CHAPTER 3:

Research Approach and Methodology

3.1 Introduction

The growing availability of various types of information offers opportunities for retrospective analyses and computer processing to integrate health, epidemiological, demographic and socio economic data for developing early warning systems for epidemic cholera (Pascual *et al.,* 2002). GIS technology provides new opportunities to study associations between social and environmental factors and the spatial distribution of disease (Tim, 1995; Croner, 1996). The ability to evaluate geospatial information provides a unique perspective of public health issues associated with an infectious disease like cholera (Waring *et al.,* 2005). GIS also provides computer mapping and analysis capabilities of integrating large quantities of geographic (spatial) data as well as linking geographic with non-geographic (study) data e.g., demographic, climatic, socio-economic information and in this case cholera incidence (Vine *et al.,* 1997). As such, a study of this nature called for an inter-disciplinary approach in the management of data relating to cholera, demography, socio-economic status and climate to realize the full potential of GIS technology towards understanding the complex questions at hand. Indeed, a number of environmental epidemiologists have made use of GIS to find possible associations between environmental drivers, socio-economic factors and spatial distributions of cholera (Craig 1988; Emch 1999; Myaux *et al.,* 1997)

3.2   Rationale

Many investigations of cholera have been approached from a clinical standpoint, thus falling short of appreciating the role of the socio-economic and natural environments in the transmission of *V.cholerae.* The magnitude of the cholera epidemic in KZN gave credence to investigate the multiple factors that may have been responsible for the outbreak of cholera, as well as those that assist in the spread of the disease. Exactly one year after the epidemic broke out in August 2001, an assessment of the cholera situation revealed that 99% of the cholera cases were reported from KZN (Dept of Health, 2002b). This raised the question as to why almost all the cases were from KZN, while during the same period; the Northern Cape province did not have a single case on record (Dept of Health, 2002b).
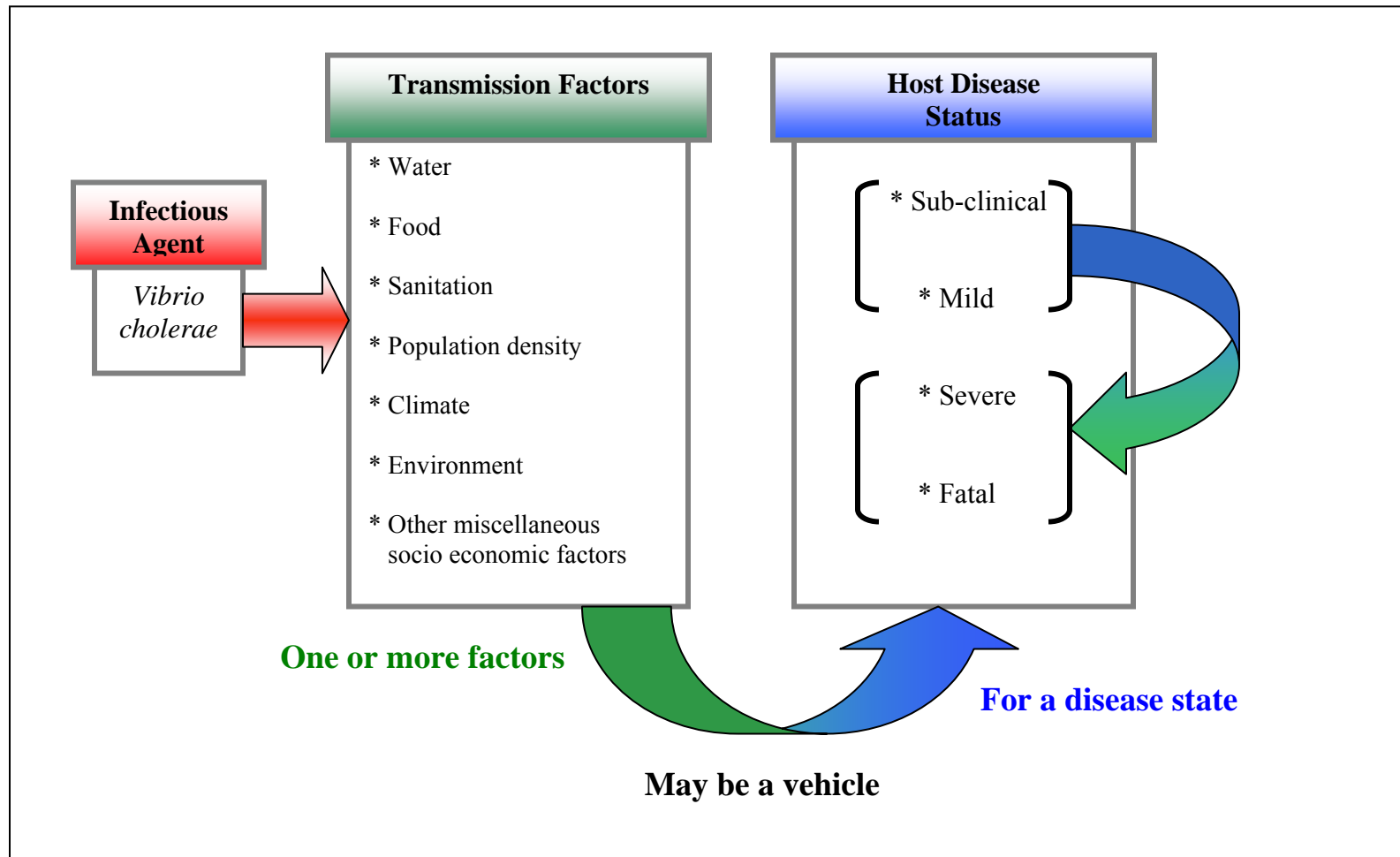
Figure 3.1: The possible relationships between the cholera pathogen, transmission factors and the human host.

