine an individual's health and result in HIs.

> Health inequities within countries are associated with a variety of factors, several of which are encountered uniformly across all countries. Examples include sex, age, economic status, education and place of residence. Other factors may be more specific to a regional or country situation, such as migrant status, race, ethnicity, caste, religion or other characteristic that can differentiate minority subgroups. A movement towards equity in health depends, at least in part, on strong health information systems that collect, analyses and report disaggregated data covering all health areas. This is recognized in SDG Target 17.18, which calls for efforts to build capacity to enable data disaggregation by a number of stratifying factors, including income, sex, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national context (WHO 2016d).

The issue of HI in SDGs is focused on five determinants: gender, age, socio-economic status, place of residence and special disadvantaged groups like refugees. However, the spectrum of health determining factors is very vast. All such factors together constitute the Determinants of health.

Determinants of health are simply defined as factors that contribute to a person's current state of health (CDC 2014). *Henceforth in this document, a single determinant of health is referred to as DoH, and multiple determinants of health are referred to as DoHs.* We can also define a DoH as a bloc of factors or conditions, which are presumed to have a general influence on people's

health, their longevity, and their level of ill health (Elstad 2014). The concept of DoHs is very broad and can be extended to include anything that influences human health.

A particular health situation is affected by multiple DoHs, and each DoH affects multiple health situations. In fact, there is no finite number of DoHs. To simplify this, DoHs are often classified in a few general categories. The WHO classifies DoHs in following three categories: *social and economic environment, physical environment, and person's individual characteristics* and *behavior* (WHO 2016b). The Center for Disease Control (CDC) classifies DoHs in following five categories: *biology and genetics, individual behavior, social environment, physical environment, and health services* (CDC 2014). Another commonly used classification of DoHs was given by Whitehead and Dahlgren in 1991 (Whitehead, Dahlgren and Gilson 2001). I have used this classification to divide determinants in later parts of this study, as it is a comprehensive and widely accepted classification. Following are the details of this classification.

Whitehead and Dahlgren divided DoHs in following five categories. (1) *Age, sex and constitutional factors* are the factors that cannot be controlled or changed, for example, genetic predispositions to congenital disorders, skin color, family history of diseases, demographic dividend and sex ratio. (2) *Individual Lifestyle Factors* are chosen by the individual and are thus a matter of choice, for example, dietary habits, sedentary lifestyle, and substance abuse. (3) *Community and Social Networks* are an accumulation of the social capital with which people live, *i.e.* how their family, community or Government supports them. For example, social security measures, resource support from parents, personal and professional network, etc. (4) *Living and Working Conditions* is the biggest group of DoHs and includes everything that defines the individual level material and non-material conditions in which people live and work on a day-to-day basis. For example, type of housing, availability of clean water, sanitation facilities, availability of health care services, employment rate, literacy rate, quality and quantity of available food, etc. (5) *General Socio-economic, Cultural and Environmental conditions* are the macro-level conditions that cannot be controlled individually. For example, economic stability of the country, predominant industry, cultural factions and associated discrimination, religious practices, air pollution, water pollution, and so on.

Figure 3: Whitehead and Dahlgren classification of DoHs

In order to use the concept of DoHs in a medical or social research, it is important that we can identify and measure the concerned determinants. However, measuring DoHs is not a straightforward task, and involves many challenges. These challenges affect the scope of a study and the methods that can be used to achieve the study objectives. Following is a brief description of challenges in measurement of DoHs:

2.1. Challenges in measurement of determinants of health

2.1.1. Multiple parameters for measurement

The number of DoHs is not finite. In addition, each DoH can be measured in multiple ways by using different parameters. For example, abusive habits are many, most common being consumption of alcohol and tobacco, and they can be measured as quantity of consumption, frequency of consumption, mode of abuse and quality of substance abused. Another example is employment, which can be measured as percentage of population unemployed, percentage of population engaged in agriculture, per-capita income and contribution to economy, and so on. It

is important to figure out what exactly needs to be measured about any DoH and how can that be actually measured? This depends a lot on the context and objectives of any study for which DoHs are being measured.

2.1.2. Qualitative DoHs

Many DoHs are hard to quantify and need careful regional and contextual understanding for measurement. Almost all cultural, religious, political and social practices pose this challenge. For example, gender roles, caste based occupations, polygamy, high fertility rates, low maternal age, preference for male child, superstition, alternate medical therapies and many such similar practices are difficult to quantify. Another aspect of qualitative DoHs is individual subjectivity involved in their measurement. For example, stress is a very common health risk. However, measuring stress can be highly subjective as sources and threshold of stress can vary from person-to-person. Thus, some of these complex qualitative determinants can only be measured though proxy measures.

2.1.3. Inter-relations between DoHs

DoHs have simple measurement and interpretation when measured as a one-on-one relation. For example, measuring and understanding the role of education in predisposing an individual to health risks is straightforward. However, one-one-one relations between a determinant and a health outcome are very likely to change when we consider the role of other determinants simultaneously. Following are some examples[14]: (i) 46% children are exclusively breastfed for the first six months after their birth in India, which is 24% higher than the global average of 37%. Thus, less children should be malnourished in India when compared with global average. However, the proportion of malnourished children in India is more than twice the global average. (ii) Prevalence of obesity in India is nearly one-seventh of global averages. Accordingly, the prevalence of cardiovascular diseases should be lower in India. However, it is found to be almost 33% higher than the global average. (iii) Prevalence of tobacco smoking in India is 27% lower than the global average. Accordingly, the prevalence of respiratory diseases should have be lower, however, the opposite is observed, with prevalence of respiratory disorders in India to be nearly three times the global average.

[14] Data Source: World Health Statistics Report (WHO 2012)

In the above examples, breastfeeding, obesity and smoking are DoHs, while prevalence of malnourishment, cardiovascular and respiratory diseases are indicators of health. The observed relations are contradictory to expectations, and are a result of other determinants that act simultaneously on determining the status of health indicators. Therefore, in reality, an individual's health is an outcome of multiple determinants acting together. Thus, we cannot ignore the role of other variables, and because of this, the one-on-one relations can change in magnitude as well as in direction of association, with incorporation of extra variables in the equation, which can be selected as per the context of a study.

When the direction of a relation reverses due to incorporation of a third variable, it is referred to as *Simpson's paradox*. *Simpson's paradox* occurs when an observed association between two variables is reversed after considering the third variable. Having two different conclusions makes this phenomenon paradoxical. Source of this paradox is the interactions between the third variable and any of the other two variables. "The third variable which causes the reversal at the direction of association is also called *confounding variable*. *Confounding variable* is defined as an extraneous variable, which tends to confound our reading and to bias our estimate for the effect studied" (Alin 2010). The *confounding variable* can also result from disaggregation of data to subpopulations. "If we partition the data into subpopulations, each representing a specific value of the third variable, the phenomena appears as a sign reversal between the associations measured in the disaggregated subpopulations relative to the aggregated data, which describes the population as a whole" (Pearl 2013). Edward H. Simpson first described this phenomenon in 1951, hence the name.

2.1.4. Levels of measurement

DoHs can be described at two different levels: *individual* and *collective*. Individual level DoHs are defined and measured for a single person. Examples of individual DoHs are employment status and literacy level of an individual. Collective DoHs are defined and measured for a group of individuals. A group of individuals can be defined at the level of an organization, a neighborhood, a municipality, a district, a nation, the whole world, or any such agglomeration of individuals. Examples of collective DoHs are pollution levels in a municipality and poverty in a country.

It should be noted that individual level DoHs can be easily *aggregated* to a relevant collective level by using simple mathematical computations like percentage and average. For example, employment status of all individuals in a group can be aggregated to define percentage of employed (or unemployed) individuals in that group, and literacy level of all individuals in a group can be aggregated to define average literacy level in that group. Such aggregation can be theoretically done for all individual level DoHs. However, the reverse is not true, and not every collective level DoHs can be easily *disaggregated* to the individual level. For example, pollution level in a municipality cannot be disaggregated to define pollution level for an individual of that municipality, but total income of a country can be disaggregated to define average income of an individual of that country.

Thus, a theoretically relevant mathematical aggregation and disaggregation of DoHs can be defined as per the context of a study. However, results obtained from aggregated or disaggregated data should be carefully interpreted, as their application at levels other than the level of their measurement can be sometimes misleading. Application of results obtained from analysis of collective (or aggregated) level data at individual (or lower) levels can sometimes lead to false interpretations, and this is commonly known as *ecological fallacy*. On the other hand, application of results obtained from analysis of individual (lower) level data at collective (or higher or aggregate) levels can also sometimes lead to false interpretations, and this is commonly known as *atomistic fallacy*. I will discuss ecological fallacy a little more in detail here as this study uses aggregate level data.

The ecological fallacy is the inappropriate assumption that relationships at the aggregate level will also hold at the individual level. It is entirely possible to find a relationship at the aggregate level that does not hold true at the lower level. This phenomenon was first highlighted by William Robinson in 1950. Robinson looked at the relationship between literacy and the proportion of immigrants across 48 states is USA. He found a correlation of 0.53, indicating that areas with lots of immigrants were highly literate. However, at the individual level, the correlation between immigrant status and literacy was -0.11. Immigrants were less likely to be literate. The correlation at the aggregate level occurred because of a tendency for immigrants to settle in areas where the native-born population was highly literate (generally, urban areas). "Robinson made this phenomenon popular, but the term ecological fallacy was later coined in

1958 by Selvin, referring to the invalid transfer of aggregate results to individuals" (Subramanian, et al. 2009). So why do we use aggregate data in the first place?

> Because they are inexpensive and easy to obtain, because they may be available under circumstances in which survey data are unavailable, and because they eliminate many of the measurement problems of survey research, data on geographic units such as counties or census tracts are often used by political scientists to measure individual behavior. This has involved us in the long-standing problem of inferring individual-level relationships from aggregate data (Phillips 1969).

Over the last few decades, it has been well established that the use of aggregate data may yield correlation and regression coefficients exhibiting considerable bias above their values at the individual level. Probably the most serious disadvantage of using aggregate data is the inherent difficulty of making valid multi-level inferences based on a single level of analysis. "Although the ecological fallacy has been widely discussed and publicized, it is still a common error in studies involving causal inference" (Clark and Avery 1976). It is not possible to avoid ecological fallacy all together. However, with advanced statistics, it is possible to reduce chances of its occurrence to a good extent.

The above-mentioned challenges demand careful measurement of DoHs with adoption of advanced statistical techniques. There is no doubt about the utility of DoHs in providing insight to many academic and non-academic issues, and this potential can be harnessed through comprehensive understanding of relation of a determinant with other determinants and health outcomes, and application of potential statistical technical tools to help in the analysis.

2.2. Social determinants of health

In recent decades, in order to understand and control HIs, the concept of DoHs has been modified into the concept of *social* DoHs. Social DoHs are the conditions, in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life. These forces and systems include economic policies and systems, development agendas, social norms, social policies and political systems. "The social DoHs are mostly responsible for HIs- the unfair and avoidable differences in health status seen within and between countries" (WHO 2016c). The Commission on social determinants of health (CSDH) was set up

by the WHO in 2005 to marshal the evidence on what can be done to promote health equity and to foster a global movement to achieve it. The final report submitted by the CSDH states that:

> The poor health of the poor, the social gradient in health within countries, and the marked HIs between countries are caused by the unequal distribution of power, income, goods, and services, globally and nationally, the consequent unfairness in the immediate, visible circumstances of people's lives- their access to health care, schools, and education, their conditions of work and leisure, their homes, communities, towns, or cities- and their chances of leading a flourishing life. This unequal distribution of health-damaging experiences is not in any sense a natural phenomenon but is the result of a toxic combination of poor social policies and programs, unfair economic arrangements, and bad politics. Together, the structural determinants and conditions of daily life constitute the *social* DoHs and are responsible for a major part of HIs between and within countries (CSDH 2008).

The word social in social DoHs should not be confused as reference to factors that are related to society in general. Rather, it gives a specific direction to interpretation of the concept of DoH, which is relevant to social and political sciences. Moreover, social DoHs can be looked upon as a sub-set as well as an extension of DoHs, a sub-set because it reduces focus on natural, in-born or justified factors behind health differences, and extension as it increases focus on policies and social systems. Thus, DoHs and social DoHs can be treated as the same set of health determining factors, with the theoretical difference that the former focuses on explaining all health differences (health inequalities) and the later focuses on explaining unjust health differences (HIs).

Treating social DoHs as the cause and HIs as the effect is not as simple as it may appear. This is because no single determinant acts in isolation to affect any single health outcome. In reality, multiple determinants act together to bring about multiple health outcomes in multiple ways. Thus, establishing *general laws* on the cause-effect relation between social DoHs and HIs is practically not possible. The best social scientists can do is to understand and explain the *mechanisms* behind this relation and help predict the consequences, with some probability but not certainty. In the following text, I will discuss this further by elaborating the mechanisms behind the relation between social DoHs and HIs. In addition, for the rest of this document, I will

use the terms determinants of health (DoHs) and social determinants of health (social DoHs) interchangeably.

3. Underlying Mechanisms

Social phenomenon can be explained, as James Coleman elaborates in his book Foundations of Social Theory, in two ways: by establishing statistical association between events and by examining the processes internal to the social system (Coleman 1990). Although Coleman does not rule out the need for statistical associations, he recommends that in social sciences, an analysis should explain the behaviors of the system by recourse to the behaviors of its parts. He explains that the explanation of the behavior of social systems entails examining processes internal to the system, involving its component parts, or units at a level below that of the system. He calls this as *internal analysis of system behavior*, and justifies his recommendation on the basis that data in social sciences in generally gathered at levels lower than the system under study (most commonly at individual level), such data is more likely to be stable, and makes more sense as the consequent interventions are implemented at lower levels as well.

3.1. Analytical Sociology

The focus on explaining *internal processes* has gained significant momentum in the past few decades, and has given rise to a specific domain of sociology called *analytical sociology*. Peter Hedström and Peter Bearman, editors of the *The Oxford Handbook of Analytical Sociology*, explains that:

> Analytical sociology is concerned first and foremost with explaining important social facts such as network structures, patterns of residential segregation, typical beliefs, cultural tastes, common ways of acting, and so forth. It explains such facts not merely by relating them to other social facts-an exercise that does not provide an explanation-but by detailing in clear and precise ways the mechanisms through which the social facts under consideration are brought about. In short, analytical sociology is a strategy for understanding the social world. (…) Analytical sociology explains by detailing mechanisms through which social facts are brought about, and these mechanisms

invariably refer to individuals' actions and their relations that link actors to one another (Hedström and Bearman 2011).

