1. METHODOLOGY

Certain methods are used for the collection, analysis and interpretation for the study of water-borne diseases in Ahmedabad agglomeration area. Mainly four types of techniques are used for identification, explanation and decision making of the water-borne diseases namely, statistical methods, graphical methods, tabular methods and cartographical methods. Both conventional and modern automated cartography with computer software, GIS, are used to describe spatial phenomena through maps. Study is based on spatial distribution of variables related to internal and external environmental conditions, socio-economic background of residents and their behavioural pattern of use of water. Variables have been digitized for superimposition, interpolation and classification of the study area for environmental quality, using the software package called Geographical Information System (GIS). The description of such spatial phenomena is helpful in knowing the locations of areas of diseases which need specific planning to control the diseases over space and time.

Methodology adopted in the study

(a) Statement and formulation of hypothesis:

After collecting some important data related to the study, efforts are
The hypothesis tested are as follows:

(i) In general the number of cases has increased over time (1989-1993).

(ii) There is a tendency of having more cases of typhoid fever during the pre-monsoon period, while more cases of gastro-enteritis and cholera are seen in the rainy (monsoon) season.

(iii) The cases of water-borne diseases are not evenly distributed over space (ward wise) in the city of Ahmedabad.

(iv) Water-borne diseases are related to the availability of water.

(v) Cases are seen more in number where the total urban environmental quality is poor, that is, in the core and eastern industrial regions.

(vi) Water-borne diseases are negatively related to income and education.

(vii) Water-borne diseases are related to younger population.

(viii) The spatial distribution of water-borne diseases is also related to the eating and drinking behaviour of people

(b) Techniques used for the analysis of data:

Several techniques are used to analyse the collected data to arrive at certain conclusions. The techniques employed can be placed under six broad sub-heads, namely;

(i) Statistical techniques
(ii) Graphical and tabular representation

(iii) Maps and diagrams

(iv) GIS software and traditional cartographic techniques

(v) Sampling techniques

(vi) Qualitative technique for studies of environmental perception.

(i) **Statistical Techniques:**

The real purpose of statistical method is to make sense out of facts and figures, to probe into the unknown, and to cast light upon the situation. Quantitative techniques are used here to group and classify data appropriately. Statistical techniques, such as mean, standard deviation, correlation-co-efficient, chi-square test etc are used to analyse the data and test the hypothesis stated earlier. Details about the statistical techniques are used for testing individual hypothesis have been given as follows:

Hypothesis (i) is tested by finding the trend line of occurrence of number of water-borne cases to know whether cases have increased over time (1989-93).

Hypothesis (ii) is tested by applying 'chi-square' test to show tendency of having more cases of hepatitis during the pre-monsoon period, while more cases of Gastro-enterities and cholera are seen in the rainy seasons.

Hypothesis (iii) is tested by plotting the special Lorenz Curve of population and cases of people affected by water-borne diseases in the various zones of Ahmedabad agglomerated areas. It has been also tested
by the method of chi-square test.

Hypothesis (iv) is tested by applying 'chi-square' test

$$X^2 = \sum \frac{(o-e)^2}{e}$$

to water-borne diseases which are related to the availability of water.

Hypothesis (v) is tested by applying 'regression' of cases

$$r = 1 - \frac{\sum d^2}{n^2 - n}$$ and $$y = mx + c$$

that are seen more in number where the total urban environmental quality is poor. The centre of gravity of cases has been calculated using,

$$\bar{X} = \frac{\sum x_i}{n}, \quad \bar{Y} = \frac{\sum y_i}{n}$$

Hypothesis (vi) is tested with the help of 'regression' to know water-borne diseases is negatively related to income and education, separately.

Hypothesis (vii) is tested by using 'chi-square' test as water-borne diseases are related to younger population.

Hypothesis (viii) is tested by regression analysis to know the spatial distribution of water-borne diseases is also related to the eating and drinking behaviour of people.