A
Overall Project Scope

Review

B
Study Approach

Review

Design Phase

C
Data Collection

D
Creation of Attribute Database

Review

Data Management Phase

E   Data Assessment

F
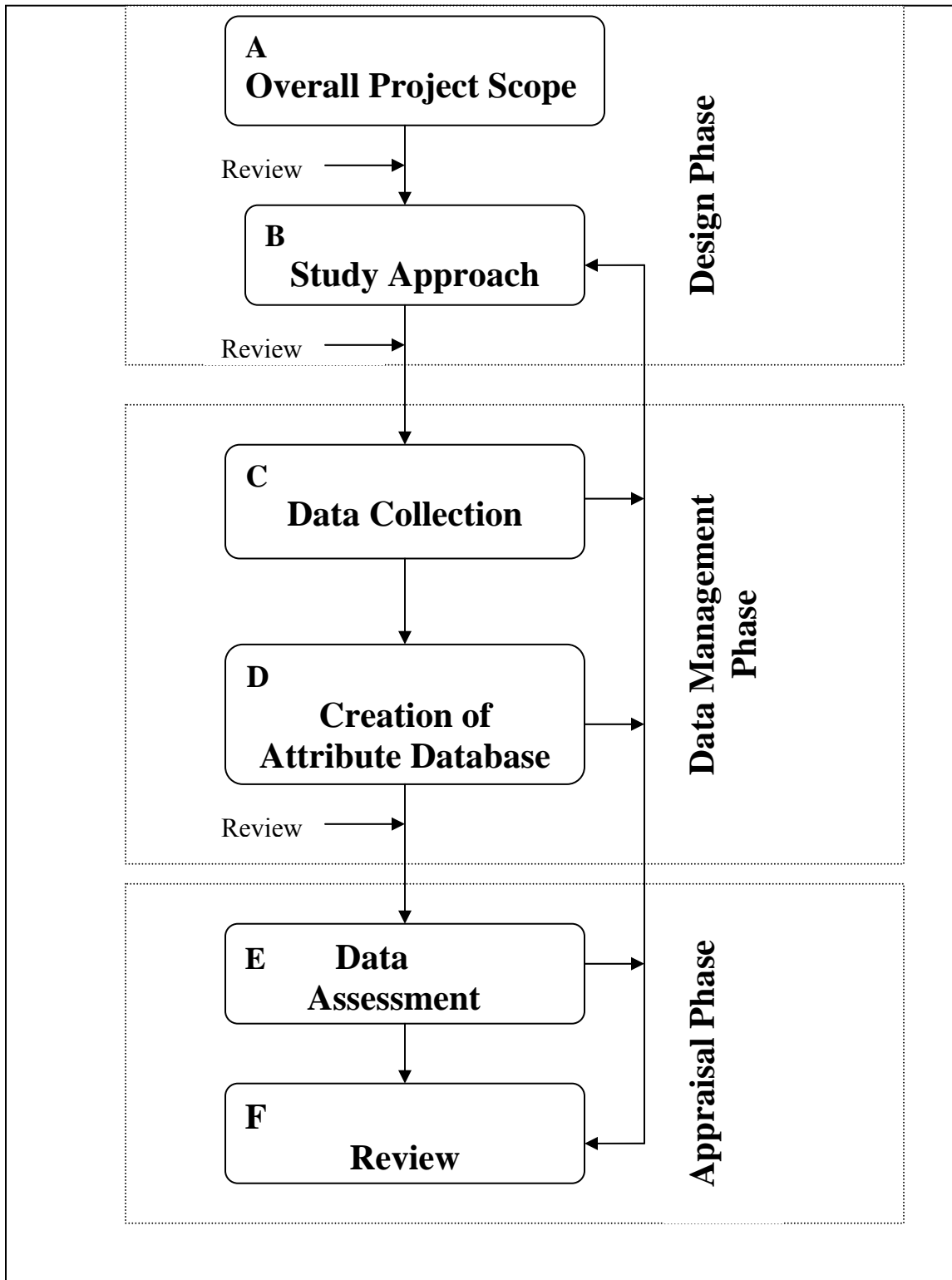**Review**

Appraisal Phase

Figure 3.2: An illustration outlining the research framework.

This state of affairs lead to the thinking that possibly the climate had a role to play in the cholera epidemic of KZN. Especially when one considers that the socio-economic status of rural areas in the other provinces was just as disadvantaged as those in KZN. The climatic aspect of the other provinces however, differs significantly to that of KZN, which is characteristically hot and humid. This observation gave impetus to the proposal to investigate the possible role of climatic conditions as the primary factors influencing the cholera outbreaks in Kwa-Zulu-Natal. Socio-economic factors were considered as enabling factors towards the spread of the disease. The anticipation was that the study findings would contribute in identifying risk factors including environmental factors that could be used as guiding principles to predict possible future cholera outbreaks within the region.

3.3 Design and development of the research framework

The research framework was drawn up to give direction to the project and in the process guide and include all the projected tasks as illustrated in Figure 3.2.

3.3.1   Design and operations framework (A)

3.3.1.1   *Designate scope (A1)*

The scope of the research was to assess the Cholera Database as well as compliment it with demographic, socio-economic and climatic databases. These would then be used as a disease management tool to determine the possible relationships between cholera and the demographic, socio-economic and climatic environment in KZN. The results would then be used to spatially map the cholera epidemic and attempt to interpret the disease trend. As such, the outcome of study would be used to guide the prediction of possible cholera outbreaks especially in areas that demonstrate risk factors associated with cholera. It will also assist to devise/design intervention measures capable of handling future possible cholera epidemics.

3.3.1.2   *Determine information requirement (A2)*

The main task was to define the types of data that were required and then locate their source and their availability thereof. As it were, all the data types required for the study already existed in
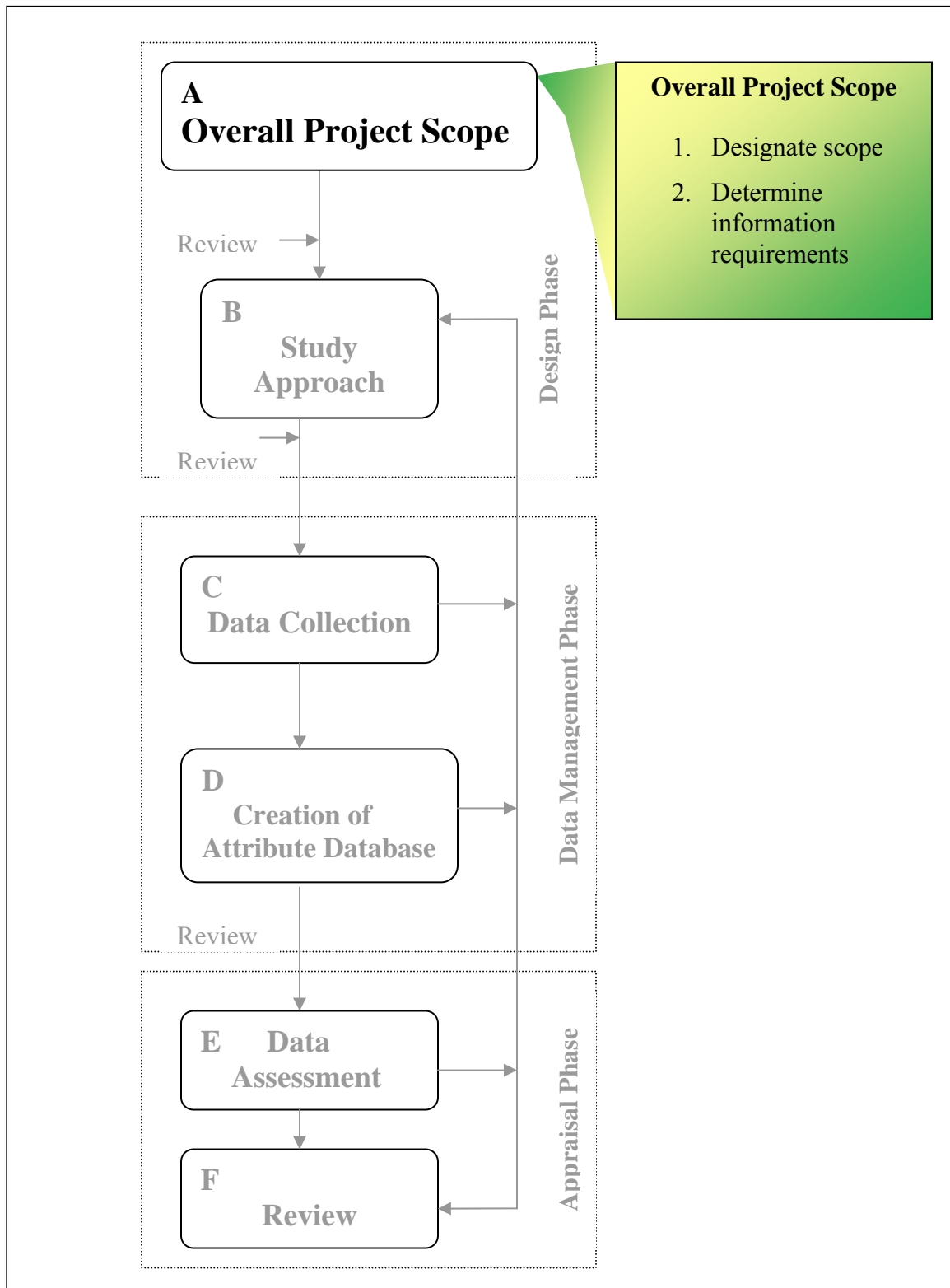
Figure 3.3: An illustration of the tasks assigned to the design and operations framework.