The traditional view in sociology is that an explanation of a phenomenon consists in showing that it is expected given a general and applicable causal law, for example the Helpel's verison of the *deductive-nomological model*. Elster explains this as follows: "to explain an event is to cite a set of initial conditions together with a statement to the effect that whenever those conditions obtain an event of type follows" (Elster 2007). "This is now rather outdated view and has never played any serious role in sociology, however, sociological explanation is all about mechanisms and statistical associations and has been so for decades" (Hedström and Bearman 2011). Elster supports this by explaining that "even when we establish a general law in sociology, this does not always result in an explanation. Often, a general law helps us predict only the direction of an effect but not its magnitude, and such laws are referred to as weak laws" (Elster 2007). This is the reason why mechanisms are often preferred over general laws in explaining social phenomenon. Further, social sciences are more concerned with explaining than predicting, and a reliance on mechanisms instead of general laws, helps us to explain better than to predict.

Daniel Little has emphasized that "analytical sociology is not just another new paradigm for sociology. Instead, it is a reconstruction of what valid explanations on sociology must look like, once we properly understand the logic of the social world" (Little 2012). He stresses on the views of Peter Demeulenaere on analytical sociology that much of the existing sociology conforms to this set of standards-but not all, and the non-conformers are evidently judged non-explanatory. Demeulenaere states that "analytical sociology should not therefore be seen as a manifesto for one particular way of doing sociology as compared with others, but as an effort to clarify (analytically) theoretical and epistemological principles which underlie any satisfactory way of doing sociology (and, in fact, any social science)" (Demeulenaere 2011).

The units of analysis in analytical sociology are mostly individuals, and sometimes groups of individuals. Hedström elaborates this by stating that "in sociological inquiries, the *core entity* always tends to be the *actors* in the social system being analyzed, and the core *activity* tends to be the *actions* of these actors" (Hedström 2005). Coleman refers to this as *internal analysis of system behavior,* Elster refers to this as *opening the black box,* and Hedström and Bearman refers to this as *achieving causal depth.* By *causal depth,* Hedström and Bearman refer to the explicit

identification of the *micro foundations*, or the social *cogs* and *wheels*, through which the social facts to be explained are brought about. "The central cogs and wheels of social life are actions and relations. Actions are important because all the things that interest sociologists are intended or untended outcomes of individuals' actions. Individuals' actions are typically oriented towards others, and therefore relations to others are central when it comes to explaining why individuals do what they do" (Hedström and Bearman 2011).

3.2. Methodological Individualism

The reliance on individuals as the units of explanation in analytical sociology is based on the doctrine of *methodological individualism.* This doctrine was introduced as a methodological precept for the social sciences by Max Weber in 1922. "It amounts to the claim that social phenomena must be explained by showing how they result from individual actions, which in turn must be explained through reference to the intentional states that motivate the individual actors" (Heath 2015). Methodological individualism has been widely associated with *rational choice theory*, which states that an individual is assumed to take account of available information, probabilities of events, and potential costs and benefits in determining preferences, and to act consistently in choosing the self-determined best choice of action.

However, more relevant to analytical sociology, is a refined version of methodological individualism called as the *structural individualism*. Hedström and Bearman describes that "structural individualism is a methodological doctrine according to which all social facts, their structure and change, are in principle explicable in terms of individuals, their properties, actions, and relations to one another. It differes from traditional notions of methodological individualsim by emphasizing the explanatory importance of relations and relational structures" (Hedström and Bearman 2011). In other words, structural individualism attributes substantial explanatory importance to the social structures in which individuals are embedded. Little refers to this as *structural localism* (Little 2012), while Demeulenaere relates this to the term *structural sociology* (Demeulenaere 2011) and states that it is concerned with the effects these structures, once created and maintained, have on the behaviour of individuals or categories of individuals.

By acknowledging the doctrine of structural individualism, we can appreciate that individuals are *nested* within groups, for example, students in classes, employees in firms, residents of geographical units, and so on. Nesting individuals within groups helps in describing the impact

that a group has on its constituent individuals. This is important as we are often encountered with situations where an individual's belongingness to its geographical or socio-economic group is an important governing factor for his/ her general characteristics like health and income. This forms the basis of explaining the mechanisms behind the relation between HIs and social DoHs, where the focus is on differences in health attributed to an individual's belongingness to a certain geographical or socio-economic group.

We have discussed before (on page 18) that in social research, the best-case scenario is that the data is collected at individual level, primarily because interventions consequent to the research are targeted at individuals. "Although variability is most commonly noticed between individuals, it can also be defined between their groups, and one may draw incorrect conclusions if no distinction is made between these different sources of variability" (Snijders 2012). This means that we should take into consideration, as to what extend the variability among individuals is attributed to their belongingness to a group. For this, data collected is individual level can be aggregated for their respective groups, and analyzed using appropriate statistical techniques. *Multi-level Analysis* (MLA) is one such technique that can be used for analysis of nested data. I have used MLA for analysis in this study, and will discuss it more in following chapters. For the rest of this chapter, I will focus on defining and describing social mechanisms.

3.3. Defining Mechanisms

The evolution of methodological individualism into structural individualism is important for analytical sociology as it forms the basis for decribing mechanisms behind social phenoenon. Various definitons of social mechanisms exist in literature. Hedström and Bearman defines it "as a constellation of *entities* and *activities* that are linked to one another in such a way that they regularly bring about a particular type of *outcome*, and we explain an observed outome by referring to the mechanism by which such outcomes are regularly brought about" (Hedström and Bearman 2011). They descibes this by stating that mechanisms can be said to consist of entities (with their properties) and the activities that these antities engage in, either by themselves or in concert with other entitites. These activities bring about change, and the type of change brught about depends upon the properties and activities of the entities and the relation between them.

Elster defines social mechanisms as "frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with intermediate

consequences" (Elster 2007). He elaborates that a mechanism explains by opening up the *black box* and showing the *cogs* and *wheels* of the internal machinery. A mechanism provides a continuous and contiguous chain of causal or intentional links between the explanans and the explanandum.

Demeulenaere also emphasizes that analytical sociology depends closely on the methodology of social causal mechanisms. "The analytical part of the phrase involves identifying separate things, and the social mechanisms idea says how these things are related. Causal mechanisms are expected to be the components of the linkages between events or processes hypothesized to bear a causal relation to each other" (Demeulenaere 2011).

3.4. Describing Mechanisms

There is no blueprint to describe a social mechanism, as even the simplest cause-effect relation can dramatically change with context. This is the challenge of social sciences, and it is in fact the reason why we need mechanisms in the first place. However, there are some basic concepts that can be used to describe mechanisms in general. They are discussed below:

Causality is the relation between an event (*the cause*) and a second event (*the effect*), where the first event is understood to be responsible for the second. Anything that affects an effect is a *factor* of that effect. There are two views towards causation: *unitary* and *plural*. The *unitary (single cause-single effect)* view has been criticized on the basis that unitary perspective betrays a narrow conception of causation and is not reflective of the broad range of causal arguments present in the fields of social science today. Instead, the fields of the social sciences are characterized by a *plurality (multiple causes-single effect)* of causal assumptions. Plurality can also be referred as the *net effect* of multiple causes. The unitary and plural views of causality are also called as *mono-causal* and *multi-causal* views, respectively. However, the terms mono-causal/ multi-causal are more commonly used in medical and epidemiological literature, while the terms unitary/ plurality are more commonly used in social sciences.

Causality can also be defined as *probabilistic* or *deterministic*. "Probabilistic causation designates a group of theories that aim to characterize the relationship between cause and effect using the tools of probability theory. The central idea behind these theories is that causes change the probabilities of their effects" (Hitchcock 2012). "Deterministic causality is, roughly

speaking, the idea that every event is necessitated by antecedent events and conditions together with the laws of nature" (Hoefer 2012).

The *causes* in a cause-effect relation can be divided into two categories, *direct factors* and *intervening factors*. Direct factors affect an effect directly, i.e. without any intervening factors, while intervening factors are the intermediate events in a cause-effect chain. Another way to classify causes (or factors) is as *necessary causes, sufficient causes and contributory causes.*

> If X is a necessary cause of Y, then the presence of Y necessarily implies the presence of X. The presence of X, however, does not imply that Y will occur. If X is a sufficient cause of Y, then the presence of X necessarily implies the presence of Y. However, another cause Z may alternatively cause Y. Thus, the presence of Y does not imply the presence of X. A cause may be classified as a contributory cause, if the presumed cause precedes the effect, and altering the cause alters the effect, regardless of whether either the cause or the effect appears only in the presence of the other (Reigalman 1979).

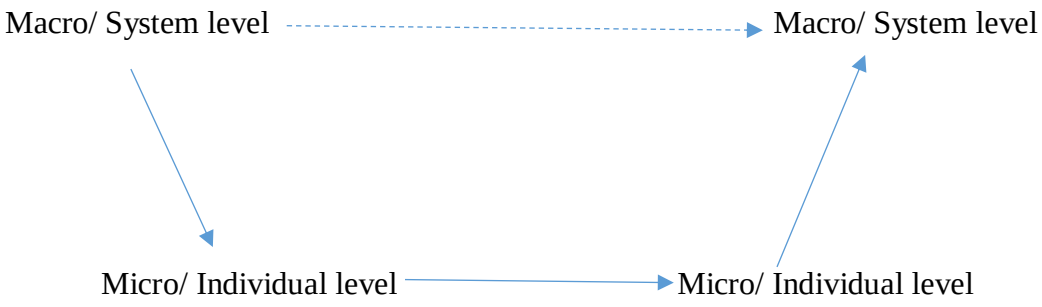
At any point of time, multiple DoHs act on multiple indicators of health, in multiple ways that change with the context and area of the study. Therefore, studying the cumulative impact of all determinants on one or more indicators is of little or no use, as the result will be too vague and will not help in designing and implementing interventions towards any desired change. Thus, there is a need to understand one-on-one relations between a determinant and an indicator, but in a multi-factorial setting, as no determinant acts in isolation, and its net effect is dependent on all other determinates acting simultaneously. This is the reason why it is nearly impossible to establish general laws regarding relations between DoHs and HIs. The best we can do is to explain the relation and establish *weak links* that can predict the direction and magnitude of the relations with some probability.

However, defining weak links in a social causal mechanism is generally not a straightforward process. This is because social causal relations are often pluralistic and probabilistic in nature, can involve complex cause-effect chains and nexuses, and can change significantly with the context and geography of study. Thus, social causes are mostly intervening or contributory in nature. They are seldom direct, necessary or sufficient in nature. "In a well-known attempt to use the terminology of necessary and sufficient causes to define what it is for one thing to be cause

of another thing, J. L. Mackie proposed that causes are at a minimum *INUS conditions*, that is, *Insufficient but Necessary parts of a condition which is itself Unnecessary but Sufficient* for their effects" (Brennan 2012). INUS conditions are important in describing causal social mechanisms as all causes can be regarded as INUS conditions, and that leads to a simpler understanding of a rather complex causal nexus.

While describing causal social mechanisms, it is also important to disintegrate the cause-effect proposition into parts. James Coleman suggested dividing a cause-effect proposition into three parts: one with the independent variable characterizing the society and a dependent variable characterizing the individual; a second with both independent and dependent variable characterizing the individual; and a third with the independent variable characterizing the individual and the dependent variable charactering the society. Thus, "the first relation is from the system level to the individual level (macro-to-micro), the second is wholly at the individual level (micro-to-micro), and the third is from the individual level to the system level (micro-to-macro). Therefore, the proposition begins and ends at macro levels, but in between, it dips to the level of individual" (Coleman 1990). The resultant depiction of these parts is referred to as *Coleman's Boat.*

Figure 4: Coleman's Boat: three parts of a cause-effect proposition



Hedström and Swedberg have classified the mechanisms pertaining to each of the three parts of Coleman's boat as *situational, action-formation,* and *transformational mechanisms* (Hedström and Swedberg 1998). Those mechanisms that explain the influence of macro forces on more micro level phenomena are situational mechanisms. Those that operate solely at the micro level linking cognition to behavior are action-formation mechanisms. Finally, those that describe how micro level factors affect the macro level are transformational mechanisms.

In addition to the general concepts of causality described above, describing causal social mechanisms will also involve knowledge of specific theories related to the topics under study. Like in this study, to describe the mechanisms behind the cause (DoHs) and the effect (HIs), theories of *generalized susceptibility* and *social mobility* can be useful.

3.5. Generalized Susceptibility and Social Mobility

The theory of *generalized susceptibility* states that individuals belonging to certain social groups are specifically advantaged or disadvantaged with respect to their health outcomes. Individuals belonging to disadvantaged groups are biologically or otherwise more vulnerable to diseases. Because of this, the theory of generalized susceptibility is also known as the theory of *generalized vulnerability* (Elstad 2014). The theory of generalized susceptibility/ vulnerability encompasses that diseases are consequence of the susceptibility or vulnerability of an individual, which is endowed upon him/ her by virtue to his/ her belongingness to a group and explained as a combination of multiple underlying factors (or DoHs).

Individuals hold different *social positions*, which determine their vulnerability towards health outcomes (generalized susceptibility). Such social positions are described by the socio-economic stratification of the society that an individual belongs to, and the health status is attributed to an individual by virtue of his/ her social position. It can be argued that an individual is born in a social position and thus his/ her social position it is a matter of *natural* random selection. Therefore, health differences based on inherent social positions should be classified as health inequalities (just) rather than HIs (unjust). However, not all individuals belonging to the same social position exhibit similar health. Thus, although social positions can explain the overall expected health status of a group of people, the individual level differences in health are determined by factors external to the concerned social position. Such *external* factors determine how the life course of individuals belonging to the same social position differs over time.