**Ratios:** A number of ratios are used to analyse the demographic features. They are as follows:

$$\text{Density per sq.km} = \frac{\text{Total population}}{\text{Total area in sq.km}}$$
Sex ratio = \[ \frac{\text{Total female}}{\text{Total male}} \times 100 \]

Growth rate = \[ \frac{(\text{Popn.1991}) - (\text{Popn.1981})}{\text{Popn.1981}} \times 100 \]

\[
L.Q. = \frac{\text{Total cases of a ward}}{\text{Total population of a ward}} \times \frac{\text{Total cases in the whole city}}{\text{Total population of the whole city}}
\]

If the ratio is >1, the concentration is high and vice versa.

(ii) Graphical and Tabular Representation:
In order to represent the data more vividly and to know the spatial and temporal distribution of diseases at a glance number of graphical and tabular methods are used in this research work. To compare the rainfall and diseases - 4th day after the rainfall and previous day of the rainfall and cases of water-borne diseases are compared with the help of compound bar-graph. Population growth of Ahmedabad agglomeration area is shown by simple line graph. Month-wise distribution of water-borne diseases (1989-93) is represented by compound bar-graph. Demographic profile of the Ahmedabad agglomeration area for 1989-93 with respect to male-female are given through compound bar. Age-wise distribution of water-borne diseases, literacy and diseases, religion and diseases are also represented by compound
Drawing of spatial Lorenz Curve and Index of Concentration is used to show ward-wise distribution of selected four water-borne diseases in relation to ward-wise distribution of population of Ahmedabad agglomeration area (1991). The procedure is similar to that for the non-spatial Lorenz Curve, it is drawn by expressing the frequency in each category as a percentage of the total frequencies, and then plotting the result geographically in the form of cumulative frequency curve. But there are two important differences:

- The curve, will become progressively less steep with greater distance from the origin, the areal units must be listed in a particular way before cumulating the data. This is done by calculation of R, the ratio between the variables to be compared, for each standard region in this curve, the ratio between the percentage of total ward-wise population and percentage of total cases of water-borne diseases. The listing of wards in descending order of R.

- Percentage for both variables are then cumulated in the order and plotted in fig. VII.12 with the 'standard' data along the horizontal axis. \(^{(49)}\)

There are few steps for the index of concentration. First step is to mark off ten equally spaced points along the horizontal axis. Then project vertically from these points to the Lorenz Curve at L1, L2, etc. After this project from L1, L2, etc. to the vertical axis at C1, C2, etc. and note the values of C1, C2, on the vertical scale. Lastly add the ten values, C1, C2 etc. to give C. The index of concentration
is now given by:

\[ IC = \frac{C - 550}{1000 - 550} = \frac{C - 550}{450} \]

This will give a maximum value of 1 (maximum concentration, when the Lorenz curve deviates as much as possible from the diagonal and when therefore \( C_1 = C_2 = C_3 = C_4 \text{ etc.} = 100 \), so that \( 10 \times 1000 \)); and a minimum value of 0 (minimum concentration, when the Lorenz curve corresponds exactly to the diagonal, and when therefore \( C_1=10, C_2=20, C_3=30 \text{ etc.} \text{ So that } C = 550 \)).

All data can not be represented by graphs. As a result, here some data are represented in tabular form. Ground-water level, length of pipeline, number of connections and water consumption pattern in Ahmedabad, separate bath-room facilities, number of slums and dilapidated house, and all other factors which are affecting the water-borne diseases are represented in table form also, which can give at a glance ideas about important factors for water-borne diseases.

(iii) Maps and Diagrams:

Any geographical research can not be completed without any map, because geography is related with space, which is represented by maps. Choropleth maps have been drawn to show the spatial pattern of water-borne diseases over a period of time. A series of maps has been drawn to show the distribution of different types of variable (nearly 33) which have affected residents to water-borne diseases. Super-imposition of distribution of different variables and diseases has been done with the help of a computer software called GIS. Location of the study area, distribution of sewerage and water-pipe
line, location of dilapidated houses and slums, water logged area, points of effluents are shown in maps.