one form or another within the various relevant institutions. Following is a list of the types of data required for the study:

1. Data on cholera cases for the 2000-2004 epidemic.
2. Climatic data specific to the study area for the defined epidemic period.
3. The demographic profile (gender and age profiles) of KZN population.
4. Data on the socio economic status of KZN population that included supply clean water, sanitation, housing, refuse collection and income in the study area.
5. Geographical and spatial information of KZN communities including the topography, the areas (of place-names) in sq km and the types of natural water sources of the region.

3.3.2   Study Approach (B)

The study approach defined the purpose of the study by proposing a hypothesis. Thereafter, the hypothesis guided the formulation of the research objectives, which set out the course of action in the formulation of the various tasks towards the fulfilment of the research project.

3.3.2.1   *Hypothesis (B1)*

The hypothesis put forward was that climatic conditions play a significant role in the outbreak of cholera in KwaZulu, Natal. In addition, socio economic variables like sanitation, clean water supply, population density and public health services, contribute to the vulnerability of communities to the risk of cholera in KwaZulu-Natal. Thus the null hypothesis was that neither the climatic conditions influence the outbreak of cholera nor the socio-economic factors contribute to the vulnerability of communities to the risk of cholera in KwaZulu-Natal.

3.3.2.2   *Objectives (B2)*

To evaluate the dynamics of the 2000-2004 cholera epidemic in KZN, with respect to the natural environment, i.e. temperature, rainfall and humidity and the socio economic status of the communities in that region.
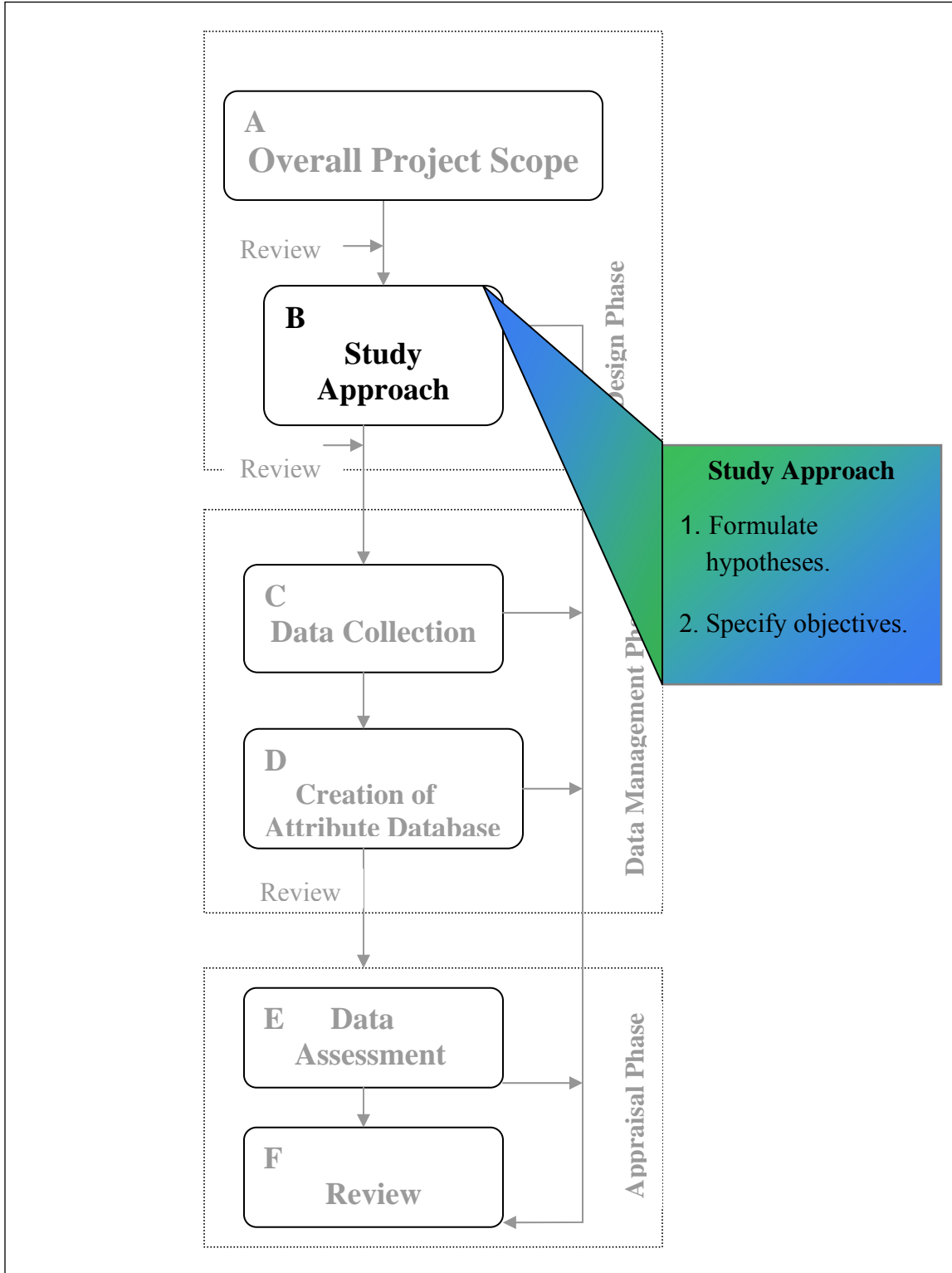
Fig 3.4: An illustration highlighting the tasks associated with the study approach.

33

Specific objectives

   a.   Investigate how factors like sanitation, water supply, population density; and other socio economic factors expose communities to the risk of cholera, through the use of retrospective data.

   b.   Use the available data as a management tool to identify the enabling factors that contribute to the spread of the disease as the basis for predicting future cholera outbreaks within the region.

   c.   Develop GIS maps that will spatially demonstrate how the important factors relate to one another in the spread of cholera.

3.3.3   Data Collection (C)

Even though the Cholera Database was the focal point of the study, the demographic, socio-economic and climatic data were just as important in the portrayal of a holistic epidemic picture. Hence, during this phase, the main research activities involved identifying the custodians of the data types required for the study and their subsequent acquisition. Basically though, data was obtained in order to obtain information that could be used in the management of cholera.