The manner in which the *social position* of an individual changes over his/ her life span is referred as *social mobility*, and that can be used to explain HIs among individuals. "People stay in, move out of, or move into, a social position over their life course" (Elstad 2014). This movement is attributed to various opportunities external to an individual's social position. The change in social position of an individual may improve, deteriorate or have no effect on his/ her

health when compared with the health of individuals from his/ her original social position. For example, if parents belonging to a lower socio-economic position educate their child so that he/ she gets a good job, then the child is likely to have a higher socio-economic position than his/ her parents and peers, which will enable him/ her to access and afford better medical facilities, and adopt a healthier lifestyle. On the contrary, an individual from a higher socio-economic position may fall to a lower position in the hierarchy due to financial and health liability of substance abuse.

DoHs can be attributed for explaining variation in health between individuals from different social groups as well as between individuals within the same social group. The variation in status of health *between* different groups can be explained as follows: multiple DoHs act together to determine the extent of susceptibility or vulnerability of individuals belonging to a group. To explain the variation in health *within* a group, DoHs can be looked upon as the external factors responsible for social mobility and its consequent influence on health.

4. Key Points from this chapter

The key points from this chapter can be summarized as follows: (i) Health is regarded as a special good because it is an important determinant of an individual's freedom, opportunities, and ability to function in a society. (ii) Unjust differences in health of individuals are common, and are referred to as health inequities (HIs). (iii) Inequities in health are observed both between and within groups of individuals, and such groups can be classified as per geographical and/ or socio-economic parameters. (iii) In India, public health policies are drafted at national level, customized at state level and implemented at district level, and inequities in health are observed both between states and within states (between districts). (iv) Factors that contribute towards the status of health of an individual are called as determinants of health (DoHs). (v) The causal relation between DoHs (cause) and HIs (effect) is pluralistic and probabilistic in nature. (vi) Multiple factors influence a health outcome at a given time, because of which, the causal relation between social determinates of health and HIs cannot be based on general laws. (vii) To explain causal social relations, like the one between DoHs and HIs, detailed understanding of the underlying mechanisms is required.

Chapter three: NEED ASSESSMENT and RESEARCH RATIOANLE

In this chapter, I will bring together the key points from the previous chapter to describe the need for this study, define the research problem, and build the hypothesis.

1. Need Assessment

In the previous chapter on theoretical background, we noticed that the Indian health administration structure is hierarchal in nature, with three important levels of administration: national, state, and district (on page 10). These three levels are geographical units with separate administration, and the administration at each level is concerned with specific roles. At the national level, health administration is concerned with drafting the National Health Policy (NHP) and allocating resources to states to meet the objectives set in the NHP. At the state level, health administration is concerned with customizing NHP as per the specific scenario in the state, consequently drafting the State Health Policy (SHP), and allocating resources to districts to meet the objectives set in the SHP under the ambit of NHP. At the district level, health administration is concerned with implementing the activities mentioned in SHP. Unlike their state level counterpart, district health administration does not customize health policies as per the specific scenario in the district.

Customization of NHP at state level means that the Indian policy makers acknowledge that there is significant variation in health indicators at state level. In other words, the Indian health policies are customized based on the belief that an individual's health can be attributed to his/ her belongingness to a state, and thus, health policies should be customized at state level. However, we noticed in the previous chapter (on page 11), that the variation in health indicators is larger at district level when compared with the corresponding variation at state level. Thus, it can be said that an individuals' health is also, and in fact more, attributed to his/ her belongingness to a district, than his/ her belongingness to a state. This is expected as districts are nested within states, and some districts in a state are likely to have a better (or a worse) health indicator than the state average.

So, when the extent of HI is higher at district level when compared with the state level, why is the NHP not customized at the district level? May be because the number of Indian districts (640) is much higher than the number of Indian states (35), and it may not be feasible to customize NHP for every district, as it will require more resources. In addition, if each district customizes its own health policy, it will be very difficult to monitor and align health interventions at district level with the overall health objectives of NHP.

However, the above reasons does not justify that the health policies should not be customized at district level, and the resulting HIs should be tolerated. By ignoring the need for customization at district level, we bear the risk of undermining the policy outcomes (ineffectiveness), and by customizing at the district level, we bear the risk of wastage of resources (inefficiency). The same can also be stated for the state level. Therefore, in order to achieve the objectives of NHP of India in an effective and resource efficient way, there is a need to identify the appropriate level at which the NHP should be customized. This leads us to the key problem addressed in this study:

In order to reduce inequities in health, at what administrative level/s should the concerned Indian health polices be customized, so that the resources are efficiently used, and outcomes are effectively obtained?

To answer this question, we should consider the following two aspects:

1.1. Multi-factorial setting

DoHs act in a multi-factorial way, and while assessing their relation with HIs, we cannot ignore the impact of other determinants. Based on common beliefs, medical facts, and other relevant historical evidence, we can expect that a DoH will result in a HI in a particular way. However, due to the role of other DoHs, many expected relations are likely to deviate from expectations in a particular setting. For example, the evidence of contradictory relations in India presented in the previous chapter (on page 17): between breastfeeding and malnourishment, between obesity and cardio-vascular diseases, and between smoking and respiratory diseases. Assessing the impact of a one determinant on one indicator[15] is relatively straightforward and can be done through simple

[15] As HIs are measured in the form of health indicators, we should look at the relation between a determinant and an indicator of health to describe how determinants of health lead to inequities in health.

regression technique. However, in a multi-factorial setting, like the one we have between determinants and indicators of health, a more complicated multiple regression will be needed. It is possible that the predictive ability of a DoH on an indicator of health can change with progression from simple to multiple regression, and such a change can be noticed in magnitude, direction or statistical significance of the relation.

In addition, health policies involve multiple aspects of public health, which needs to be addressed though interventions designed around multiple DoHs. Thus, while designing and implementing a health intervention, we need to consider the relation between a determinant and a health indicator. As each determinant will affect each health indicator in its own unique way, it is very likely that a determinant-indicator relation will vary from another determinant-indicator relation. Thus, for an effective and efficient implementation of health interventions, we need to identify which DoHs can predict[16] the indicator of health in a multi-factorial setting, and how does this vary from one determinant to another. This leads us to the first research hypothesis:

Hypothesis-1-null: The concerned determinant of health cannot predict the concerned indicator of health in a multi-factorial setting.

1.2. Multi-level setting

Judging that an individual's health is a matter of his/ her belongingness to a state or a district, by referring to the *range* of variations observed at each level is misleading (refer Table 2). This is because we do not know as to what extent the belongingness to a state or a district affects an individual's health. The higher range of variation at district level could be due to a few under-performing or over-performing districts. Moreover, the population in such outlier districts may not be big enough to significantly affect the overall state average. It is also unknown as to what extent districts vary within a state and across the states. It is possible that health indicators at district level vary extensively in one state, while they may not vary that much in another state. Thus, the actual impact of belongingness to a district may not be strong enough to justify the customization of NHP at district level, even when the range of variation is higher at district level.

[16] In this study, I have given preference to predictability rather than causality as describing causality requires a detailed analysis between each determinant and each indicator, and the data and scope of this study does not allow that. However, describing predictability will act as the first step towards identifying and describing causality.

Whether or not a health policy should be customized at state or district level (or both), is a matter of quantification of the extent to which an individual's belongingness to a level is significant enough to act a factor for influencing his/ her health. In other words, belongingness to a level may act as a variable in determining an individual's health, and we need to measure this to bring about the desired health outcomes in an effective and resource efficient way.

Thus, for an effective implementation of health interventions, we need to quantify as to what extent does a determinant-indicator relation varies at state and district levels, and how does that variation differ from one determinant-indicator relation to another. In effect, it is possible that a determinant-indicator relation may vary with a state, with a district, with both, or with none, and the second determinant-indicator relation may differ from the first determinant-indicator relation in this regard. Advanced statistical techniques like multi-level analysis can be used to quantify the impact of belongingness to a level. This leads us to the second research hypothesis:

Hypothesis-2-null: The concerned administrative level cannot predict the concerned indicator of health in a multi-level setting.

1.3. Defining the specific research objective and hypothesis

Indian health policies are now paying a strong attention on reducing HIs, and on targeting DoHs to attain the desired health outcomes, and the same can be noticed in the objective and strategies of 2015 draft NHP of India. The interventions in NHP are rolled out in the form of *disease specific* National Health Missions (NHM). For example: Revised National Tuberculosis Control Program, National Program for Control of Blindness, National Leprosy Eradication Program, National Iodine Deficiency Disorder Control Program, National Vector Borne Diseases Control Program, and so on. Within each NHM, interventions are carried out over multiple relevant DoHs. Because of this disease specific nature of implementation of NHP, I have focused on measures of *morbidity* as the relevant indicators of health in this study.

Morbidity has been defined as any departure, subjective or objective, from a state of physiological or psychological well-being (Porta 2014). In practice, morbidity encompasses disease, injury, and disability. The simplest measures of morbidity are *prevalence* and *incidence*.

The prevalence of a disease is the proportion of a population that are cases[17] at a point in time, while the incidence of a disease is the rate at which new cases occur in a population during a specified period (Coggon, Rose and Barker 1997). Incidence is a longitudinal phenomenon with observations recorded at minimum two different points in time. Prevalence is a cross-sectional phenomenon and involves observations at a single point in time.

It should be noted that diseases (morbidity) can be classified in many ways, for example, underlying cause (etiology), mechanism by which the disease develops (pathogenesis), symptoms of the disease, or the organs affected by the disease. However, these classifications are more relevant from a medical and therapeutic standpoint. In social and political sciences, diseases are most commonly classified as *acute* or *chronic*. While referring to a disease, the term *acute* refers to a disease with sudden onset, often brief and not necessarily clinically severe. Most infections are classified as acute diseases. The term chronic in this reference means a disease lasting a long time. Most lifestyle related diseases like Hypertension and Diabetes are classified as chronic diseases. It should be noted that the terms acute and chronic diseases are often used synonymously with *communicable* and *non-communicable diseases* respectively.

In India, the total burden of acute diseases is 58% higher than the global average, while the total burden of chronic diseases is 24% higher than the global average[18]. In addition, the range of variation in both acute and chronic diseases is higher among districts than among states (refer Table 2). Thus, I will test the research hypothesis for prevalence of acute and chronic diseases as the concerned health indicators, and state and district as the concerned administrative levels. Consequently, the research question and hypothesis can be stated as follows:

Research question: In order to reduce prevalence of acute/ chronic diseases in India, at what administrative level/s should the concerned health polices be customized (state/ district/ both/ none), so that the resources are efficiently used, and outcomes are effectively obtained?

Hypothesis-1-null ($H_{1\text{-}0}$): The concerned determinant of health cannot predict the prevalence of acute/ chronic diseases in a multi-factorial setting.

[17] The term case refers to an individual affected by the disease or a phenomenon under consideration.

[18] Data Source: World Health Statistics Report (WHO 2012)

Hypothesis-2-null ($H_{2\text{-}0}$): The concerned administrative level cannot predict the prevalence of acute/ chronic diseases in a multi-level setting.

It is expected that by testing the Hypothesis-1, we will be able to identify the determinants that can predict the prevalence of acute (or chronic) diseases in a multi-factorial environment. It is expected that by testing the Hypothesis-2, we will be able to identify the level of administration (state, district, both, or none) at which a health policy, targeted at regulating the prevalence of acute (or chronic) diseases, should be customized.

Chapter four: METHODOLOGY

I have started this chapter with a description of the data used in this study. This is followed by a systematic process for building the appropriate statistical model for the required data analysis.

1. Metadata

The Government of India, through its Open Government Data (OGD) platform, provides data from various government surveys to be freely used by public. The data available through the OGD platform is aggregated data, collected through household and individual surveys, and aggregated at the level of districts and states. Various government departments use the same data for designing and implementing their respective policies. For this study, I have sourced data from the OGD platform, and following are its details:

1.1. Data on indicators of health

In this study, the targeted indicators of health are prevalence of acute and chronic diseases, the data on which is obtained through the District Level Household and Facility Survey (DLHS). DLHS is a nationwide survey covering all the 640 districts, and covers various socio-economic, lifestyle and health related themes. The fourth round of DLHS (DLHS-4) was carried out during 2012-13, and it covered 350,000 households. I chose data from DLHS-4 as its period coincides with the period of the other source of data in this research. Through OGD platform, DLHS-4 data on prevalence of acute and chronic diseases could be obtained for 274 districts from 21 different states, and the same has been used for this study.

1.2. Data on determinants of health

The first hypothesis of this study revolves around testing the predictive ability of DoHs in a multi-factorial setting. As we know, there is no finite number of determinants. Thus, with every additional determinant that can be incorporated in the analysis, the results will be more valid. However, the determinants that should be included in the analysis should be theoretical linked to prevalence of acute and chronic diseases, and relevant to the Indian context. Moreover, the selected determinants should cover all the five categories of the Whitehead and Dahlgren's

classification of DoHs to give a comprehensive coverage to the analysis. It should also be noted that the data on all selected determinants should be coherent in its source and comparable in terms of its sample characteristics. Considering these factors, data on 17 different DoHs could be gathered for this study. The data was sourced from the OGD platform, but was originally gathered through two different surveys: DLHS-4 and the Census of India 2011.

Data on 8 of the 17 determinants was gathered from the DLHS-4, while data on the rest 9 determinants was gathered through the Census of India. Although both DLHS-4 and Census covered all districts in the country, DLHS-4 was based on a stratified sample of 350,000 households, while the Census covered every household and represented the whole population. From the OGD platform, data on these 17 determinants could be obtained for all the 640 districts. However, data on acute and chronic diseases could only be gathered for 274 districts, and thus, the data on the 17 determinants from the same 274 districts was used for analysis.

As mentioned before, the data used in this study is aggregated data, collected at individual level and aggregated at district level. It should be noted that by using aggregate data, there is an incorporated risk of ecological fallacy. However, aggregated data was used here because the scope of this study demands data from multiple states and districts, but does not allow time and resources for primary collection at individual level. Thus, the only available choice is the data from OGD platform, and it is a good option, as it is very credible, aggregated from a huge sample size, used by the government authorities themselves, and is free of cost.