**Spatial Analysis:** Weighted mean of spatial centres of diseases for five years (1989 to 1993) have been calculated separately. Then those five points have been plotted to know whether the centre was at the same place in the city or it has shifted during these five years. The following method has been used:

By locating the central point in each 63 ward by eye estimation and numbering them as per serial number the co-ordinates of 63 points in x and y axis have been found out. Then Multiple $X_1 \times N_1 + X_2 \times N_2 + X_n \times N_n$ ($N$ = number of diseases in a ward in a year). Multiple $Y_1 \times N_1 + Y_2 \times N_2 + Y_n + N_n$. Then weighted means of $X$ and $Y$ were found by

$$\bar{X} = \frac{\sum x \times N}{Total \ cases}, \quad \bar{Y} = \frac{\sum y \times N}{Total \ cases}$$

There will be total of 5 points as weighted mean centres for 5 years (number I, II, III, IV, V) shown in a map.

**Choropleth maps:**

A large number of choropleth maps have been prepared both by traditional and modern cartographic methods. Thirty three variables have been used to depict the environmental quality of the city with sixty-three wards. Each variable has been given weightage, ranging from 1 as "very poor" to 5 as "very good" in environmental condition. Each variable has been depicted as choropleth map. A method of normalisation for making choropleth maps has been based on the sample taken for primary data. Since the sample size was variable (as per the no. of cases of disease), the data was brought to equal sample size per word by multiplying by 16 (as explained below) and dividing by
sample size. It was decided to have one thousand sample (which is 6% of the total cases of 1986), randomly picked up. Later the number of cases to be sampled in a ward was decided by area and its population.

If equal number of sample is to be drawn from 63 wards, the number of sample would have been 16 (1000/63 = 16) per ward. As the concentration of cases and area are not equal in each ward, the number of sample is not 16 in each ward; somewhere it is more and somewhere it is less than 16. A method of "normalisation" has been adopted having the base as 16 (average sample size). For example, 20 sample in a ward with 33 variables equal to 700 score But 16 sample in a ward with 33 variables would have been = 700/20x16 = 560

Choropleth maps have been made with normalised weighted score.

With the help of normalised weighted total score of 33 variables, an integrated choropleth map of environmental quality has been prepared. 33 variables have been first combined to form six major environmental factors which have been overlaid with the help of the computer software called GIS to have a composite choropleth map of "environmental quality" of Ahmedabad. If normalisation was not done, the relation between environmental condition and cases of diseases was found to be abnormal. That is, with good environmental quality more cases were found which is not true in the field. This is due to the variation in area and population of the ward. If ward is poor in environmental condition with a large area and population, the total weightage would be large to depict it as a good area which may not be true. Hence normalisation of weightage is essential to give an appropriate picture of the environment at a micro-level.

Seasonal variations of water-borne diseases are represented through
multi-ternary diagrams. A multi ternary diagram represents the
distribution of water-borne diseases of five years (1989-93), with
main three seasons of summer, winter and monsoon At sub-regional
levels, different types of diseases in different seasons are shown.
Distribution of water-borne diseases according to income-group is
depicted by a pie-diagram.

(iv) GIS Software:

Geography is the science of spatial relationships. Maps form a major
constituent to geography as they are means of representing spatial
relationships. Cartography, the science of map making, is a technology
for the organization of geographical information into a map. However,
a map by itself is of no value unless it is put to use, with proper
interpretation. Geographical Information System or GIS can be manual
or automated. Manual methods are traditional and not adequate when
handling large data sets or when repeated analysis is required or
several alternative scenarios are to be generated. Hence the need of
computer software like Geographic Information System (GIS) is felt in
map making. GIS is a powerful tool, which needs a complete
understanding and careful application.

A Geographical Information System (GIS) is very helpful as a system
which facilitates to capture, display, manage, manipulate, analyse and
retrieve of spatial and aspatial data into digital format.