3.3.3.1   *The KZN Cholera Database (C1)*

The GIS Unit of the KZN provincial DOH in Pietermaritzburg established the surveillance of morbidity and mortality statistics of the 2000 - 2004 cholera epidemic. The aim of gathering this data was to convey an understanding of the distribution and status of cholera within KwaZulu-Natal (KZN Cholera 2000). Thus, this exercise of systematically collecting information on all the cholera cases reported at health institutions in KZN led to the establishment of the Cholera Database. The Cholera Database was vital to the study in that it provided a sound scientific foundation to study the cholera epidemic. A similar study in Delhi highlighted the usefulness of surveillance data to identify age groups, geographical localities and seasons with increased risk to cholera and to allow for focussed control measures (Singh *et al.,* 1998). The database has an
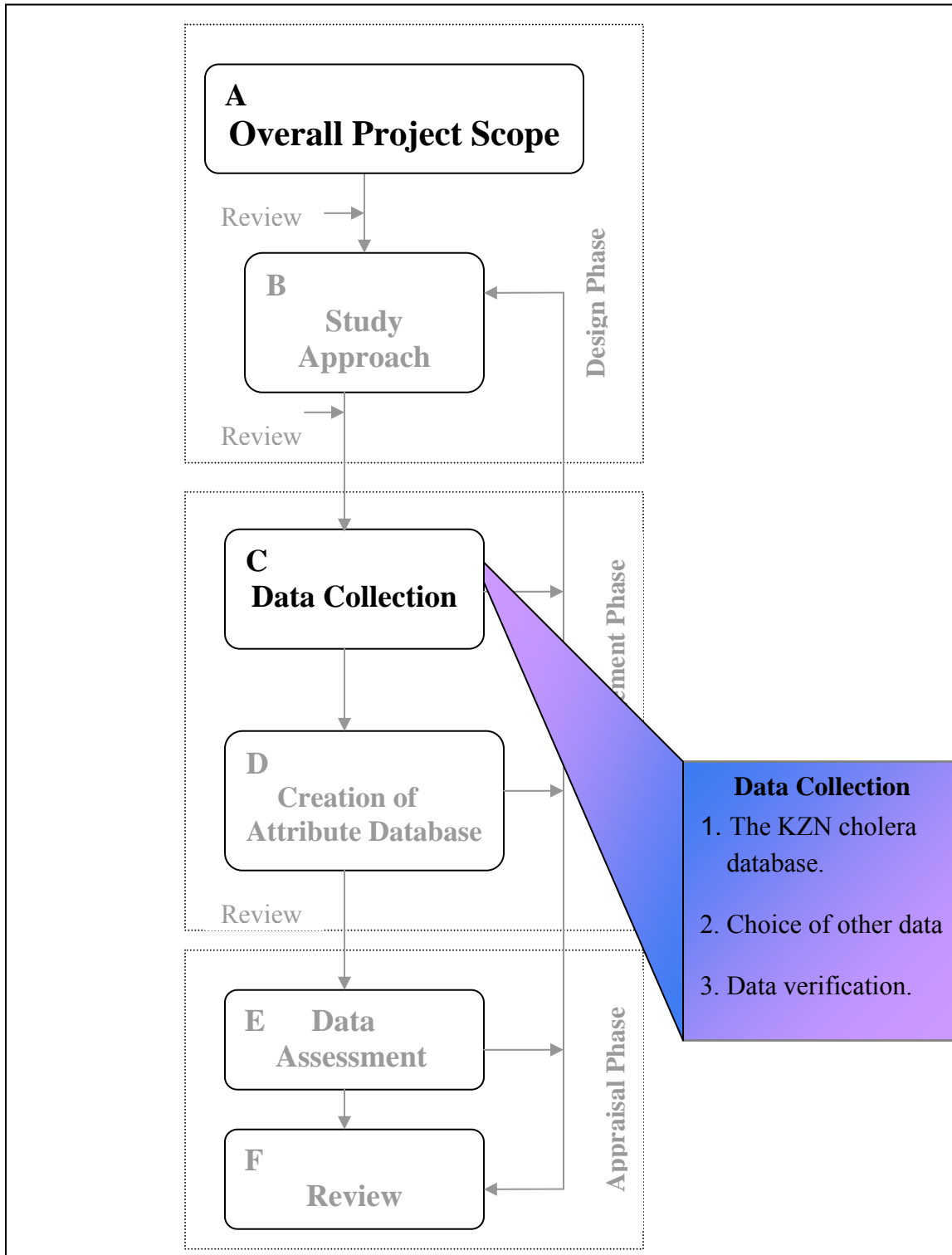
Figure 3.5: An illustration of the tasks assigned to the data collection & verification task.

Health Institution

"Line Listing Form" – patient information.

Fax form to Communicable Disease Control (CDC) Coordinator for the institution.

CDC Co-ordinators

Faxed form entered into the KZN Cholera Database developed by the GIS Unit, DOH, KZN. Database emailed to the centrally based GIS Unit. The GIS Unit assembles all the databases into one database.

GIS Unit

ArcView 3.2 GIS Software is used to produce many different kinds of maps based on the information captured into the database.