1.3. Defining the variables

The determinants of health are the theoretical causes behind the recorded values of indicators of health. Thus, DoHs are the *independent variables* (IVs) and indicators of health are the *dependent variables* (DVs). In this study, the DVs ware prevalence of acute diseases (DV1) and prevalence of chronic diseases (DV2), while the IVs were various DoHs. As explained above, the data for analysis was obtained by gathering data on 2 dependent variables (DVs) and 17 independent variables (IVs), aggregated at the district level, and sourced from the OGD platform. All IVs and DVs were defined in the surveys through which data on them was originally collected, and a brief of those definitions is provided in table no. 4 and 5 below.

In addition, the data heads (DVs and IVs) were coded for simplicity in data analysis and presentation of findings. Coding was done as per the following rules: the 2 DVs were coded with initials DV and a number. Number 1 represented prevalence of acute diseases while number 2 represented prevalence of chronic diseases. In addition, the 17 IVs were coded with initials IV, an alphabet and a number. The alphabet represented the group that an IV belonged to (as per the Whitehead and Dahlgren classification[19]), while the number represented a simple sequence of concerned IV within its group. The names of variables, their definitions and codes are as follows:

Table 4: Names, codes and definition of dependent variables

Name	Code	Definition
Acute Diseases	DV1	Percentage of households that reported incidence of acute diseases during the last 15 days of the survey.
Chronic Diseases	DV2	Percentage of households that reported incidence of chronic diseases during the last 1 year of the survey.

Table 5: Names, codes and definition of independent variables

Name	**Code**	**Definition**
Sex Ratio	IVA1	The number of females per 100 males.
Under 15	IVA2	The percentage of population under the age of 15 years.
Tobacco Smokeless	IVB1	The percentage of population above the age of 15 years that consumes tobacco in smokeless forms.
Tobacco Smoking	IVB2	The percentage of population above the age of 15 years that smokes tobacco.
Alcohol	IVB3	The percentage of population above the age of 15 years that consumed alcohol.
Gender Inequality	IVC1	The difference in literacy and employment rate between males and females[20].

[19] (A) Constitutional Factors, (B) Individual Lifestyle Factors, (C) Community and Social Networks, (D) Living and Working Conditions, (E) Overall Socio-Economic and Cultural Factors.

[20] Measured in percentage points, and taken as average of the difference in literacy rate and employment rate.

Scheduled Population	IVC2	The percentage of population that belonged to either scheduled caste or scheduled tribe[21].
Electricity	IVD1	The percentage of electrified households.
Drinking Water	IVD2	The percentage of households with access to clean drinking water.
Toilet	IVD3	The percentage of households with toilets.
Clean Cooking Fuel	IVD4	The percentage of households with access to clean cooking fuel.
Literacy Rate	IVD5	The percentage of literate population above the age of 7 years.
Working Population	IVD6	The percentage of employed population.
Agricultural Workers	IVE1	The percentage of employed population engaged in agriculture.
Marginal Workers	IVE2	The percentage of working population engaged in marginal work[22].
Household Size	IVE3	The number of people in a household.
Marriage Age	IVE4	Percentage of currently married people, who were married below the legal age[23].

1.4. Defining IV-DV relations

This study involves 2 DVs (prevalence of acute and chronic diseases), and 17 IVs (DoHs), and I plan to test the relations between them, and will refer to them as the IV-DV relations. As each of the 17 IVs can be theoretically linked to both the DVs, there are 34 IV-DV relations to be studied. Most simply, an IV-DV relation can be broken into two aspects: *direction* of the relation and *power* of the relation.

The direction of a IV-DV relation can be *positive* or *negative*, based on the sign of the correlation or regression coefficient. A positive direction of relation means that the indicator changes in the same direction as the direction of change in the determinant. Thus, the indicator increases with increase in determinant, and decreases with decrease in determinant. Contrary to this, negative direction of relation means that the indicator changes in the direction opposite to the direction of

[21] Scheduled castes and tribes are socio-economically backward sections of the society.

[22] Marginal work is defined as employment for less than 6 months, during the preceding 1 year to the survey.

[23] Legal age for marriage for women is 18 years, and for men it is 21 years.

change in the determinant. Thus, indicator increases with decrease in the determinant, and decreases with increase in the determinant.

Power of a IV-DV relation refers to the magnitude of change in indicator with per unit change in determinant. Power of a IV-DV relation can be very low, low, moderate, high and very high, based on the value of the correlation[24] or regression coefficient.

2. Appropriate statistical techniques

The predicted value of a IV-DV relation can be explained in terms of the direction and power of the relation. To predict the direction and power of a IV-DV relation, bivariate correlation or simple regression can be used. The resultant correlation coefficient or regression coefficient will describe the one-on-one relation between a determinant and an indicator, without the influence of any other determinant.

The first hypothesis requires that IV-DV relations are tested in a multi-factorial setting. As all the 17 determinants act simultaneously to influence the prevalence of acute and chronic diseases in a direct or indirect way, multiple regression with all 17 determinants as independent variables and acute or chronic diseases as the dependent variable can be carried out. The resultant regression coefficient for an IV-DV relation will differ from the regression coefficient predicted earlier by simple regression. By analyzing the differences in the regression coefficient obtained by simple and multiple regression analysis, we will be able to identify the IVs that can affect the DV even in the multi-factorial setting. It should be noted that the multiple regression analysis should be adjusted for multi-collinearity and heteroscedasticity, before arriving at conclusions.

The second hypothesis requires that IV-DV relations are tested in a multi-level setting, so that the effect of belongingness to a state or a district can be measured. For this, multi-level analysis with random intercepts can be carried out, and the level which shows significant effect on the IV-DV relation can be justified as the level at which concerned policies should be customized.

[24] As per the classification given by Cohen and Holliday in 1982

These methods were applied to the gathered data, and the model for analysis was build up as described below.

2.1. Building the model for data analysis

The data analysis was a combination of *multiple regression* and *multi-level analysis* techniques. Multiple regression was used to test the first hypothesis, while multi-level analysis was used to test the second hypothesis. The models for both multiple regression and multi-level analyses were built in a stepwise manner. At each step, theoretically relevant and statistically valid decision rules were followed to reach conclusions. The Models were build-up in the similar manner, but separately, for both the DVs. Following are the details:

2.1.1. Building the regression model: testing the first hypothesis

The model for testing the first hypothesis was divided into two steps: simple regression analysis, and multiple regression analysis.

Simple regression analysis was used to test the one-on-one IV-DV relations. Each IV was regressed against each DV, so that one-on-one IV-DV relations were tested for statistical significance by referring to the p-value[25], and the direction and power of their relation was measured in the form of *regression coefficients.* However, the results from simple regression analysis will not be used for further statistical analysis, but will be referred to while discussing the findings of this study.

With multiple regression, we have the opportunity to test the effect of an IV on a DV, while controlling for other IVs. Thus, all IVs were regressed simultaneously against each of the two DVs, and new p-values and regressions coefficients were observed. However, as the IVs with high degree of correlation with each other may render the results invalid (multi-collinearity), *collinearity diagnostics* were included in the multiple regression analysis, and *problematic*

[25] If p-value (Sig.) is less than or equal to 0.05 for an IV, then the predictive effect of that IV on the concerned DV is statistically significant, and the direction and power of the corresponding regression coefficient can be used. Also, please note that the confidence interval is kept at 95% throughout the data analysis.

(highly correlated) IVs were removed[26]. The multiple regression was re-run without the problematic variable, and IV-DV relations with statistically significant p-value were identified.

In addition, scatter plots between *standardized residuals* and *standardized predicted values* were made for both DV1 and DV2, and cone shape typical of heteroscedastic data was seen in both the plots (refer to Annexure no.1). This means that the residuals in IVs increase with higher value of DV. As SPSS was used for data analysis, there was no option to correct for heteroscedasticity. Thus, results from multiple regression analysis corrected for multi-collinearity were used to identify the IVs that can predict the DVs in a multi-factorial setting. This was used to test the first hypothesis. The $H_{1\text{-}0}$ was rejected for IVs that showed statistically significant effect on DVs in above mentioned multiple regression analysis. The IVs for which the $H_{1\text{-}0}$ was rejected, were classified as *predictive IVs*, and were used for further analysis.

Figure 5: Regression analysis model: to test the first hypothesis

Identification of determinants of health (IVs) that can predict the indicators of health (DVs) in a multi-factorial setting: for testing the first hypothesis.		
Step 1: Simple Regression analysis		
Carried out between each DV and each IV (regressed one-by-one).	Carried out between each DV and each IV (regressed one-by-one).	Carried out between each DV and each IV (regressed one-by-one).
Step 2: Multiple Regression analysis		
Carried out between each DV and all IVs (regressed together).	Record the p-value for identification of statistically significant IV- DV relations in a multi-factorial setting.	Record the regression coefficients for explaining statistically significant one-on-one IV- DV relations.
Reject the Hypothesis-1-null for IVs with statistically significant p-value. These IVs can be stated as predictive IVs, as they can predict the DVs in a multi-factorial setting.		

[26] If the variation inflation factor (VIF) was greater than or equal to 5, then the concerned IV was classified as a problematic IV. A common practice is to exclude variables with VIF greater than or equal to 10. In this study, the VIF threshold level was kept low (at 5) as the number of variables is high (17) and sample size is not so big (274). Problematic IV can also be identified by comparing Pearson's r-value in a correlation matrix.

2.1.2. Building the multi-level analysis model: testing the second hypothesis

The predictive IVs identified in part 2 of the regression model, were used for *multi-level analysis* with districts as the base level and states as the second level. As districts are nested within states, by using multi-level analysis, we could test the extent to which variability in IV-DV relations observed at district level was attributed to belongingness to their respective states. If we were using individual level data, then we could have used individuals at the base level, districts at the second level, and states at the third level, to assess the extent to which variability at individual level was attributed to belongingness to their respective district and state. With the limitations of aggregated data, we could assess the variability due to belongingness to a state, and attributed the un-explained variation in the analysis to belongingness to a district. This will be clearer in the following text as I explain the model for multi-level level analysis done to test the second hypothesis.

The first step in building the multi-level analysis model was to define the *null model*. All predictive IVs for a DV were included in the null model for that DV. In addition, the level 2 variable (or the contextual variable) was included in the model. In this analysis, states were the contextual variable. The second step was to check *intra-class correlation* (ICC)[27] by introducing random intercepts in the null model without introducing fixed effects. This is done to assess whether or not the variation attributed to the contextual variable is high enough to justify building a multi-level analysis. The ICC was calculated using the following formula:

ICC = $\{s^2(b_n) / [s^2(b_n) + s^2(w_n)]\}$, where s^2 (w) is the pooled variance within states, and s^2 (b) is the variance of the trait between states in the null model with random intercepts (step 2).

In the third step, *fixed effects* were introduced in the null model to describe the IV-DV relations without accounting for variability due to belongingness to respective states. The p-values, regression coefficients and the *chi-square statistic* at this step were noted. The p-values[28] were

27 If intra-class correlation (ICC) was greater than or equal to 0.05, then a substantial amount of variability was attributed to the contextual variable (states).

28 The regression coefficients for each IV-DV relation in a multi-level model with fixed effect should theoretically show the same value (and sign) as the corresponding regression coefficient obtained through a multiple regression analysis with only those variables that tested negative for the $H_{1\text{-}0}$ (predictive IVs). However, a slight difference can

noted to verify if the predictive IVs hold their statistical significance when only the predictive IVs (and not all the 17 IVs) were included in the model. The regression coefficients were noted for reference while discussing findings of this study. The chi-square statistic will help in predicting the improvement in fit when compared with the similar statistic from the next step. Fourth step was to add *random intercepts* to the fixed effects, and the consequent chi-square statistic and p-values were noted, and *pseudo-R-square value* was calculated.

Chi-square statistics obtained in the third and fourth steps were compared to assess improvement in overall fit of the analysis[29]. An improvement in fit means that we should reject the $H_{2\text{-}0}$. As the second level (contextual) variable in this model was states, an improvement in fit will testify the hypothesis only towards states (and not districts). This means that if the fitness of model is improved with introduction of random intercepts in the fixed effects, belongingness to the contextual (state) level can predict the DV. In addition to the chi square statistic, another way to check the improvement in fit is through the R-square value. However, multi-level analysis does not provide an R-square value, and thus pseudo R-square values were computed. Apart from assessing goodness of fit, pseudo R-square value can also be used to measure the amount of variation in DV that has been explained in the model or the amount of variation that still needs explanation by further analysis[30]. It was calculated by using the following formula:

Pseudo r-square = $\{[s^2(b_f) - s^2(w_f)\} + \{s^2(b_n) - s^2(w_n)]\} / s^2(b_f) + s^2(b_n)$, where $s^2(w_f)$ is the pooled variance within states, and $s^2(b_f)$ is the variance of the trait between states in the fitted model (step 4), while $s^2(w_n)$ and $s^2(b_n)$ are the corresponding values in the null model with random intercepts (step 2).

Thus, both chi-square statistic and pseudo-R-square values were used to assess the predictive ability of belongingness to states. As the used data is aggregated at the level of districts, there

be noted if the multi-level analysis is carried with maximum likelihood setting. If we change the setting to restricted maximum likelihood, the regression coefficients will be same as those obtained through multiple regression.

[29]The change in degrees of freedom by inclusion of random intercepts was 1, and the corresponding significant value for change in -2LL value at 95% confidence interval was 3.84. Thus, if the change in chi-square statistic (-2LL) was greater than 3.84, then the later model had a better fit, and was thus the preferred model.

[30] Higher the pseudo R-square value, higher will be the amount of variability that is explained by the random intercept. The difference between the pseudo R-square value and 1 will be the amount of variation that still needs explanation.

was no option to assess the predictive ability of belongingness to districts. However, the p-values obtained in the fourth steps can be used as a proxy indicator to identify the IVs that have an effect on DV even after accounting for variation attributed to belongingness to the contextual variable. With some probability, this will testify the hypothesis towards districts, but only for IVs with statistically significant p-value in the fourth step of the multi-level analysis model described here.