Geographic Information System (GIS) can be defined as a set of
integrated activities which provide us a tool to:

• integrate geographic data received from different sources such
  as maps, charts, tables, aerial photographs, satellite
imagery, in digital environment,

- attach thematic information to the geographic details,
- analyse results and to build up queries based on spatial information and to
- get the result in a desired mapped form. In addition, it provides other facilities such as data sharing, data security, data transportability, improved data quality and automatic repetitive process. Because of these capabilities GIS has become increasingly important in various disciplines for decision makers, planners, administrators, scientists, academicians.

Essentially there are two contrasting but complementary way of representing spatial data in the computer. They are referred to as explicit and implicit way of describing spatial entities. Explicit representation means that the form of an object is built from a set of points on a grid or RASTER. The implicit representation makes use of a set of lines, defined by starting and end points and some form of connectivity and is referred to as vector representation.

**Raster Data Structure:**

The simplest raster data structures consist of an array of grid cells and each grid cell contains a number, representing the type or value of the attribute being mapped. Each co-ordinate space of an object quantized for the entire object.

**Vector Data Structure:**

The vector data representation of an object is an attempt to represent the object as exactly as possible. The co-ordinate space, unlike in
the case of raster space, is a continuous one. All the positions, lengths and dimensions of the object are defined very precisely.77

A major reason for the growing importance of GIS is that once all data are stored, both the digital and data base can be manipulated simultaneously in two ways as, querying of the data base to generate a digital map for display and querying of a digital map to retrieve information from the data base.

GIS allows arithmetic as well as logical overlay operations. A GIS data base contains map information in the form of separate layers associated with attribute and data. Any query to a particular theme, a layer will result in answers to that particular theme only. For example, if queries are put to the landuse information layer, then the answers obtained will pertain to the landuse only. Similar is the case with other themes. But if the information on the other themes are also to be obtained simultaneously, a set of spatial overlay operations has to be performed on all the thematic maps. The overlay operation generates a new layer with information on the thematic maps of our choice causing no alterations to the input of layers. The overlay functions differ in their operations based on the type of data set, i.e. raster or vector, chosen for the analysis.77

There are so many packages for GIS. Every package has different facilities. GRAM ('Geo-Referenced Area Management') package has been used in this study. It is a software package to store and analyse spatial data on a low cost PC configuration. Evolved from experience gained in Natural Resources Data Management System (NRDMS) project of the Department of Science and Technology (DST), Government of India,
GRAM has been developed primarily to promote application of spatial data technologies to problems of resource management and planning. Developed by a multi-disciplinary team of research scientists at the centre for studies in Resource Engineering, Indian Institute of Technology (IIT), Bombay. GRAM is a product of many years of effort in geographic data handling, digital image processing and spatial modeling.

GRAM is based on DOS operating system and provides a user-friendly environment for carrying out the different operations. Most of the functions of the package are menu driven. Simple entries are needed for different systems at the time of execution of these functions. Help screens are provided for on-line support. GRAM is modular and can be tailored to suit specific applications.

GRAM package provides tools for developing information system to cater to the needs of planner and decision makers involved in spatial planning. The package is capable of handling both raster and vector inputs and can integrate data gathered from both remote sensing and ground based data using GEOTRANS function of the module which further enhances the utility of GRAM package for resource management and planning applications. Other sophisticated tools available with the module include mathematical and logical operations, overlying and proximity analysis. Resultant maps can be displayed and taken out as hard copies.

Possible applications of GRAM package are Landuse Planning (Rural Urban Planning), Land capability analysis, water-land studies, thematic mapping, environmental studies, hazard zonation, irrigation
scheduling, watershed planning, spatial modeling, facility location, demographic studies etc.\(^{(45)}\)

In the present study, the GIS package is widely used for the assessment of the distribution of water-borne diseases with relation to the various factor where the overlay of 15 maps are done to get a composite environmental quality map. The overlay function along with the regroup function is very helpful in handling a number of maps, with

<table>
<thead>
<tr>
<th>Classes</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>IV</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>V</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Source: Prepared by Author.

different classes. The overlay function is done in the following way:

Individual maps for different variables are prepared with weightage first manually, then digitized in the GIS (GRAM) and each map is rasterised after removing all the errors. These maps are overlayed in the overlay function, two at a time and after many steps the final map is obtained, as shown in the flow chart (Fig. II.1). All the parameters are divided into five classes and when two maps are overlayed, there are possibilities of getting nine categories. (Table II.1) In each overlay the classes go on increasing, hence after each overlay the classes are regrouped into five classes. After each overlay, the maps is regrouped into five categories. Then two regrouped maps are again overlayed, step-by-step as shown in the flow chart to get the final composite diseases and variables map (Fig. II.2).
There are main six factors namely ecological, demographic, socio-economic, drinking water, environmental and behavioural factors taken into the study. But this model focuses mainly on three main important spatial factors, namely drinking water, environmental factor and behavioural factor. These three factors vary from region to region within Ahmedabad and have a distinct spatial pattern; while other uncounted factors in this model are ecological, demographic and socio-economic factor which have similar impact in all regions, and do not give so much spatial variations. These three factors have similar effects on each region. To find out the importance of the factors according to region this model has been prepared for this study. Here

### Table II.2

<table>
<thead>
<tr>
<th>Region</th>
<th>% of 3 major factors by weighted average of cases (1993)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Behavioural</td>
</tr>
<tr>
<td>Core Region I</td>
<td>52</td>
</tr>
<tr>
<td>Eastern Region II</td>
<td>31</td>
</tr>
<tr>
<td>Western Region III</td>
<td>37</td>
</tr>
<tr>
<td>Eastern suburb region IV</td>
<td>41</td>
</tr>
<tr>
<td>Western suburb V</td>
<td>30</td>
</tr>
<tr>
<td>Total city</td>
<td>25</td>
</tr>
</tbody>
</table>

Source: Prepared by author with the help of field data.

Method of calculation is same for all factors in each region. Calculation has been done for each factor by region separately e.g. from the composite, overlay map of drinking water the number of wards with various intensity of water quality was found out. Later a map of cases with various intensity was compiled. After that % of no. of wards of a certain quality has been multiplied by % of no. of cases in
Fig II.1 Use of GIS: Flow chart of variables related to water-borne diseases in Ahmedabad (1993) using the software of Geographic Information System (GIS)

Variable Index (1993)

<table>
<thead>
<tr>
<th>Am1</th>
<th>Quality of water</th>
<th>Am2</th>
<th>Availability of water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Am3</td>
<td>Amount of water needed</td>
<td>Am4</td>
<td>Inner Environment</td>
</tr>
<tr>
<td>Am5</td>
<td>Outer Environment</td>
<td>Am6</td>
<td>Overall cleanliness of area</td>
</tr>
<tr>
<td>Am7</td>
<td>Purification of drinking water</td>
<td>Am8</td>
<td>Throwing daily garbage</td>
</tr>
<tr>
<td>Am9</td>
<td>Cleaning water tank</td>
<td>Am10</td>
<td>Habits of eating out</td>
</tr>
</tbody>
</table>

Source: Prepared by Author
Overlaying Data Base

Disease maps
Different variables maps

Drinking water
Quality of water
Amount of water needed
Availability of water

Inner Environment
Total of following variables
- Housing amenities
- Room density
- Cleanliness of house
- Types of house

Outer Environment
Overall Cleanliness of water
Total of following variables
- Water logging by sewage water
- Water logging by rain water
- Garbage accumulation
- Problems of mosquito/fly

Purification of drinking water
Throwing out daily garbage

Cleaning water tank

Habits of eating out drinking water

Fig 11.2
that ward according to intensity of very good to very poor. Then
average weightage (AW) has been found out.

\[ A.W. = \frac{\% \text{ of cases} \times \% \text{ of wards of region}}{\text{Total no. of wards of the region}} \]

This process has been done for all five regions for three factors.
From which tables have been prepared (Table II.2).