The database, maps and various reports are disseminated via email, CD, FTP and Internet
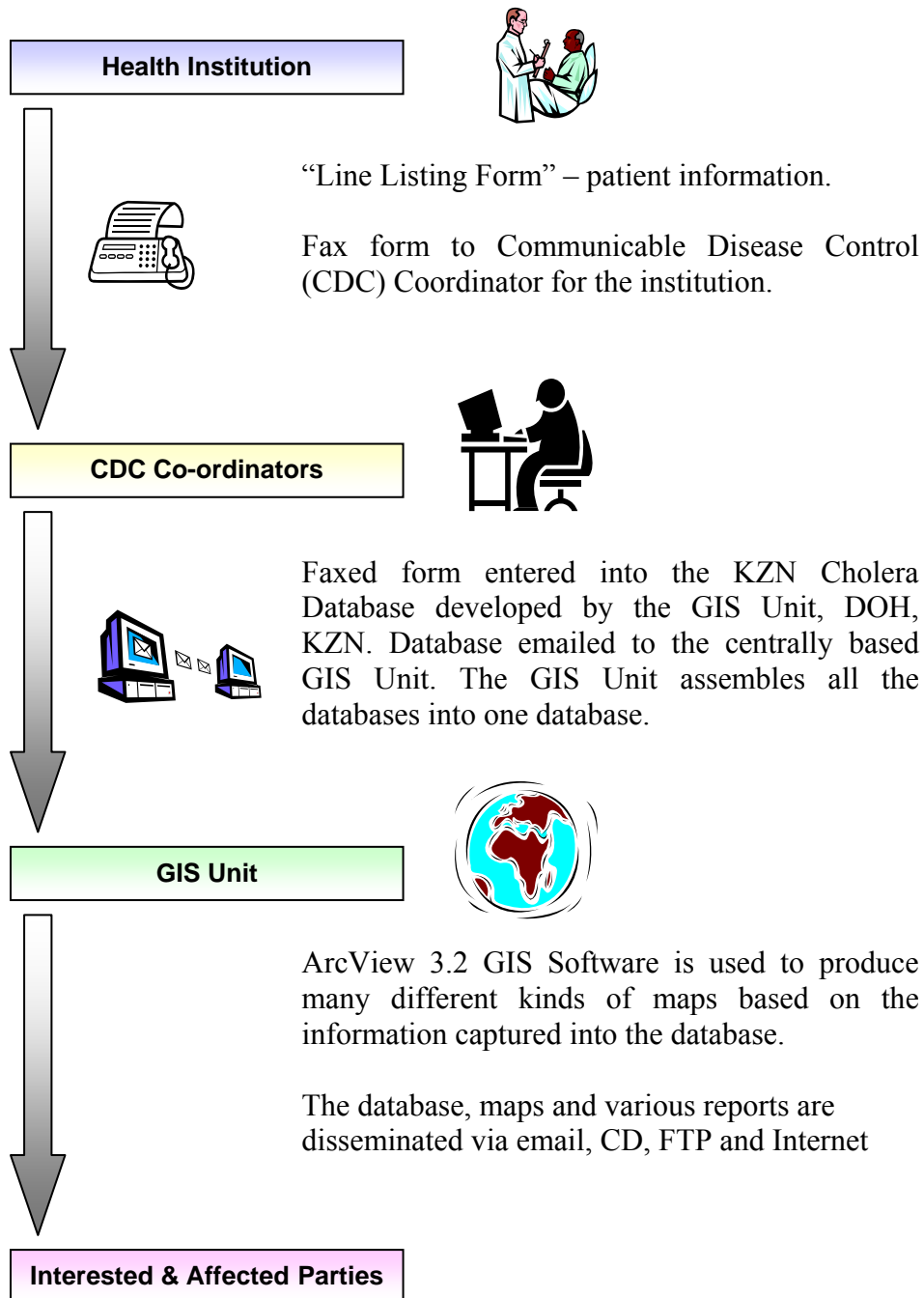
Interested & Affected Parties

Fig 3.6: An illustrative representation of how the Cholera Database was assembled.

Reproduced by permission from Shannon Rushworth, GIS Unit, DOH, KZN.

36

inventory of all the cholera patients registered at the various health institutions that are linked to the central database collecting system in KZN. This data collection network, which is an ongoing activity, is illustrated in Figure 3.6. Each cholera patient represented a data entry, as shown in the sample of the "line listing" form (Table 3.1). Information on the patient's age; gender; place of residence; the date they fell ill; and the hospital they reported to, were also captured in the database. The patient's personal information was therefore captured when they (patients) sought treatment (both on an in-patient and on an out-patient basis) at health centres. Some data entries had additional "follow-up" information collected on the progress of the disease, such as whether cholera was diagnostically confirmed and in terminal cases, the date when death occurred. Considering that data collection on cholera reports is an ongoing exercise, it is important to note that the Cholera Database referred to in this study, is meant to reflect the period from August 2000 to February 2004. Notwithstanding, this time frame afforded the study sufficient data entries to investigate annual trends between the years 2000 - 2003.

3.3.3.2   *Choice of data and selection of parameters (C2)*

In addition to the cholera data that formed the foundation of the study, there were other major factors considered in this study. These included the demographic, socio economic, geographical and climatic factors. All the data types collected, including the cholera data were processed to introduce a geo-spatial link using ArcView GIS 3.3.

Demographic and socio-economic data

Demographic data of KZN was important to the study in that it reflected the profile and the proportion of the resident population affected by cholera with the respect to the overall KZN population. While the socio economic aspects of the KZN population, like availability of piped water, sanitation services, source of energy; income and housing complemented the demographic profile of this population. All data under this category were sought from Census 1996 data from Statistics South Africa (STATSSA, 1998). A point to note is that the study used the 1996 census data for all its demographic and socio economic data. Several interrelated issues led to this decision. The cholera epidemic in question started in the year 2000 before the national census of 2001 was carried out. One of the main information types that the GIS Unit of the KZN

37

provincial DOH needed in the creation of the Cholera Database were all the geographical place-names. This was for the obvious reason that they will be able to identify at any time where cholera is reported from; as each patient will be required to indicate where they came from (the geographical place-name). As such the GIS Unit of the KZN provincial DOH used the census data that was available at the time as their source of place-names, which was Census 1996 (Personal communication). Thus the demarcated place-names as listed in Census 1996 are the same as those listed in the Cholera Database of August 2000 - February 2004. Albeit, during the course of this study, Census 2001 data did become available (in 2003), though there one major issue of concern. It became apparent that some of the magisterial demarcations, place-names, geographical codes etc. of KZN, had changed, and this would have provided confusion in correctly placing the cholera patients who had already been included in the Cholera Database thus far. As one newly demarcated place name may now be overlapping into more than one previously demarcated place-name(s). It can also be assumed that new cholera patients at the grass root level most probably still provided information of where they come from with the old place-names as they may have been oblivious of the newly demarcated areas. Thus to provide consistency with the Cholera Database, which had used the Census 1996 demarcations for communities, towns and village during the data collection exercise, it was decided that the Census 1996 dataset would also be used for all the other demographic and socio-economic data types.

Geographical data

The place-names, their P-codes (place codes), magisterial districts and district councils represented geographical parameters. The source of the data on these parameters was the Cholera Database itself, making their retrieval straightforward. This information was documented on the "Line listing form" that was used to record the required information on all the patients presenting with cholera at health facilities in KZN. The surface area of place-names in square kilometres was sourced from Census 1996 data (STATSSA, 1998) while their elevation above sea level was processed using ArcView GIS 3.3. The types of water resources (perennial rivers, dams, pans, wetlands and canals) were sourced from Environmental Potential Atlas (ENPAT, 2001).

Climatic data

As previously explained, the notion behind the inclusion of climatic data is based on the climatic environment in KZN as a possible major driver of *V.cholerae* transmission. Thus, the hypothesis that, climate probably plays a pivotal role in the transmission of *V.cholerae* in KZN. Climate data collected over the same period as that of the cholera epidemic, i.e. from August 2000 to February 2004 was requested from the South African Weather Service (SAWS). The SAWS is the national organisation responsible for the provision of meteorological data, and the custodian of such data at the national level. The climatic data included minimum, maximum and average temperature, rainfall, and humidity.

The unprocessed climatic data of all climatic parameters mentioned above, was made available (from SAWS) as data collected on a daily basis. Considering the study had envisaged that the climatic variations would be assessed on a monthly basis, the daily climatic data was therefore processed to give monthly averages. Thus, the climatic data was summarized per weather station to include the average relative humidity per month, average temperature per month, average minimum and maximum temperatures per month and average rainfall per month. The calculations to process the average monthly climatic data were performed in MS Excel 2000. The spatial analysis tool in ArcView GIS 3.3 was used to interpolate the various data points and eventually geo-process them to add valid maximum, minimum and average monthly temperatures, average monthly humidity values and average monthly rainfall numbers to each place-name polygon. Thus, every weather station was assigned a set of values for the period August 2000 to February 2004 on a monthly basis. To derive seasonal weather data for the individual DCs in KZN, each place-name was allocated to a DC and thereafter summarised to display seasonal weather data for each DC in KZN.