Figure 6: Multi-level analysis model: to test the second hypothesis

<table>
<tr><td colspan="3">Identification of the level of aggregation that can predict the indicators of health (DVs) in a multi-level setting: for testing the second hypothesis.</td></tr>
<tr><td colspan="3">Step 1: Define the null model for multi-level analysis</td></tr>
<tr><td colspan="2">Carried out between each DV and all predictive IVs (taken together).</td><td>Include states as the contextual variable (level 2).</td></tr>
<tr><td colspan="3">Step 2: Introduce random intercepts in the null model.</td></tr>
<tr><td colspan="2">Calculate the intra-class correlation (ICC).</td><td>Assess the variation attributed to the contextual variable, to justify building multi-level model.</td></tr>
<tr><td colspan="3">Step 3: Introduce fixed effects in the null model.</td></tr>
<tr><td>Record the p-value to verify the predictive IVs.</td><td>Record the regression coefficients for reference in discussion.</td><td>Record the chi-square statistic for comparison in step 4.</td></tr>
<tr><td colspan="3">Step 4: Add random intercepts to the fixed effects obtained in step 3.</td></tr>
<tr><td>Record the chi-square statistic and compare it with the one obtained in step 3, to judge improvement in fit of the model.</td><td>Calculate the pseudo-R-square value, to assess improvement in fit and amount of explained variation.</td><td>Record the p-value to identify the IVs with unexplained variation.</td></tr>
<tr><td colspan="3">Reject the Hypothesis-2-null for the contextual variable (state level) if the model shows improvement in fit. Thus, belongingness to the state level can predict the DVs, and this will be valid for all IVs in the model.</td></tr>
<tr><td colspan="3">Reject the Hypothesis-2-null for the district level, only for the DVs with statistically significant p-value in step 4. Thus, belongingness to the district level can predict the DV, but this will be valid only for the concerned IVs.</td></tr>
</table>

Chapter five: FINDINGS

To test the research hypothesis, the gathered data was analyzed as per the model described in the previous chapter. IBM SPSS 23 software was used for this. In this chapter, I have presented the findings.

1. Findings from the multiple regression model: testing the first hypothesis

The regression coefficients obtained by simple regression analysis were statistically significant for 13 out of 17 IVs for DV1, and for 10 out of 17 IVs for DV2. Irrespective of the statistical significance, regression coefficient for all the 34 IV-DV relations were recorded for reference in discussion on findings of the study.

Table 6: Findings from the simple regression analysis

IV Code	IV Name	DV1		DV2	
		p-value	*Regression coefficient*	*p-value*	*Regression coefficient*
IVA1	Sex Ratio	0.006	-0.165	0.000	-0.265
IVA2	Under 15	0.000	0.254	0.641	-0.028
IVB1	Tobacco Smokeless	0.000	0.302	0.000	-0.269
IVB2	Tobacco Smoking	0.000	0.452	0.744	-0.020
IVB3	Alcohol	0.000	0.339	0.299	-0.063
IVC1	Gender Inequality	0.835	-0.013	0.000	0.480
IVC2	Scheduled Population	0.008	0.160	0.000	-0.294
IVD1	Electricity	0.000	-0.486	0.713	-0.022
IVD2	Drinking Water	0.000	-0.264	0.009	0.156
IVD3	Toilet	0.156	0.086	0.000	0.220
IVD4	Clean Cooking Fuel	0.000	-0.219	0.003	0.179
IVD5	Literacy Rate	0.000	-0.234	0.195	-0.078
IVD6	Working Population	0.006	-0.166	0.000	-0.389

IVE1	Agricultural Workers	0.175	0.082	0.019	-0.142
IVE2	Marginal Workers	0.089	0.103	0.674	-0.025
IVE3	Household Size	0.000	0.211	0.315	0.061
IVE4	Marriage Age	0.000	-0.219	0.000	-0.215

The regression coefficients obtained by multiple regression analysis were statistically significant for 7 out of 17 IVs for DV1 and for 9 out of 17 IVs for DV2. With inclusion of collinearity diagnostics in the multiple regression analysis, IVC1 (Gender Inequality) was identified as the problematic IV as it had a VIF 5.150. The VIF will be identical for both DVs as the concerned DVs and their data in their respective analysis is same. Finding from VIF was also crosschecked using bivariate correlation, and it was found that IVC1 had a Pearsons r-value greater than 0.4 with 7 other IVs. Thus, multiple regression was rerun without the problematic IV- IVC1. In the revised multiple regression analysis done with 16 IVs, all IVs had a VIF of less than 5. Thus, all problematic variables were removed, and the concerned multiple regression analysis can be used for further analysis.

Table 7: Findings from the multi-collinearity diagnostics

		DV1 and DV2	
		With all IVs	*Without the problematic IV- IVC1*
IVA1	Sex Ratio	2.482	1.538
IVA2	Under 15	1.690	1.672
IVB1	Tobacco Smokeless	3.065	3.050
IVB2	Tobacco Smoking	3.155	3.152
IVB3	Alcohol	2.312	2.195
IVC1	Gender Inequality	5.150	NA
IVC2	Scheduled Population	4.360	3.762
IVD1	Electricity	2.267	2.256
IVD2	Drinking Water	2.384	2.342
IVD3	Toilet	2.858	2.834
IVD4	Clean Cooking Fuel	2.810	2.805
IVD5	Literacy Rate	3.167	3.076

IVD6	Working Population	3.631	2.384
IVE1	Agricultural Workers	3.946	3.818
IVE2	Marginal Workers	1.919	1.918
IVE3	Household Size	2.908	2.498
IVE4	Marriage Age	2.246	2.243

In the revised multiple regression analysis, the regression coefficients were statistically significant for 8 out of 16 IVs for DV1 and for 10 out of 16 IVs for DV2. Thus, in a multi-factorial setting, the number of IVs that had a statistically significant effect on DVs reduced from 23 to 18. For these 18 IV-DV relations, $H_{1\text{-}0}$ could be rejected. This means that these IVs can predict the DV in a multi-factorial setting. The corresponding p-values and regression coefficients are as follows.

Table 8: Findings from the multiple regression analysis

IV Code	***IV Name***	**DV1**			**DV2**		
		p-value	*Regression coefficient*	*Reject $H_{1\text{-}0}$*	*p-value*	*Regression coefficient*	*Reject $H_{1\text{-}0}$*
IVA1	Sex Ratio	0.045	-0.111	Yes	0.000	-0.281	Yes
IVA2	Under 15	0.369	0.059		0.720	0.019	
IVB1	Tobacco Smokeless	0.270	0.024		0.004	-0.050	Yes
IVB2	Tobacco Smoking	0.000	0.199	Yes	0.001	0.142	Yes
IVB3	Alcohol	0.007	0.095	Yes	0.501	-0.019	
IVC1	Gender Inequality	NA	NA		NA	NA	
IVC2	Scheduled Population	0.000	-0.101	Yes	0.000	-0.088	Yes
IVD1	Electricity	0.000	-0.169	Yes	0.133	-0.056	
IVD2	Drinking Water	0.051	-0.068		0.155	-0.040	
IVD3	Toilet	0.003	0.073	Yes	0.000	0.070	Yes
IVD4	Clean Cooking Fuel	0.993	0.000		0.186	0.025	
IVD5	Literacy Rate	0.275	-0.059		0.003	-0.127	Yes
IVD6	Working Population	0.002	-0.209	Yes	0.000	-0.322	Yes
IVE1	Agricultural Workers	0.005	0.086	Yes	0.000	0.099	Yes

IVE2	Marginal Workers	0.432	0.035		0.017	0.086	Yes
IVE3	Household Size	.0940	-1.3190		.0084	-1.6702	Yes
IVE4	Marriage Age	.0851	-.4931		.2070	.2894	

Table 9: List of Predictive IVs

DV1			**DV2**		
S.no.	*IV Code*	*IV Name*	*S.no.*	*IV Code*	*IV Name*
1	IVA1	Sex Ratio	1	IVA1	Sex Ratio
2	IVB2	Tobacco Smoking	2	IVB1	Tobacco Smokeless
3	IVB3	Alcohol Consumption	3	IVB2	Tobacco Smoking
4	IVC2	Scheduled Population	4	IVC2	Scheduled Population
5	IVD1	Electricity	5	IVD3	Toilets
6	IVD3	Toilets	6	IVD5	Literacy Rate
7	IVD6	Working Population	7	IVD6	Working Population
8	IVE1	Agricultural Workers	8	IVE1	Agricultural Workers
			9	IVE2	Marginal Workers
			10	IVE3	Household Size

2. Findings from the multi-level analysis model: testing the second hypothesis

The 18 IV-DV relations for which $H_{1\text{-}0}$ was rejected (predictive IVs) were included in the multi-level analysis. This means that only those IVs, which predict DVs in a multi-factorial setting, were analyzed further to test the impact of levels. As the first step of multi-level analysis, the null model was established. The null model does not provide any finding *per se* but acts as the base for further steps. In the second step, with inclusion of random intercepts (without inclusion of fixed effects), the ICC for DV1 and DV2 were calculated using the formula mentioned in the previous chapter (on page 44), and they were found to be 0.66 and 0.60 respectively. Thus, the ICC for both the DVs was more than the threshold value of 0.05, and it could be concluded that a substantial amount of variability is attributed to the contextual variable and it makes sense to

build the *multi-level analysis model*. Observed variances, used for calculation of ICC are mentioned in the table no. 12 below.

With inclusion of fixed effects in the null model, as the third step, we could record the p- values, regression coefficients and chi-square statistic. The p-values for all the IV- DV relations were statistically significant and this confirmed that predictive IVs identified through the multiple regression analysis can predict the DVs even when only the predictive IVs are included in the model. The regression coefficients of all IV-DV relations were also recorded, and they will be used for reference in discussion of findings. The p-values and regression coefficients are mentioned in the table below. In addition, at this step, chi-square statistic was noted for comparison with the respective value from the next step.

Table 10: Findings from step 3 of multi-level analysis: fixed effects

DV1			**DV2**		
IV Code	*Regression Coefficient*	*p-value*	*IV Code*	*Regression Coefficient*	*p-value*
IVA1	-0.118	0.012	IVA1	-0.242	0.000
IVB2	0.239	0.000	IVB1	-0.035	0.027
IVB3	0.106	0.001	IVB2	0.147	0.000
IVC2	-0.096	0.000	IVC2	-0.086	0.000
IVD1	-0.226	0.000	IVD3	0.070	0.000
IVD3	0.048	0.015	IVD5	-0.117	0.003
IVD6	-0.197	0.000	IVD6	-0.296	0.000
IVE1	0.099	0.000	IVE1	0.079	0.000
			IVE2	.060	.045
			IVE3	-1.518	.008

With random intercepts in the fourth step, the chi-square statistic (-2LL) value for DV1 and DV2 changed by 110.177 and 88.990 respectively. Thus, for both DV1 and DV2, the change in -2LL value was higher than the required significant value of 3.84. Therefore, the model with random intercepts had a better fit, and it could be considered as the preferred model. This means that we could accept the $H_{2\text{-}0}$ for all the 18 IV-DV relations included in the multi-level model, but could

apply it only for the contextual variable *i.e.* states. This means that policies concerned with any of these 18 IV-DV relations should be customized at the state level.

Table 11: Findings from step 3, 4 of multi-level analysis: chi-square statistic

-2 log likelihood (-2LL) value		
	DV1	*DV2*
Step 3: Fixed effects only	1617.757	1482.392
Step 4: Fixed effects with random intercepts	1507.580	1393.402
Change in -2LL value	110.177	88.990
Improvement in fit	Yes	Yes
Reject $H_{2\text{-}0}$ (state level), applicable for all IVs	Yes	Yes

The pseudo R-square values for DV1 and DV2 were calculated using the formula mentioned in the previous chapter (on page 45), and they were found to be 0.413 and 0.250 respectively. This is satisfactorily high value. Thus, the improvement in fit as confirmed by the chi-square statistic was re-confirmed by pseudo-R-square value. Also, based on the pseudo-R-square value, it can be stated that the model has been able to explain 41% of state level variation in DV1, and 25% of state level variation in DV2. To explain more variation, further analysis could be done, for example inclusion of random slopes.

Table 12: Findings from step 2, 4 of the multi-level model: ICC, and pseudo-R-square value

Estimate of covariance parameters			
Step of the multi-level model	*Parameter*	*DV1*	*DV2*
Step 2: Random intercept only	Variance within states, $s^2(w_n)$	12.782	9.095
	Variance between states, $s^2(b_n)$	25.331	13.653
Step 4: Fixed effects with random intercepts	Variance within states, $s^2(w_f)$	11.698	8.110
	Variance between states, $s^2(b_f)$	16.396	9.294
Calculating ICC $\{s^2(b_n) / [s^2(b_n) + s^2(w_n)]\}$		[25.331 / (25.331 + 12.782)] = 0.66	[13.653 / (13.653 + 9.095)] = 0.60

Calculating pseudo-R-square $\{[s^2(b_f) - s^2(w_f)\} + \{s^2(b_n) - s^2(w_n)]\} / [s^2(b_f) + s^2(b_n)]$	$\{[(16.396 - 11.698) + (25.782 - 12.782)] / (16.396 + 25.331)\} = 0.413$	$\{[(9.292 - 8.110) + (13.653 - 9.095)] / (9.294 + 13.653)\} = 0.250$

As a proxy indicator of the predictive ability of belongingness to districts level on DVs, p-values for IV-DV relations in the fourth step of the multi-level analysis were recorded, and they were statistically significant for 5 IVs against DV1 and for another 4 IVs against DV2. For these 9 IV-DV relations, we could reject the $H_{2\text{-}0}$ and apply it to the district level. This means that policies concerned with any of these 9 IV-DV relations should be customized at the district level also.

Table 13: Findings from step 4 of the multi-level model: IVs with statistically significant effect

DV1			**DV2**		
IV	*p-value*	*Reject $H_{2\text{-}0}$ (district level)*	*IV*	*p-value*	*Reject $H_{2\text{-}0}$ (district level)*
IVA1	.980		IVA1	.264	
IVB2	.012	Yes	IVB1	.008	Yes
IVB3	.965		IVB2	.134	
IVC2	.010	Yes	IVC2	.144	
IVD1	.032	Yes	IVD3	.008	Yes
IVD3	.004	Yes	IVD5	.532	
IVD6	.122		IVD6	.049	Yes
IVE1	.001	Yes	IVE1	.066	
			IVE2	.456	
			IVE3	.002	Yes

Chapter six: DISCUSSION

Majority of this chapter includes a detailed discussion on findings of this study, with an emphasis on their possible applications in health and other public policies. Towards the end of this chapter, I have briefly mentioned the concluding remarks along with limitations and scope of this study.