(a) By multiplying no. of wards with % of cases when concentration of
cases is high in particular category (like 'poor'), the total score is
high as in Core Region (171). When % of cases is low in dominant
category, which may be 'Fair' or 'good', the total score is low as in
Western suburbs (87). Other 3 regions are intermediate to the above
(b) To display model of importance of 3 factors, with equal size of
parallelopipeds, these scores are normalised to 100 (by multiplying
each score by 100 and dividing by total score of region).
(c) When analysing overall city, wards are approximately equally
divided into all 5 categories (VP, P, F, G, VG), hence weighted
average becomes ordinary average, which will be generally 20% (as the
cases would be also approximately equally divided in 5 categories) So
the total score in whole city will be also low (ideally 3x20% = 60%
but actually it is 78%). Hence it is important to examine region wise
(meso scale) and ward wise (micro level) data to understand the
spatial variations.

Now Region-wise % of three factors are found. When the total is made,
it will not be 100%, because each factor is not completely exclusive
on the area but it affects jointly with other factors. So to remove
this problem, again region wise % is found out. Here we find each
region separately as 100% for each factor and with the help of above
table following new percentages have been found out (Table II.3). On
the basis of this table II.3 blocks have been drawn in each region for
the Environmental Behavioural Model. (Fig. VII.14).

Table II.3
Conversion of data of the model into % from Table II.2

<table>
<thead>
<tr>
<th>Region</th>
<th>% of 3 major factors by weighted average of cases (1993)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Behavioural</td>
</tr>
<tr>
<td>Core Region</td>
<td>30</td>
</tr>
<tr>
<td>I Eastern</td>
<td>26</td>
</tr>
<tr>
<td>II Western</td>
<td>45</td>
</tr>
<tr>
<td>Region III</td>
<td>36</td>
</tr>
<tr>
<td>Eastern suburb</td>
<td>35</td>
</tr>
<tr>
<td>Total city</td>
<td>32</td>
</tr>
</tbody>
</table>

Prepared by author with the help of field data.

(vi) Environmental Perception of residents and Perception Index (PI)

It has been tried to evaluate the environmental perception of residents with four important variables like cleanliness of houses, overall cleanliness of an area, availability of water, quality of drinking water. Perception Index has been evolved with a ratio of rating of environmental quality of residents with that of the Investigator.

\[
PI = \frac{Perception \ of \ residents}{Perception \ of \ investigator}
\]

P.I. can be equal to 1, when rating of residents and of that of investigator is equal, rating can be more than one (\(>1\)) where residents have over rated the environmental quality and it may be less than 1, when residents have under rated quality of the environment.

There are three main groups:

1. Over-rated group having PI more than 1. Here three sub-divisions can
be made.

(a) Highly over-rated group having P.I. more than 1.41
(b) Moderately over-rated group having P.I. 1.1 to 1.40
(c) Slightly over-rated group having P.I. 1.01 to 1.09

2. Second is similarly rated group having P.I. = 1
3. Third group is under-rating group having less than 1 P.I.

This method is useful to find out the environmental perception of the residents but demerits are there. This method is highly subjective in giving score. It differs from person to person. The quality of environment is also based on subjective evaluation which may change by time and by person. This method fully depends on the perception of the investigator. Result depends on ratings of residents as well as investigator.

(vii) **Sampling Techniques:**

The information needed for this study occurs in a great variety of forms. The geographical study is that of the so-called 'data explosion', there are just too many isolated pieces of information concerned with spatial distribution for it to be possible to even start to analyse them all.

Sometimes it becomes difficult as well as impossible to study the characteristic properties of the whole population into consideration, only definite representative number of samples selected at random are to be analysed and the results obtain out of that are to be treated valid for the whole population. It is difficult to study each and every case due to short span of time, man power and money, hence sampling techniques are used here.
In 1985 nearly 6,000 cases occurred for water-borne diseases in the records of government hospitals. So this study is based on 6% sample size for study purpose. Here 1000 cases had been taken into account. It means sample size is 1000. In short, stratified random sampling had been done to get primary data by area or ward.

Field survey is essential as there is no individual data on the behavioural and awareness of patients, environment and houses they live-in and socio-economic status of residents. This study also uses interview method for collecting above information. Sample had been in proportion to the occurrences of cases or space by region.