3.3.3.3 Data verification (C3)

It was important to inspect the databases and confirm that all the relevant information was present in every data entry. In addition, information details such as geographical names and age entries had to be checked for possible spelling mistakes and/or other errors before any data

analyses could be undertaken. Most importantly, the data verification exercise ensured that the results and conclusion drawn from the data were a true interpretation of it.

Incomplete Cholera Database entries

One of the main shortcomings of the Cholera Database was incomplete data entry. Although the Cholera Database recorded a total of 158 896 cases, only 136 262 could be included in the study. The remaining 22 634 cholera case records had one or more data entries missing, thus could not be used in the study. The incomplete entries made up 14.2% of the Cholera Database. Entries with incomplete information had one or more aspects of the patient personal information omitted. Missing information was encountered under all the variables i.e. age, gender, place name, p-code, and date of notification. An extract of the "line listing" capture form with its comprehensive list on patient information is shown in Table 3.1. Among the omitted parameters, the p-codes and the GIS place-names were the most common, followed by age and gender respectively. A total of 4 825 (3%) cholera case entries from the original 158 896 of the Cholera Database had both the place-name and p-code information missing. In the event that either the p-code or the place name was available, some of the data entries could be salvaged, as the p-code could be linked to a place name and vice versa. However, if all the geographical parameters were missing, it made placing of data entries into their respective DCs and MDs particularly difficult. Data with missing information on age, gender and/or date of notification could not be salvaged as such, as this meant tracing patients through hospital records to verify the information. An exercise, that was beyond the scope of the study.

Table 3.1: Extract of a "line listing form" used to capture information on cholera patients.

Age (Yrs)	Gender	Area Patient Sickened	GIS Place Name	Date of Notification	Admit Yes /No	Institution Admitted to	Stool Result	Died Yes /No	Death Date
23	Male	Ezitendeni	Weenen NU	2001/12/20	Yes	Estcourt Hospital	Not Taken	No	
36	Male	Mngwenya	Weenen NU	2001/12/19	Yes	Estcourt Hospital	Not Taken	No	
60	Female	Ngodini	Weenen NU	2001/12/12	Yes	Estcourt Hospital	Not Taken	No	

A
Overall Project Scope

Review

B
Study Approach

Design Phase

Review

C
Data Collection & Verification

Management Phase

D
Creation of Attribute Database

Review

Creation of Attribute Database

1. Selection and categorisation of parameters.

2. Compilation of the Attribute Database.

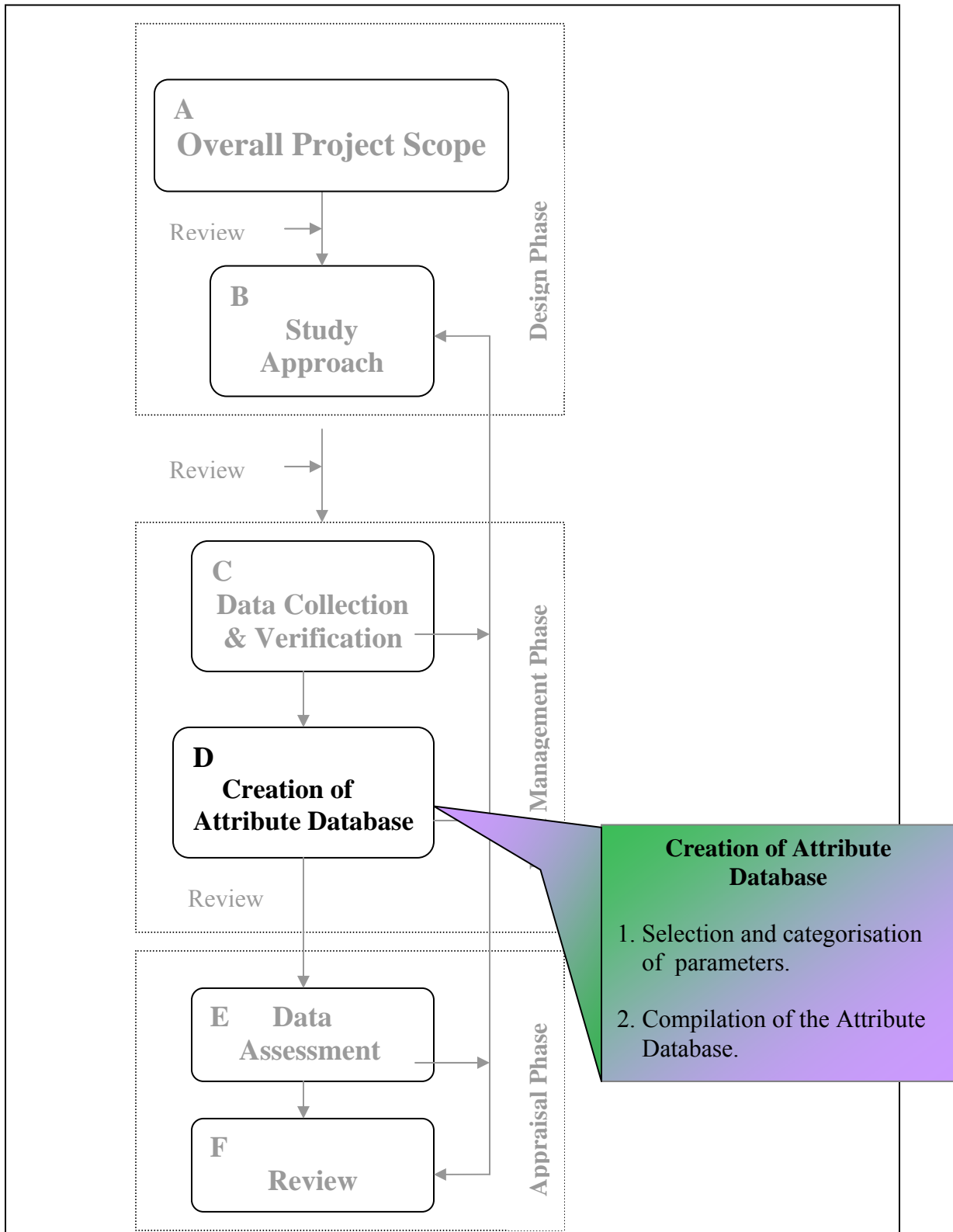E   **Data Assessment**

Appraisal Phase

F
Review

Figure 3.7: An illustration of the assigned tasks that lead to the creation of the study database.

41

3.3.4 Creation of the Attribute Database (D)

The reliability and compatibility of the data sources was of fundamental importance. This was particularly important because eventually all data types collected were reconciled and merged to assemble an all-inclusive study database, hereby referred to as the Attribute database. The exercise thus culminated in the generation of a study database that was used for all the graphic and spatial representations, and as well as for statistical evaluation. The choice of the parameters from each of these groups is explained below.

3.3.4.1 Selection and categorisation of parameters (D1)

Once the type of parameters relevant for the study were identified and subsequently verified, categorisation of the data types followed. This is because the magnitude of the database dictated that some of the variables within the different groups of data be categorised for ease of assessment. The types of variables in each data group follows, and where categories were introduced, the rationale behind it is provided.