In this study, I tried to assess the predictive ability of 17 different DoHs on prevalence of acute and chronic diseases in a multi-factorial and a multi-level setting, and in context of present day public health policy administration in India. The NHP of India strives to reduce the prevalence of acute and chronic diseases, as the total burden of acute diseases in India is 58% higher than the global average, while the total burden of chronic diseases is 24% higher than the global average[31]. For this, various disease specific NHMs are rolled out under the ambit of NHP, with interventions targeted at relevant DoHs.

However, to execute the NHP and NHMs in an effective and resource efficient way, only those determinants, which can predict the prevalence of diseases, should be targeted. Thus, we must identify the relevant determinants, and this assessment should be done in a multi-factorial and multi-level setting, the reasons for which have been explained in chapter 3 (on page 32-35). Accordingly, the research objectives were defined, hypothesis were developed, and data analysis was completed.

The findings of the analysis were presented in the previous chapter, and the same are discussed in detail in this chapter. As the analysis was carried out separately for acute diseases (DV1) and chronic diseases (DV2), and further divided as per the multi-factorial (Hypothesis 1) and multi-level (Hypothesis 2) context of the study, the discussion in this chapter is organized in a similar manner.

[31] Data Source: World Health Statistics Report (WHO 2012)

1. Discussion on Acute Diseases

1.1. Determinants of health that can predict prevalence of acute diseases in a multi-factorial setting.

The regression coefficients obtained by simple regression analysis were statistically significant for 13 DoHs. This means that these 13 determinants can predict the prevalence of acute diseases, without influence of other factors. On introducing the influence of other factors, regression coefficients obtained through the multiple regression were statistically significant for 8 determinants. Thus, these 8 determinants can predict the prevalence of acute diseases in a multi-factorial setting, and theoretically, if targeted in the policies concerned with reducing the prevalence of acute diseases in India, these determinants are likely to produce desired results in more effective and efficient way. They are: Sex Ratio, Tobacco Smoking, Alcohol Consumption, Scheduled Population, Electricity, Toilets, Working Population and Agricultural Workers.

The statistical significance, measured by the p-value, merely confirms that these determinants can predict the prevalence of acute diseases. However, their actual impact can be assessed by reading the direction and magnitude of their respective regression coefficients, obtained in the third step of the multi-level analysis (fixed effects without random intercept). Refer to table no. 10 for the concerned regression coefficients. The same are also mentioned and discussed in the text below.

While discussing the regression coefficients obtained in this study, it is important that the conventional relation of the discussed determinants with prevalence of acute diseases are also taken into consideration, in order to identify how and why the findings in this study differ from the conventional expectations. To explain the findings, especially the unexpected findings, it is important to refer back to underlying mechanisms. With these in mind, I will now discuss the determinants, which were identified to have predictive ability on prevalence of acute diseases in a multi-factorial setting.

1.1.1. Sex Ratio

The determinant sex ratio was defined in the data as the number of females per 100 males. The regression coefficient for the determinant *sex ratio* was found to be -0.118. This means that with each extra female per 100 males, prevalence of acute diseases decreases by 0.118 units. It should

be noted that the constant in the concerned multiple regression was 41.15. In addition, prevalence of acute diseases in the data represents the percentage of households that reported incidence of acute diseases during the last 15 days of the survey. Thus, it can be said that in a population with a sex ratio of 90, and value of all other IVs as zero, 30.53% households reported incidence of acute diseases (41.15 minus 10.62). Moreover, if the sex ratio in that population is changed from 90 to 91, *ceteris paribus,* the prevalence of acute diseases will reduce by 0.118 to 30.41%. Therefore, if the sex ratio increases, the prevalence of acute diseases is likely to reduce in India. However, there is no direct causal relation between sex ratio and acute diseases, because sex ratio in itself does not cause diseases. Therefore, we cannot state weather the observed relation between sex ratio and prevalence of acute diseases is expected or unexpected. However, we can explain this observation, and for this we need to dig into the underlying mechanisms, and the same are discussed below.

The observed relation between sex ratio and prevalence of acute diseases is not as simple and straightforward as it may sound, mainly because both sex ratio and prevalence of acute diseases are macro-level phenomenon, measured at individual level but aggregated at macro-level. Thus, the actual effect of aggregate change in sex ratio can be understood by drilling it down to the individuals (macro to micro), understanding its diffusion among individuals in the target group (micro to micro), and aggregating the prevalence of acute diseases back to the macro level (micro to macro). Refer back to Coleman's boat discussed in chapter 2 on page 28. Numerous factors will govern this macro-micro-micro-macro transition, and they can be understood in detail as per the context to explain how changing sex ratio can change the prevalence of acute diseases.

For example, in India, sex ratio is skewed in favor of men, *i.e.* the number of women is less than the number of men. Some of the common reasons attributed to this are cultural preferences for male child, and emigration of males for work. However, there has been a positive trend in this, and the sex ratio has increases from 933 (females per thousand males) in 2001 to 940 in 2011[32]. In the same period, the percentage of deaths attributed to acute (communicable) diseases reduced from 40% to 28% in India[33]. One possible explanation could be as follows: due to socio-

[32] Data Source: Census of India (Ministry of Home Affairs, Government of India 2011d)

[33] Data Source: World Bank Open Data (World Bank 2016b)

economic development of India during the reference period, preference for male child and emigration has reduced, consequent to which sex ratio has increased. The same socio-economic development has also led to healthier living conditions and better medical facilities, both of which helped in reducing the prevalence of acute diseases. Thus, there may not be a direct causal link between sex ratio and prevalence of acute diseases, but both may be influenced by common third factors like socio-economic development. This is an example of how underlying mechanisms are important in explaining the predictive ability of DoHs on indicators of health. The data and scope of this study are insufficient to investigate this further and reach conclusive arguments, but opens the scope for another study. The scope of this study allows us to state that sex ratio has an inverse relation with prevalence of acute diseases. Thus, policies that aim at increasing the sex ratio are likely to reduce the prevalence of acute diseases indirectly.

1.1.2. Tobacco Smoking

The determinant tobacco smoking was defined in the data as the percentage of population above the age of 15 years that smokes tobacco. The regression coefficient for the determinant tobacco smoking was found to be 0.239. This means that if the percentage of smokers in the population decreases by one unit, prevalence of acute diseases decreases by 0.239 units. Thus, it can be said that in a population where 20% people smokes tobacco, and value of all other DVs as zero, 36.37% households reported incidence of acute diseases (41.15 minus 4.78). In addition, if the percentage of smokers in that population is changed from 20 to 19, *ceteris paribus*, the prevalence of acute diseases will reduce by 0.239 to 36.13%. So, in order to reduce the prevalence of acute diseases, public health policies in India should target at reducing smoking tobacco. This is an expected relation as smoking tobacco increases an individual's susceptibility to many diseases like acute respiratory infections, asthma, bronchitis, lung cancer, emphysema, pneumonia, etc. Although most of these are chronic diseases, acute exacerbations and attacks can be aggravated due to smoking.

Since tobacco smoking is directly (as well as indirectly) linked to acute diseases, the role of third factors in describing the predictive relation between tobacco smoking and prevalence of acute diseases cannot be underestimated, and we need to investigate into the underlying mechanisms for explaining this. For example, there are different forms of tobacco smoking in India: cigarettes, beedis, hookah, and reverse smoking. Use of cigarettes is more common in urban

areas as they are expensive and urban areas have higher per capita income. Use of beedis, hookah and reverse smoking is more common in rural areas as they are inexpensive and rural areas have lower per capita income. Rural areas typically show a higher incidence of acute diseases, when compared with urban areas. In addition, cigarette smoking is medically less harmful that the other three forms of smoking. Thus, it possible that the predictive ability of tobacco smoking on prevalence of acute diseases in a combination of direct medical effects and common underlying factors like socio-economic conditions of the area. Briefly, although there is evidence that policies targeted at reducing prevalence of acute diseases can include interventions for reducing tobacco smoking, a deeper understanding of local socio-economic conditions may be needed for better explanations.

1.1.3. Alcohol Consumption

The determinant alcohol consumption was defined in the data as the percentage of population above the age of 15 years that consumed alcohol. The regression coefficient for the determinant alcohol consumption was found to be 0.106. This means that if the percentage of people consuming alcohol decreases by one unit, prevalence of acute diseases also decreases by 0.106 units. So, in order to reduce the prevalence of acute diseases, public health policies in India should target at reducing alcohol consumption. Similar to tobacco consumption, alcohol consumption can be linked directly to acute diseases, like gastro-intestinal problems, and indirectly to acute exaggerations of chronic diseases like liver disorders, diabetes and cardiovascular problems. In addition, in India, more expensive and relatively less unhealthy distilled alcohol is consumed more in richer urban areas, while less expensive and relatively unhealthier country liquor is consumed more in poorer rural areas. Thus, although there is evidence that policies targeted at reducing prevalence of acute diseases can include interventions for reducing alcohol consumption, because of the role of common third factors, a detailed understanding of local socio-economic conditions may be needed for better explanations.

1.1.4. Scheduled Population

The determinant scheduled population was defined in the data as the percentage of population that belonged to either scheduled caste or scheduled tribe. Scheduled caste and tribes are socio-economically backward sections of Indian society. The regression coefficient for the determinant scheduled population was found to be -0.096. This means that in an area with higher proportion

of scheduled population, prevalence of acute diseases is slightly lower. This is contrary to the expectation that socio-economically backward population will have a higher prevalence of acute diseases. However, the simple regression between the determinant scheduled population and prevalence of acute diseases showed a statistically significant regression coefficient of 0.160. This means that the expected relation between scheduled population and acute diseases holds contradictory relevance in a multi-factorial environment. To assess the reversal of this relation, a more specific analysis with different variables can be conducted.

Similar to the determinant sex ratio discussed above, the determinant scheduled population is a macro-level demographic variable that does not have any direct causal relation with prevalence of acute diseases. It is due to other common factors like poverty, lifestyle, and working conditions that a correlation is observed between scheduled population and prevalence of acute diseases. For example[34], only 2.4% of urban population belongs to scheduled tribes, while 10.4% of rural population belongs to scheduled tribes. Thus, we can say that the data used and analysis done in this study is inconclusive to establish a predictive or causal relation between scheduled population and prevalence of acute diseases.

1.1.5. Electricity

The determinant electricity was defined in the data as the percentage of electrified households. The regression coefficient for the determinant electricity was -0.226. This means that for every one-unit increase in proportion of households with electricity, prevalence of acute diseases decreases by 0.226 units. Although absence of electrification is not a direct cause of acute diseases, electrified households are better equipped to deal with unhealthy living conditions and add comfort to life. Approximately 8% villages in India are not electrified even today[35]. As electrification is not a part of health policies, relevant policies that aim at electrifying more and more households, are likely to reduce the prevalence of acute diseases as an indirect long-term effect.

1.1.6. Toilets

[34] Data Source: Census of India (Ministry of Home Affairs, Government of India 2011c)

[35] Data Source: DLHS 4 (Planning Commission 2014)

The determinant toilets was defined in the data as the percentage of households with toilets. The regression coefficient for the determinant *toilets* was 0.048. This means that for every one-unit increase in proportion of households with toilets, prevalence of acute diseases increases by 0.048 units. However, it is a well-established fact that access to safe sanitation will reduce acute (infectious) diseases. However, the findings in this study contradict this. In addition, the simple regression between the determinant toilets and prevalence of acute diseases did not show a statistically significant result. The evidence in this study is not enough to refute the common belief that better sanitation reduces acute diseases. Thus, we can say that the concerned relation is inconclusive in this study. However, it is important to assess the reasons for the contradictory relation observed here. Simplest reason could be that the data used here was exceptional. However, a more practical reason is to attribute this contradictory relation to influence of other related variables. Following is an example:

The status of sanitation is very poor in India. In 2011, only 46.9% of India's total population, and only 30.7% of India's rural population, used improved sanitation[36]. UNICEF India states that diarrhea and respiratory infections are the number one cause for child deaths in India. Hand washing with soap, particularly after contact with excreta, can reduce diarrhea by over 40% and respiratory infections by 30%[37]. Thus, it is possible that proper hand washing is more important than using a toilet for controlling acute diseases. As interventions to build toilets are often accompanied by other aspects of sanitation like hand washing, there is a possibility that the relation between toilets and acute diseases is governed by other factors. The data in this study was not sufficient to establish this, but the observed contradictory relation provides a good reason to investigate it further.

1.1.7. Working population

The determinant working population was defined in the data as the percentage of employed population. The regression coefficient for the determinant *working population* was -0.197. This means that for every one-unit increase in proportion of working population, prevalence of acute diseases decreases by 0.197 units. This is an expected relation. However, similar to the

[36] Data Source: Census of India (Ministry of Home Affairs, Government of India 2012)

[37] Data Source: UNICEF India (UNICEF India 2016)

determinant sex ratio and scheduled population discussed above, the determinant working population is a macro-level variable that does not have any direct causal relation with prevalence of acute diseases. It is due to other common factors like poverty, lifestyle, and working conditions that a correlation is observed between working population and prevalence of acute diseases. Thus, we can say that to establish causal links between working population and prevalence of acute diseases, we need to study the underlying mechanisms. However, employment policies aimed at increasing employment rate will indirectly help in reducing the prevalence of acute diseases in the long term.

1.1.8. Agricultural workers

The determinant agricultural workers was defined in the data as the percentage of employed population engaged in agriculture. The regression coefficient for the determinant *agricultural workers* was 0.099. This means that for every one-unit increase in proportion of working population engaged in agriculture, prevalence of acute diseases increases by 0.099 units. This is an expected relation as in India, agriculture is the primary occupation in rural areas, and most rural areas are characterized by under-developed living and working conditions. In addition, Indian agricultural practices are primitive when compared with global standards, and are labor intensive. As the predictive ability of agricultural workers as observed here is as per the conventional expectations, we can say that employment policies aimed at reducing the proportion of agricultural workers in the country, while increasing the overall employment rate, will indirectly lead to a decrease in prevalence of acute diseases in long term.