Sampling technique applied for personal interview is with the help of field questionnaire. Intensity of diseases is not equal in all wards as well as in regions. Sometime area and population of a ward is low and cases of water-borne diseases is high due to environmental and other factors. In those areas more interview and samples had been taken. Samples are not equally distributed.

<table>
<thead>
<tr>
<th>Table IX.4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample taken from each region is as follows:</strong></td>
</tr>
<tr>
<td>Region</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

N.B. Region 1, Khadia 1 & 2 and Kalupur N & S have taken together so that region I has 10 wards.

Source: Prepared by the author.
2. DATA ASPECTS

The word 'data' is a general term used in statistics to denote the figures or other information collected. Data is the heart of the research. Without any type of data one cannot come to any conclusion. This aspect is related with source of data, nature of data, problem and limitations of data; which are as follows:

(a) **Source and nature of data**:

The sources of data can be either primary or secondary, of which the former is a direct one collected by the investigator himself or herself. Whereas the latter is the type of published and other data already collected by someone else.

The important secondary data collected in this study are as follows:

(i) Data on cases of four types of diseases (gastroenteritis, viral-hepatitis, typhoid and cholera) are usually available in the patients' registers maintained by various Ahmedabad Municipal Corporations Hospitals (i.e. V.S. Hospital, Shardaben Hospital, L G Hospital and I.D. Hospital) and from the public health department of Ahmedabad Municipal Corporation.

(ii) All demographic factors for 1991 such as population, literates, SC/ST population, are collected from census Handbooks of Ahmedabad. For the years of 1989 to 1993 all detail about demographic data such as population, sex ratio, literates and about housing types of houses like dilapidated and slums are collected from yearly outline of statistical handbook of Ahmedabad Municipal Corporation.

(iii) Data on rainfall 1989-1993 are collected from Agricultural Department.
(iv) Data for water quality are collected from Ahmedabad Municipal Laboratories.

(v) Location of out fall points of effluents areas area collected from Ahmedabad Municipal Corporation.

Temporal span of five years from 1989 to 1993 are considered for the study and collected from the above sources.

Primary data is very useful as it covers individual characteristics of socio-economic data, environmental quality as well as behavioural factors which were not available in published form. For that the researcher has conducted intensive fieldwork to collect relevant information through a detailed one thousand questionnaires (Appendix No.II.1) which include following information.

* Family information as details of family members according to age, sex.
* Information on socio-economic aspects like income, education, religion, occupation etc.
* Health aspects like source of water, quality of water, quantity of water available as well as needed per person per day.
* Information about inner and out environment, inner environment includes housing amenities, room density, cleanliness of house, types of houses etc. Outer environment includes water logging by sewerage and rain water, cleanliness of an area, problem of mosquito/fly etc.
* Information about behavioural factor which is important for this study. Different types of habits like purification of water, cleaning habits viz. washing hand before eating and after latrine, cleaning of water-tank, throwing of daily garbage, habit of eating out, and awareness about taking bath, precaution and medicine uses,
cleaning of clothes and dishes of patients separately with boiled water.

The data collected for the study are both qualitative and quantitative in nature and have been collected from primary as well as secondary sources.

(b) Problem of Data:

(i) Source of data:
Source of data has a number of shortcomings. Data collection is the most important factor for researchers and data collection itself is a big problem. So limitations of the data need to be recorded for rectification in future.

(ii) Base Map:
To prepare a base map itself was a problem. Municipality has prepared the city map for their own purpose i.e. for election work. Census has the same outer boundaries but inner ward boundaries differ from municipality's ward boundaries. So it was very difficult to prepare ward boundary map of Ahmedabad city, and match the data about population. Recently municipality has prepared a new ward boundary map of Ahmedabad accordingly to census.

(iii) Lack of Data:
Here only Government hospitals and Municipal Health Department hospitals considered as the main source of data for diseases. Private hospitals and private dispensaries do not maintain registers about patients' records. So it is very difficult to find out the exact number of cases of all diseases. Thus the data used here are not
necessarily published by the Public Health Department of Ahmedabad Municipal Corporation. So it may give different picture about cases and places.