Cholera data

The creation of the Cholera Database is already explained in 3.3.3.1. The categories formed for each of the demographic, socio-economic and climatic variables were independent of each another. The categories directly related to the cholera cases are explained below:

i. Cholera per age groups

The cholera database used in this study had 158 895 entries, each representing the personal record of a cholera patient. The first exercise that was undertaken was to sort the database by age into age groups compatible and comparable with the demographic data of Census 1996, which had age presented into groups of 5-year intervals thus cholera entries by age followed the same grouping starting from 0-4, 5-9,10-14 up to and above 100 years. The basis for keeping consistency with this age grouping was to be able to subsequently perform age related comparisons. One difference worthy of note was that the Census 1996 age groups gave a count

that was the sum of both the genders while the cholera data had an option of presenting the cholera cases by age groups of individual genders or as a sum of both.

ii   Gender

The Cholera Database specified the gender of each cholera patient. Thus, for spreadsheet analysis and graphical representation of the data, the genders could be analysed separately. For statistical and spatial analyses though, correlations were based on age groups that represented both the genders. This is because the study did not have Census 1996 population data that had further divided the age groups according to gender.

iii   Calculation of Percent Cumulative Infection Rate (%CIR)

The %CIR represents the infection rate per 100 individuals in a defined population. Statistically, the cumulative infection rate is the risk of individuals in the population getting the disease during a specified period (Beaglehole *et al.,* 1993). In this study, the %CIR was used as a measure of vulnerability to the incidence of cholera. The %CIR was used to standardise the infection rates for each DC and MD in that it reflected the percentage of the population in question that got the disease and not the actual number of cases. This also ensured that the comparisons and correlations based on the different MDs were standardized.

 Thus the %CIR was calculated using the formula outlined below by Beaglehole *et al.* (1993):

CI = <u>   Number of people who get a disease during a specified period   </u>   x 100

    Number of people free of the disease in the population at risk during a specified period

Demographic indicators

Demographic profiles were used to investigate their broader relationships with the cholera epidemic. In this case the demographic indicators were those from the Census 1996 database as already explained in 3.3.3.2. These included age and gender. Subsequently, the corresponding

cholera cases were sorted likewise by similar age groups and gender. This exercise was done using Pivot tables in MS Excel (Refer to Annex A- Methodology).

i   Age

There were 21 sets of age groups at five year interval starting from 0-4, 5-9, 10-14, up to +100 years. For statistical analysis, the age groups were sorted to give nine categories, to represent the different phases an individual normally goes through from birth to adult life. The basis for each of the following categories is explained below:

- Age 0-4 years

  This group represents the paediatric age, a phase during which an individual is susceptible to diseases because of their underdeveloped immune system.

- Age 5-9 years

  The pre-adolescence stage is characterised by individuals who are starting to attend formal education. They are semi independent and their susceptibility to diseases is dependent on the environment around them.

- Age 10-14 years

  This is the age of puberty where adolescence starts to set in. Individuals in this age group notice physiological changes as they progress through the puberty stage. There may also be psychological changes as individuals start perceiving themselves as semi-independent young adults.

- Age 15-19 years

  The mid to late teenage group contributes significantly to the reproductive base of expansive populations such as the one of KZN (MDG, 2004, PRB, 2005).

- Age 20-29 years

  Such individuals are considered to be adults. Unless immuno-compromised by chronic disease states, normal individuals of this age group have a healthy immune status operating at its peak. This age group contributes the most to the reproductive base of an expansive population, such as that of KZN. They are also the caregivers, especially the

females. Such responsibilities constantly expose them to a high risk of infection as they take up the responsibility of looking after the young and the sick.

- Age 30-39 years

  Most of the productive/active work force belongs to this group. It is also quite possible that this group include the many individuals that leave their homes to become migrant workers elsewhere. The state of being migratory means that their immune systems are constantly challenged by new infections, with the possibility of being overwhelmed at one time or another (Wilson, 1995a).

- Age 40-59 years

  The middle age group would normally have an established family set up and in most cases expected to have a steady income.

- Age 60-74 years

  Retirement starts at this age group. Health also becomes an ongoing concern as the immune system of such individuals start waning.

- Age 75- over 100 years

  This age group comprises the elderly. These may be nutritionally disadvantaged in poor communities. Their immune system has weakened with age and they are thus susceptible to a wide range of illnesses.

Socio economic indicators

The indicators in this category, include types of dwellings, income and services such as clean water supply, sanitation, refuse collection and electricity. The groups and their resultant categories, if any, are explained in the following sub sections. Most of the explanations are in whole or in part as put forward by the Census 1996 report (STATSSA, 1998).

i. Water services

Issues surrounding availability and usage of water are always a point of focus when it comes to cholera epidemics. Therefore, the information on types of water services as listed in Census 1996 database were not grouped into categories. It was considered important that each water

service/utility be assessed and correlated independently to the incidence of cholera. There were seven distinct types of water services and utilisations listed. A brief explanation of each is given below:

- Piped water in dwelling

Listed under this group are households with piped tap water in their dwelling. The assumption here is that the water they received was treated using conventional water treatment methods before reaching them. The group also includes those who had the same service but were accommodated in hostels or other types of communal dwellings.

- Piped water on site

    Households listed under this group received water from a standpipe, which was installed within their yard. It also included individuals who had the same service, but resided in hostels or other types of communal dwellings.

- Public tap

    As the name suggests, households who had this service made use of taps shared with other households/residents in their communities. Public taps were also assumed to deliver clean treated water to the residents they served.

- Water-carrier/tanker

    Water tankers are carrier trucks that carry and distribute water in areas where there is a shortage of piped water services. The water brought in by such tankers is clean treated water. This service is more often than not, run on a commercial basis. It is thus only available to those who can afford it, unless it is a special extension of water services to communities by their local authority. Especially in emergency situations like droughts, floods and epidemics.

- Borehole/well/ rain-water tank

    Households under this group mostly use water from underground sources or that, which is collected from precipitation. These sources suggest no water treatment takes place except perhaps by natural. Borehole and well water undergo natural filtration through soil layers to considerably reduce the microbial load. Well water may however, be polluted from surface contaminants if left exposed or due to surface run-off during heavy

rains or floods. Harvested rainwater often carries high contamination loads depending on how it was collected and the vessels used for storage. Most often rainwater does not meet drinking water standards, especially with respect to the bacteriological water quality (Gould, 1999).

- Dam/river/stream/spring

Under this option are households that make use of surface water sources that do not undergo natural filtration or chemical treatment. Depending on the locality and the associated land use activities in the area, such as agriculture, industry and domestic use, such water sources usually carry high microbial loads.

- Other water sources

Households listed here use water sources that are not specified. It may be that they use all or any of the above listed water services/sources whenever and wherever they become available to them.

ii. Sanitation services

Sanitation parallels water as a necessary basic service required within a community. With respect to this study, sanitation refers to the development and application of measures for the safe collection, transportation, treatment and disposal of human wastes. Sanitation is an important public health measure essential for the prevention of infectious diseases. The types of sanitation listed by STATSSA (1998) during the Census 1996 data collection are thus explained:

- Flush toilet

This group of households had flush toilets that made use of water cisterns. This meant that the households were connected to a piped water system that supported the operation of such toilets. In addition there must have been a sewer system supported by the local authority or individual sewer pits for the safe disposal of sewage.