The discussion on predictive ability of above-mentioned 8 determinants, on prevalence of acute diseases in a multi-factorial setting, can be summarized as follows: a determinant is expected to influence the prevalence of acute diseases based on popular belief and historical evidence. This expected relation can be positive or negative, and was explained in chapter 4 (on page 40). However, the observed findings in this study could be in accord with that expectation or not. Thus, the observed predictive ability of the determinant can be classified as expected or unexpected. In addition, the relation between the determinant and prevalence of acute diseases can be direct when the determinant directly causes acute diseases, or indirect when the determinant is simply correlated to acute diseases. In case of unexpected and/ or indirect relations, it is important that we investigate the underlying mechanisms to explain the reasons for

deviation from expectation and correlations. Based on these, the discussion on related findings can be summarized as follows:

Table 14: Important aspects of the discussion on prevalence of acute diseases

Determinant	Expected relation	p-value	Regression Coefficient	Observed relation	Causality
Sex Ratio	Negative	0.012	-0.118	Expected	Indirect
Tobacco Smoking	Positive	0.000	0.239	Expected	Direct
Alcohol Consumption	Positive	0.001	0.106	Expected	Direct
Scheduled Population	Positive	0.000	-0.096	Unexpected	Indirect
Electricity	Negative	0.000	-0.226	Expected	Indirect
Toilets	Negative	0.015	0.048	Unexpected	Direct
Working Population	Negative	0.000	-0.197	Expected	Indirect
Agricultural Workers	Positive	0.000	0.099	Expected	Indirect

The NHP of India can intervene at the level of determinants that have a direct causal relation with acute diseases. However, determinant with indirect causal relation are intervened under other (non-health) policies, for example, employment policies, rural development policies, agriculture policies, etc. To assess the impact of any policy (health or otherwise) on acute diseases, we can refer to the findings in this study. However, the determinants with unexpected and/ or indirect relations with acute diseases, as mentioned in the table above, should be carefully looked upon with emphasis on their underlying mechanisms.

1.2. Administrative level that can predict prevalence of acute diseases in a multi-level setting.

The above discussion was based on the finding that the concerned 8 determinants showed statistically significant results in a multi-factorial setting. However, the effect of level of aggregation was also analyzed as an additional factor for prediction of prevalence of acute diseases. As the used data had the administrative level- districts as the base level and the administrative level- states as the second level, predictive ability of belongingness to states on prevalence of acute diseases could be analyzed using the multi-level analysis. In the multi-level

analysis, the ICC was observed to be higher than the threshold value and the introduction of random intercepts improved the fit of the model (verified by the chi-square statistic and pseudo-R-square value). Thus, we can say that belongingness to states can predict the prevalence of acute diseases.

However, 3 out of the 8 relations under consideration have direct causal relation with acute diseases, and they can be targeted for intervention in health policies. Thus, the concerned health policies should be customized at state level. For the remaining 5 determinant that have an indirect causal relation with acute diseases, we should assess the indirect impact of their concerned policies on prevalence of acute diseases in a multi-level setting while encompassing the role of belongingness to states.

It should also be noted that even after inclusion of random intercepts, there was a considerable proportion of unexplained variation in the model, and 5 out of the concerned 8 determinants still showed predictive ability on acute diseases as they demonstrated statistically significant regression coefficients. As the used data was aggregated at district level, we cannot analyze the actual impact of belongingness to districts through multi-level analysis, however, the 5 determinants which demonstrated statistically significant regression coefficients are very likely to account for the unexplained variation in the model. They are Tobacco Smoking, Scheduled Population, Electricity, Toilets, and Agricultural Workers.

2 of the above-mentioned 5 determinants, tobacco smoking and toilets, have direct causal relations with acute diseases. Thus, with some probability, it can be said that the Indian public health policies concerned with reduction of prevalence of acute diseases by intervening at the level of these 2 determinants, should be customized at district level as well. For the remaining 3 determinant that have an indirect causal relation with acute diseases, we should assess the indirect impact of their concerned policies on prevalence of acute diseases in a multi-level setting while encompassing the role of belongingness to districts as well. Following is the summary of this discussion:

Table 15: Summary of the discussion on prevalence of acute diseases

	Concerned policies	Customization of the policy for intervention	Assessment of indirect impact of the policy interventions
Sex Ratio	Other policies		At state level
Tobacco Smoking	Health policies	At state and district levels	
Alcohol Consumption	Health policies	At state level	
Scheduled Population	Other policies		At state and district levels
Electricity	Other policies		At state and district levels
Toilets	Health policies	At state and district levels	
Working Population	Other policies		At state level
Agricultural Workers	Other policies		At state and district levels

2. Discussion on Chronic Diseases

The findings related to prevalence of chronic diseases can be discussed in a manner similar to the discussion on prevalence of acute diseases. As the governing concepts will remain the same, I will keep the discussion on chronic diseases specific and short.

2.1. Determinants of health that can predict prevalence of chronic diseases in a multi-factorial setting.

The regression coefficients obtained by multiple regression analysis were statistically significant for 10 DoHs. Thus, these 10 determinants can predict the prevalence of chronic diseases in a multi-factorial setting, and theoretically, if targeted in the policies concerned with reducing the prevalence of chronic diseases in India, these determinants are likely to produce desired results in more effective and efficient way. They are: Sex Ratio, Tobacco Smokeless, Tobacco Smoking, Scheduled Population, Toilets, Literacy Rate, Working Population, Agricultural Workers, Marginal Workers, and Household Size. The concerned p-values, regression coefficients, expected relation (positive or negative), observed relation (expected or unexpected), and causality (direct or indirect) are mentioned in the table below:

Table 16: Important aspects of the discussion on prevalence of chronic diseases

Determinant	Expected relation	p-value	Regression Coefficient	Observed relation	Causality
Sex Ratio	Negative	0.000	-0.242	Expected	Indirect
Tobacco Smokeless	Positive	0.027	-0.035	Unexpected	Direct
Tobacco Smoking	Positive	0.000	0.147	Expected	Direct
Scheduled Population	Negative	0.000	-0.086	Expected	Indirect
Toilets	Negative	0.000	0.070	Unexpected	Direct
Literacy Rate	Negative	0.003	-0.117	Expected	Indirect
Working Population	Negative	0.000	-0.296	Expected	Indirect
Agricultural Workers	Negative	0.000	0.079	Unexpected	Indirect
Marginal Workers	Negative	.045	.060	Unexpected	Indirect
Household Size	Negative	.008	-1.518	Expected	Indirect

2.1.1. Sex Ratio

The regression coefficient for the determinant *sex ratio* was found to be -0.242. This means that with each extra female per 100 males, prevalence of chronic diseases decreases by 0.242 units. It should be noted that the constant in the concerned multiple regression was 51.70. Also, prevalence of chronic diseases in the data represents the percentage of households that reported incidence of chronic diseases during the last 1 year of the survey. Thus, it can be said that in a population with a sex ratio of 90, and value of all other DVs as zero, 29.92% households reported incidence of chronic diseases (51.70 minus 21.78). In addition, if the sex ratio in that population is changed from 90 to 91, *ceteris paribus,* the prevalence of chronic diseases will reduce by 0.118 to 29.68%. However, both sex ratio and prevalence of chronic disease are macro level variables, and there is no direct causal link between them. Thus, it is important that the correlation observed between them be explained through underlying mechanisms.

2.1.2. Tobacco Smokeless

The determinant tobacco smokeless was defined in the data as the percentage of population above the age of 15 years that consumes tobacco in smokeless forms. The regression coefficient

for the determinant tobacco smoking was found to be -0.035. This means that if the percentage of smokers in the population decreases by one unit, prevalence of chronic diseases increases by 0.036 units. However, it is expected that consumption of smokeless tobacco will give rise various chronic diseases, but the analysis shows that otherwise. The simple regression between the determinant tobacco smokeless and prevalence of chronic diseases showed a regression coefficient of -0.269. This means that the observed relation between tobacco smokeless and chronic diseases is contradictory to the expected relation in both one-on-one scenario and multi-factorial scenarios. To assess this unexpected relation, a more specific analysis with different variables can be conducted.

2.1.3. Tobacco Smoking

The regression coefficient for the determinant tobacco smoking was found to be 0.147. This means that if the percentage of smokers in the population decreases by one unit, prevalence of chronic diseases also decreases by 0.147 units. The observed relation is similar to the expected relation. Thus, it can be said that in a population where 20% people smoke tobacco, and value of all other DVs as zero, 48.76% households reported incidence of chronic diseases (51.70 minus 2.94). In addition, if the percentage of smokers in that population is changed from 20 to 19, *ceteris paribus,* the prevalence of chronic diseases will reduce by 0.147 to 48.61%. As there is direct causal relation between smoking and chronic diseases, in order to reduce the prevalence of chronic diseases, public health policies in India should target at reducing smoking tobacco.

2.1.4. Scheduled Population

The regression coefficient for the determinant scheduled population was found to be -0.086. This means that in an area with higher proportion of scheduled population, prevalence of chronic diseases is slightly lower. This is expected as lifestyle related chronic diseases like diabetes and hypertension, are generally less prevalent in population belonging to lower socio-economic segments of the society. The observed relation is in accord with the conventional expectations. As both scheduled population and prevalence of chronic diseases are macro level variables, and as there is no direct causal relation between them, the observed correlation can be assessed by explaining the underlying mechanisms and breaking them down in the macro-micro-micro-macro transition. Also important is the role of other common factors like socio-economic conditions and lifestyle factors. For example, the development policies aimed at integration and

upliftment of scheduled population may result in increase in prevalence of chronic diseases as more people may adopt lifestyle attributed to chronic diseases.

2.1.5. Toilets

The regression coefficient for the determinant toilets was 0.070. This means that for every one-unit increase in proportion of households with toilets, prevalence of chronic diseases increases by 0.048 units. However, it is expected that access to safe sanitation will reduce diseases, but most of these diseases are acute in nature with some of them having chronic progression. On the other hand, urban areas have a higher prevalence of toilets, but they also have a higher prevalence of lifestyle related chronic diseases. Thus, we cannot really say for sure what should be the expected relation, and whether or not the observed relation contradicts it. In absence of a detailed analysis, I would like to give the benefit of doubt to the direct relation between toilets and chronic diseases *i.e.* they are inversely related. The simple regression between the determinant toilets and prevalence of chronic diseases showed a statistically significant regression coefficient of 0.220. This means that the observed relation between toilets and chronic diseases is contradictory to the expected relation in both one-on-one scenario and multi-factorial scenarios. As this is not enough evidence to refute the conventional relation between toilets and chronic diseases, finding in this study provides some starting points to undertake a detailed analysis for assessing the reason behind this.

2.1.6. Literacy Rate

The determinant literacy rate was defined in the data as the percentage of literate population above the age of 7 years. The concerned regression coefficient was found to be -0.117. This means that if the literacy rate in the population increases by one unit, prevalence of chronic diseases decreases by 0.117 units. This is an expected relation but the causality between literacy rate and prevalence of chronic diseases is indirect. Thus, education policies aimed at improving the literacy rate in India will have an indirect effect of reducing the prevalence of chronic diseases in a long term.

2.1.7. Working Population

The concerned regression coefficient was -0.296. This means that for every one-unit increase in proportion of working population, prevalence of chronic diseases decreases by 0.296 units. This

is an expected relation but the causality between working population and prevalence of chronic diseases is indirect. Thus, employment policies aimed at improving the employment rate in India will have an indirect effect of reducing the prevalence of chronic diseases in a long term.

2.1.8. Agricultural Workers

The concerned regression coefficient was 0.079. This means that for every one-unit increase in proportion of working population engaged in agriculture, prevalence of chronic diseases increases by 0.099 units. As the relation between proportion of agricultural workforce and prevalence of chronic diseases is indirect, we cannot really say for sure what should be the direction of their expected relation. To assess this further, we should include the impact of third factors. For example, urban areas have a lower proportion of agricultural workforce but a higher prevalence of chronic diseases, when compared with the respective values from rural areas. Due to the limited scope of this study, I would like to give the benefit of doubt to negative relation between agricultural workforce and chronic diseases. Thus, the observed relation is contradictory.

2.1.9. Marginal Workers

The determinant marginal workers was defined in the data as the percentage of working population engaged in marginal work. Marginal work is defined as employment for less than 6 months, during the preceding 1 year to the survey. The concerned regression coefficient was found to be 0.060. This means that if the proportion of marginal workers in the population increases by one unit, prevalence of chronic diseases also increases by 0.060 units. However, it is more likely that the proportion of marginal workers will be lower in economically developed areas, and such areas tend to have a higher proportion of chronic diseases. However, this is not conclusive and need a detailed analysis with focus on underlying mechanisms. For this study, I will give the benefit of doubt to the negative direction of relation between marginal workers and chronic disease, and will state that the observed relation is unexpected.

2.1.10. Household Size

The determinant household size was defined in the data as the number of people in a household. The concerned regression coefficient was found to be -1.518. This means that households with more people are less likely to get chronic diseases. The average household size in rural parts of

India is 4.9, and in urban parts, it is 4.6[38]. As the prevalence of chronic diseases in higher in urban areas, it is expected that there is a negative relation between household size and chronic diseases. Therefore, the observed relation is as expected. However, as the causality between household sized and chronic diseases is indirect, a more specific and detailed analysis on underlying mechanisms is deemed.

The NHP of India can intervene at the level of determinants that have a direct causal relation with chronic diseases. However, determinant with indirect causal relation are intervened under other (non-health) policies, for example, employment policies, rural development policies, agriculture policies, etc. To assess the impact of any policy (health or otherwise) on chronic diseases, we can refer to the findings in this study. However, the determinants with unexpected and/ or indirect relations with chronic diseases, as mentioned in the table no. 16 above, should be carefully looked upon with emphasis on their underlying mechanisms.

2.2. Administrative level that can predict prevalence of chronic diseases in a multi-level setting.

The above discussion was based on the finding that the concerned 10 determinants showed statistically significant results in a multi-factorial setting. However, the effect of level of aggregation was also analyzed as an additional factor for prediction of prevalence of chronic diseases. In the concerned multi-level analysis, the ICC was observed to be higher than the threshold value and the introduction of random intercepts improved the fit of the model (verified by the chi-square statistic and pseudo-R-square value). Thus, we can say that belongingness to states can predict the prevalence of chronic diseases.