(iv) **Accuracy of Data:**

The adequacy and reliability of morbidity data, are questionable. One can say confidently that, due to the acuteness of the symptoms, the major water-borne diseases, viz., Gastro-enteritis, Viral Hepatitis, Typhoid (Enteric-Fever) and Cholera, patients are compelled to attend doctors. Every person does not like to take treatment in municipality hospitals. Sometimes the wealthy people are treated by private practitioners and hence remain out of record. Sometimes because of economic condition and less knowledge about acuteness of diseases a negligible percentage does not reach to any kind of medical help and it increases the mortality figures.

(v) **Lack of information of quality of water:**

Data about the water quality is a great problem. For finding out the effect of water quality on diseases, data about quality of water is needed, which is not available in proper way. Ahmedabad Municipal Corporation has its own laboratories. The data on quality of water, collected from the Ahmedabad Municipal Corporation's Public Health Department is not adequate. They are collecting water for testing of quality, every month from different wards. They also accept complaints about water and they try to test the water quality. They also give service to people for personal water testing by paying fees.

Water-borne diseases do not fully depend upon water-quality. Mainly Gastro-enteritis cases are coming up because of contamination of water.
with that of gutter. Municipality laboratories take care of it, but whatever report they have collected about unfitness of water is not adequate prior to 1994 as only the presence of E Coli bacteria has been considered for determining the fitness of the water sample. Records on water testing do not show the levels of unfitness, as other pollutants are present in the sample. Records on water testing do not show the levels of unfitness, as other pollutants are present in the sample. But whatever method and record they have applied is sufficient for them. They have no data according to ward as well as they have no clarity about which particular water sample is not fit for drinking purpose. So for the years earlier to 1994 it is difficult to correlate the water quality and cases of water-borne diseases. The detailed data on water-quality of 1994, was available, which covers Ahmedabad Municipal Corporation's limits. Another problem is detailed water testing is very much expensive, so it is very difficult for researcher to test the water according to time and space.

(vi) Mismatch of Data:
Next data problem is about frequency of cases. Ahmedabad Municipal Corporation has also collected some data about this four particular diseases according to time (year-wise) and space (ward-wise). Researcher has also collected cases of diseases data from four municipality hospitals (V.S. Hospital, Shardaben Hospital, L G Hospital, I.D.H.). Municipal Corporations Health Department has collected data about diseases from four hospitals. There is mismatch which is due to some following reasons. Data on patients were available continuously only for years 1989-1993 from the above hospitals.

* Municipal health department has records of civil hospital and other
hospitals too; which researcher do not have due to lack of available patients, registers. Civil Hospital does not maintain patients' registers because of heavy rush of patients and lack of man power. So for researcher, it's difficult to get accurate data from Civil Hospital. They have all records in case paper but within a short period of time, data about five years on these diseases is impossible to find out, as case papers are not maintained properly

- Ahmedabad Municipal Corporation has started to collect the health data from about 1988. Due to heavy rain in 1991 some hospital's patients' registers are washed away.

- Ahmedabad Municipal Corporation's Health Department has collected health data only from corporation's limit area, but researcher's study area is Ahmedabad Agglomeration area. Naturally frequency of data are not same.

- Outer limits of AMC area have noted less number of cases This study has taken four AMC hospitals which are far from the suburban area of Ahmedabad. So people generally go to the nearest private hospital which is located nearby their residence.

- Municipality Health Department is collecting data about cases of diseases every-day evening. Hospital gives their diagnosis about patient at a time, which may not match with the patients' diseases later. At evening Municipal Health Department has a patient's record under Malaria diseases, but after some laboratory testing patient may be replaced under Typhoid. Hence frequency of cases according to municipality differs from that of researcher's data. This study is based on the four hospitals patients' register, which is the final diagnosis after investigation by doctors

The data on behavioural aspects have as usual problems related to the
accuracy of responses obtained through field work. However, there is no other way that behavioural data can be obtained. Without the data on human behaviour, causes of health problems cannot be understood...