However, 3 out of the 10 relations under consideration have direct causal relation with chronic diseases, and they can be targeted for intervention in health policies. Thus, the concerned health policies should be customized at state level. For the remaining 7 determinant that have an indirect causal relation with chronic diseases, we should assess the indirect impact of their concerned policies on prevalence of chronic diseases in a multi-level setting while encompassing the role of belongingness to states.

[38] Data Source: Census of India (Ministry of Home Affairs, Government of India 2011b)

It should also be noted that even after inclusion of random intercepts, there was a considerable proportion of unexplained variation in the model, and 4 out of the concerned 8 determinants still showed predictive ability on chronic diseases as they demonstrated statistically significant regression coefficients. They are Tobacco Smokeless, Toilets, Working Population, and Household Size.

2 of the above-mentioned 4 determinants, tobacco smokeless and toilets, have direct causal relations with chronic diseases. Thus, with some probability, it can be said that the Indian public health policies concerned with reduction of prevalence of chronic diseases by intervening at the level of these 2 determinants, should be customized at district level as well. For the remaining 2 determinant that have an indirect causal relation with chronic diseases, we should assess the indirect impact of their concerned policies on prevalence of chronic diseases in a multi-level setting while encompassing the role of belongingness to districts as well. Following is the summary of this discussion:

Table 17: Summary of the discussion on prevalence of chronic diseases

	Concerned policies	**Customization of the policy for intervention**	**Assessment of indirect impact of the policy interventions**
Sex Ratio	Other policies		At state level
Tobacco Smokeless	Health policies	At state and district levels	
Tobacco Smoking	Health policies	At state level	
Scheduled Population	Other policies		At state level
Toilets	Health policies	At state and district levels	
Literacy Rate	Other policies		At state level
Working Population	Other policies		At state and district levels
Agricultural Workers	Other policies		At state level
Marginal Workers	Other policies		At state level
Household Size	Other policies		At state and district levels

3. Concluding remarks

The key research problem in this study was as follows:

In order to reduce prevalence of acute/ chronic diseases in India, at what administrative level/s should the concerned health polices be customized (state/ district/ both/ none), so that the resources are efficiently used, and outcomes are effectively obtained?

The answer to this, based on the findings and discussion, can be simply states as follows:

It depends upon the determinant of health under consideration.

We started the data analysis with 34 determinant-indicator relations, 17 each for the two health indicators under consideration: prevalence of acute diseases and prevalence of chronic diseases. With the help of multiple regression, we could conclude that in a multi-factorial setting, the prevalence of acute and chronic diseases is predicted by only 8 and 10 determinants respectively. Based on popular beliefs and historical evidence, these 18 determinants are expected to have a causal relation with the prevalence of acute or chronic diseases, which can be classified as direct or indirect.

Determinants like tobacco and alcohol consumption have direct causal relation with prevalence of diseases, and therefore, they form an important part of the interventions in the Indian health policy. On the other hand, determinants like sex ratio and working population are related to prevalence of diseases in an indirect way, and their specific causality can be assessed by explaining their respective underlying mechanisms. All such indirect factors form integral parts of other (not health) policies like development policies and employment policies, and their long-term (but indirect) impact on prevalence of diseases can be assessed.

Weather we customize health policy (for determinants with direct causality) or assess the impact of other policies (for determinants with indirect causality), we should take into consideration the effect of belongingness to a state, as a significant amount of variation was found to be attributed to an individuals' belongingness to his/ her state. This is applicable for all the 18 determinant-indicator relations that were found to have a predictive effect on prevalence of acute and chronic diseases in a multi-factorial setting. In addition, belongingness to a district cannot be ignored

either, but it was found to be applicable for only 9 out of the 18 concerned determinant-indicator relations.

Coming back to the research problem, it can be concluded that, in order to reduce prevalence of acute diseases in India, health policies should be customized at the state as well as district level for interventions on tobacco smoking and use of toilets. However, for intervention on alcohol consumption, policies should be customized only at the state level. In addition, in order to reduce prevalence of chronic diseases in India, health policies should be customized at the state as well as district level for interventions on consumption of tobacco in smokeless form and on use of toilets. However, for intervention on tobacco smoking, policies should be customized only at the state level. With this customization, we can expect to execute the health policy in an effective and resource efficient way.

4. Limitations

The findings of this study are based on aggregate data, which means that there is an incorporated risk of ecological fallacy. Thus, the findings may not truly represent the situation at individual level. However, there is no way to determine that unless data at individual level is used for a similar analysis. In addition, many determinants used in the analysis do not have a direct causal relation with prevalence of diseases. This means that in order to understand their causality, we should try to explain their underlying mechanisms, and that could not be included in this study.

5. Scope

First, to overcome the limitations mentioned above, further studies with individual level data and/ or a more determinant specific analysis can be conducted. Second, the approach for measurement of the predictive ability of determinants in a multi-factorial setting using multiple regression analysis can be duplicated for use in other similar contexts. Third, the approach for measurement of the predictive ability of the level of aggregation in a multi-level setting using multi-level analysis can be duplicated for use in other similar contexts. Lastly, multi-level analysis model used in this study can be further extended to include random slopes so that more determinant specific results can be obtained.

References

AFMC. "Illness Behavior and Need for Care." *ww.afmc.ca.* 2009. http://phprimer.afmc.ca/Part1-TheoryThinkingAboutHealth (accessed November 13, 2016).

Alin, Aylin. "Simpson's paradox." *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010: 247-250.

Anand, Sudhir. "Th concern for equity in health." *Journal of Epidemiology and Community Health*, 2000: 485-487.

Bettcher, Douglas. "International public health instruments." In *Oxford textbook of public health*, by Roger Detels, Robert Beaglehole and Mary Ann Lansang, 368-382. Oxford: Oxford University Press, 2009.

Brennan, Andrew. "Necessary and Sufficient Conditions." In *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition)*, by Edward N Zalta. 2012.

Canadian Institute for Health Information. *Indicators.* 2016. https://www.cihi.ca/en/health-system-performance/performance-reporting/indicators (accessed November 13, 2016).

CDC. *Social Determinants of Health.* 2014. http://www.cdc.gov/socialdeterminants/Definitions.html (accessed November 13, 2016).

Coggon, D, G Rose, and DJP Barker. *Epidemiology for the Uninitiated.* London: BMJ Publishing Group, 1997.

Coleman, James S. *Foundations of Social Theory.* Cambridge: The Belknap Press of Harvard University Press, 1990.

CSDH. *Closing the gap in a generation: Health equity through action on the social determinants of health.* Final Report of the Commission on Social Determinants of Health, Geneva: World Health Organization, 2008.

Demeulenaere, Pierre . *Analytical Sociology and Social Mechanisms.* New York: Cambridge University Press, 2011.

Eisenberg, Leon. "Disease and illness Distinctions between professional and popular ideas of sickness." *Culture, Medicine and Psychiatry*, 1977: 9-23.

Elstad, Jon I. "Health determinants, equity, and social inequalities in health." Oslo, September 18, 2014.

Elster, Jon. *Explaining Social Behavior: More nuts and bolts for the social sciences.* Ney York: Cambridge University Press, 2007.

Heath, Joseph. "Methodological Individualism." In *The Stanford Encyclopedia of Philosophy (Spring 2015 Edition)*, by Edward N Zalta. 2015.

Hedström, Peter , and Peter Bearman. *The Oxford Handbook of Analytical Sociology.* New York: Oxford University Press, 2011.

Hedström, Peter. *Dissecting the Social: On the Principles of Analytical Sociology.* New York: Cambridge University Press., 2005.

Hedström, Peter, and Richard Swedberg. *Social Mechanisms: An Analytical Approach to Social Theory.* New York: Cambridge University Press, 1998.

Hitchcock, Christopher. *The Stanford Encyclopedia of Philosophy.* 2012. http://plato.stanford.edu/archives/win2012/entries/causation-probabilistic (accessed November 13, 2016).

Hoefer, Carl. *The Stanford Encyclopedia of Philosophy.* 2012. http://plato.stanford.edu/archives/spr2016/entries/determinism-causal (accessed November 13, 2016).

Hox, Joop J. *Multilevel Analysis Tehniques and Applications.* New York: Routledge Taylor & Francis Group, 2010.

Kant, Immanuel. *Critique of pure reason.* 1787.

Lin, Ann Chih. "Bridging Positivist and Interpretivist Approaches to Qualitative Methods." *Policy Studies Journal*, 1998: Vol. 26, No. I, (162-180).

Little, Daniel. "Ananlytical Sociology and the rest of Sociology." *Sociologica*, 2012: 1-47.

Ministry of Home Affairs, Government of India. "Latrine Facility." *http://censusindia.gov.in/.* 2012. http://censusindia.gov.in/2011census/hlo/Data_sheet/India/Latrine.pdf (accessed November 13, 2016).

—. *Metadata.* 2011a. http://www.censusindia.gov.in/Meta-Data/data/metadata.html (accessed November 13, 2016).

—. *Normal Households By Household Size.* 2011b. http://www.censusindia.gov.in/2011census/hh-series/hh01.html (accessed November 13, 2016).

—. *Scheduled castes and scheduled tribes.* 2011c. http://censusindia.gov.in/Census_And_You/scheduled_castes_and_sceduled_tribes.aspx (accessed November 13, 2016).

—. *Sex Ratio.* 2011d. http://censusindia.gov.in/Census_Data_2001/India_at_glance/fsex.aspx (accessed November 13, 2016).

MoHFW, Government of India. *http://www.mohfw.nic.in/.* December 2014. http://www.mohfw.nic.in/showfile.php?lid=3014 (accessed November 13, 2016).

—. "Indicators from Annual Health Survey 2010-11." *www.data.gov.in.* April 28, 2016. https://data.gov.in/catalog/indicators-annual-health-survey (accessed November 13, 2016).

—. "Reported Prevalence of Morbidity - DLHS IV." *www.data.gov.in.* December 29, 2014. https://data.gov.in/catalog/district-level-household-and-facility-survey-dlhs-4 (accessed November 13, 2016).

Murzi, Mauro. "Logical Positivism." In *The New Encyclopedia of Unbelief*, by Tom Flynn. Prometheus Books, 2007.

Pearl, Judea. "Understanding Simpson's Paradox." *www.ucla.edu.* December 2013. http://ftp.cs.ucla.edu/pub/stat_ser/r414.pdf (accessed November 13, 2016).

Planning Commission. "Villages electrified." *www.data.gov.in.* April 28, 2014. https://data.gov.in/catalog/villages-electrified (accessed November 13, 2016).

Porta, Miquel. *A dictionary of epidemiology, 6th edition.* New York: Oxford University Press, 2014.

Reigalman, R. "Contributory cause: Unnecessary and insufficient." *Postgraduate medicine*, 1979: 66 (2): 177–179.

Rogers, Everett M. *Diffusion of Innovations, 5th edition.* New York: Free press, 2003.

Roth, Wendy D, and Jal D Mehta. "The Rashomon Effect: Combining Positivist and Interpretivist Approaches in the Analysis of Contested Events." *Sociological Methods & Research*, 2002: Vol. 31 No. 2, pp131-173.

Saracci, R. "The World Health Organization needs to reconsider its definition of Health." *ritish Medical Journal* 314 (1997): 1409-10.

Scully, Leach Jackie. "What is a Disease?" *EMBO Reports* (EMBO Reports), 2004: 650-653.

Sen, Amartya. "Why Health Equity?" *Health Econmics*, 2002: 659-666.

Snijders, Tom AB. *Multilevel Analysis An Introduction to Basic and Advanced Modelling.* London: SAGE Publications Ltd., 2012.

Tobin, James. "On Limiting the Domain of Inequality." *The Journal of Law and Economics*, 1970: Vol. 13, No. 2 (Oct., 1970), pp. 263-277.

UNICEF India. "Water, Environment and Sanitation." *www.unicef.in/.* 2016. http://unicef.in/story/1125/water--environment-and-sanitation (accessed November 13, 2016).

United Nations. "The Universal Declaration of Human Rights." *http://www.un.org/.* December 10, 1948. http://www.un.org/en/universal-declaration-human-rights/ (accessed November 13, 2016).

Üstün, B, and R Jakob. "Calling a spade a spade: meaningful definitions of health conditions." *Bullettin of World Health Organization*, 2005: 802.

Whitehead, Margaret . *The concepts and principles of equity and health.* Discussion Paper, WHO Regional Office for Europe, Copenhagen: WHO, 1990, 5-7.

Whitehead, Margaret, Goran Dahlgren, and Lucy Gilson. "Developing the policy response to inequities in Health: a global perspective." In *Challenging inequities in health care: from ethics to action.*, by Timothy Evans, Margaret Whitehead and Finn Diderichsen, 309-322. New York: Oxford University Press, 2001.

WHO. "About WHO." *http://www.who.int/.* July 22, 1946. http://www.who.int/about/mission/en/ (accessed November 13, 2016).

WHO. *International statistical classification of diseases and related health problems.* Manual, Malta: WHO, 2011.

WHO. *Preamble to the Constitution of the World Health Organization.* Official Records of the World Health Organization, New York: WHO, 1946, 100.

—. *Social Determinants of Health.* 2016b. http://www.who.int/social_determinants/en/ (accessed November 13, 2016).

—. *The Determinants of Health.* 2016a. http://www.who.int/hia/evidence/doh/en/ (accessed November 13, 2016).

—. *What are social determinants of health?* 2016c. http://www.who.int/social_determinants/sdh_definition/en/ (accessed November 13, 2016).

WHO. *World Health Statistics.* Geneva: World Health Organization, 2012.

WHO. *World Health Statistics.* Geneva: World Health Organization, 2016d.

World Bank. *Cause of death, by non-communicable diseases (% of total).* 2016b. http://data.worldbank.org/indicator/SH.DTH.NCOM.ZS (accessed November 13, 2016).

—. *Health.* 2016a. http://data.worldbank.org/topic/health?view=chart (accessed November 13, 2016).

—. *Health Nutrition and Population Statistics.* 2012. http://databank.worldbank.org/data/reports.aspx?source=health-nutrition-and-population-statistics (accessed August 10, 2016).

Annexure No. 1: Cone shape typical of heteroscedastic data