- Pit latrine

Pit latrines, once built, are a low maintenance sanitation option. The system is as the name implies, a pit dug in the ground that serves as a semi permanent sewage disposal unit. The pit latrines considered here are of the VIP type (Ventilated Improved Pit), with

47

a vent pipe to take odours away from the enclosure of the latrine. This vent pipe is also covered with a mesh filter to prevent flies from getting in.

- Bucket toilet

  This option offers households the use of a bucket as a disposal unit that has to be frequently emptied by the local authority. It is a high maintenance system and in the event that the local authority fails to collect and empty the bucket, the service is easily compromised.

- Other types of toilet

  Households included under this group did not have any of the sanitation options listed above but could have used other rudimentary options.

iii. Refuse disposal

In the context of this study, refuse disposal is the collection and disposal of solid wastes/rubbish (excluding sewage) of a household.

- Refuse collected once a week

  The local authorities removed refuse at least once a week.

- Refuse collection irregular

  Refuse was removed by local authority less often and on an irregular basis.

- Refuse disposal in community dump site

  Refuse was disposed at a dumpsite used by several households of a particular community.

- Refuse disposal in own dump site

  This group of households disposed of their refuse in their own dumpsites, implying the dumpsite must have been located in an area the household had authority over.

- No disposal of refuse

  Most probably this group refers to households that had no particular area assigned for rubbish disposal.

- Other types of refuse disposal

  Though unspecified, the households included here may have used other elementary type of refuse disposal such as incineration.

iv. Income groups

Income is defined as all money received from salary, wages or own business, as well as money from additional work activities, remittances from family members living elsewhere, pensions or grants and income from investments (STATSSA, 2004). The Census 1996 socio-economic data had a total of 16 income groups, including the "None income" group. Subsequently these income groups were clustered into eight categories. The income of the different groups below is given in South African currency (Rand) per month.

- No income

  According to the definition given by Census 1996, the people under this group were considered those who were employable but did not have the opportunity to earn an income through formal employment.

- Low income

  The group of individuals earning between R 1-R 1,500.

- Medium income

  The group of individuals earning between R 1,501-R 6,000.

- High income

  The group of individuals earning between R 6,001-R 16,000.

- Highest income

  The group earning between R 16,001 to above R 30,000.

- Unknown income

  There was no information available about the net income of this group.

v. Household size and type

The type of dwelling refers to the various structures used for accommodation. Such structures include houses, traditional houses, townhouses, flats/apartments, hostels, huts, informal dwellings such as shacks, semi-detached houses, etc. These were grouped into four categories.

- Traditional houses

  A traditional dwelling is one made of clay, mud, thatch or other traditional materials. Traditional dwellings may be found as single units or in clusters.

49

- Conventional house

  All the house types built of standard brick and mortar were grouped under this category. This included houses, townhouses, flats/apartments and semi-detached houses.

- Informal

  Informal dwellings included shacks, whether in backyards or in squatter settlements. These types of dwellings are made out of basic materials like poles, cardboard and plastic covers for protection against the elements.

- Temporary shelters

  Hostel, caravan/tents were considered as temporary shelters, in that the residents had a transitory occupancy of the accommodation.

- Homeless

  Homeless persons are those who had no form of shelter and no known address. Such persons are typically found spending nights on street pavements, in alleyways, or at railway stations to name a few.

vi. Energy source

The types of energy sources considered here are those used for domestic purposes such as cooking, heating and lighting. There were a total of eight sources of energy groups listed in the Census 1996 database. These sources were subsequently grouped into three categories:

- Electricity

  Electricity supplied directly by the municipality/local authority or electricity from another source, e.g. generators, solar cells.

- Petrochemicals

  The energy sources making up this group are of petrochemical nature and these include gas, paraffin and candles.

- Other energy sources

  These energy sources included coal, wood and animal dung.

vii.   Health facilities

Information on the health facilities in KZN was provided together with the Cholera Database. A total of 1 101 health facilities were listed by type and by their administrative authority.

a.   Health facilities by authority

The administration of the health facilities fall under the authority of any one of the following:

- Provincial

  Provincial health facilities get their funding from the provincial government budgets.

- Local authority

  Health facilities that fall under this category are clinics funded by their municipalities.

- Private facilities

  Many of these hospitals are owned and managed by a consortium of private physicians or by large business organisations. Private hospital fees are generally higher than those of provincial hospitals.

b.   Health facility type

The health facilities were classed according to the type of service they rendered to the public. The definitions of the different types of health facilities found in KZN, are in accordance with the National Department of Health (DOH, 2003-b).

- Hospital

  A public health facility, which receives referrals from and provides generalist support to clinics and community health centres with health treatment, administered by general health care practitioners or primary health care nurses (KZN-DOH, 2000).

- Clinic

  A public health facility at and from which a range of primary health care services are provided and that is normally open eight or more hours a day based on the need of the community for such services. There are no facilities to admit a patient on an in-patient basis (KZN-DOH, 2000).

- Mobile base

  A mobile clinic is a temporary point of service providing primary health care (PHC) services. As it can be moved, it is set up for the convenience of the patients (DOH, 2003).

- Satellite clinic

  Such a health facility represents the main clinic of a particular area. It is usually placed at a distance from the main clinic to serve the communities that cannot easily access the main clinic.

Climatic data

To make the weather station data compatible with the study database, GIS interpolation and calculations was used to derive a monthly value to each place-name (P-code) as explained in 3.3.3.2. This means prior to these calculations, the climate data from the weather stations did not give specific data for each place name but the general climatic data for a specific area that represented many place-names. In addition to monthly climatic variables, seasonal extrapolations were also performed. Summer months were considered to be from December to February; autumn; from March to May; winter; from June to August and spring; from September to November. The calculation of average and seasonal climatic data was also for the purpose of introducing consistency within the Attribute database. This would then allow monthly cholera cases to be compared with monthly climatic variables and seasonal cholera cases to be compared with seasonal climatic variables. Below is a brief explanation of the individual climatic variables used in the study.

i. Temperature

Temperature is measured in degrees Celsius. The SAWS weather stations recorded minimum, maximum, and average temperatures on a daily basis.

- Minimum temperature

  The lowest temperature recorded in a 24 hour period is referred to as the minimum temperature for that day. Averages of monthly minimum temperatures were calculated

by adding all the minimum temperatures recorded for the month, divided by the number of days the minimum temperature readings were taken for that month.

- Maximum temperature

The highest temperature recorded in a 24 hour period is referred to as the maximum temperature for that day. Averages of monthly maximum temperatures were calculated by adding all the maximum temperatures recorded for the month, divided by the number of days the maximum temperature readings were taken for that month.

- Average temperature

This is calculated as the mean of the maximum and the minimum temperatures. Averages of monthly average temperatures were calculated by adding all the average temperatures recorded for the month, divided by the number of days the average temperature readings were taken for that month.

Table 3.2:   The definition of temperature ranges in South Africa as classified by the South African Weather Service.

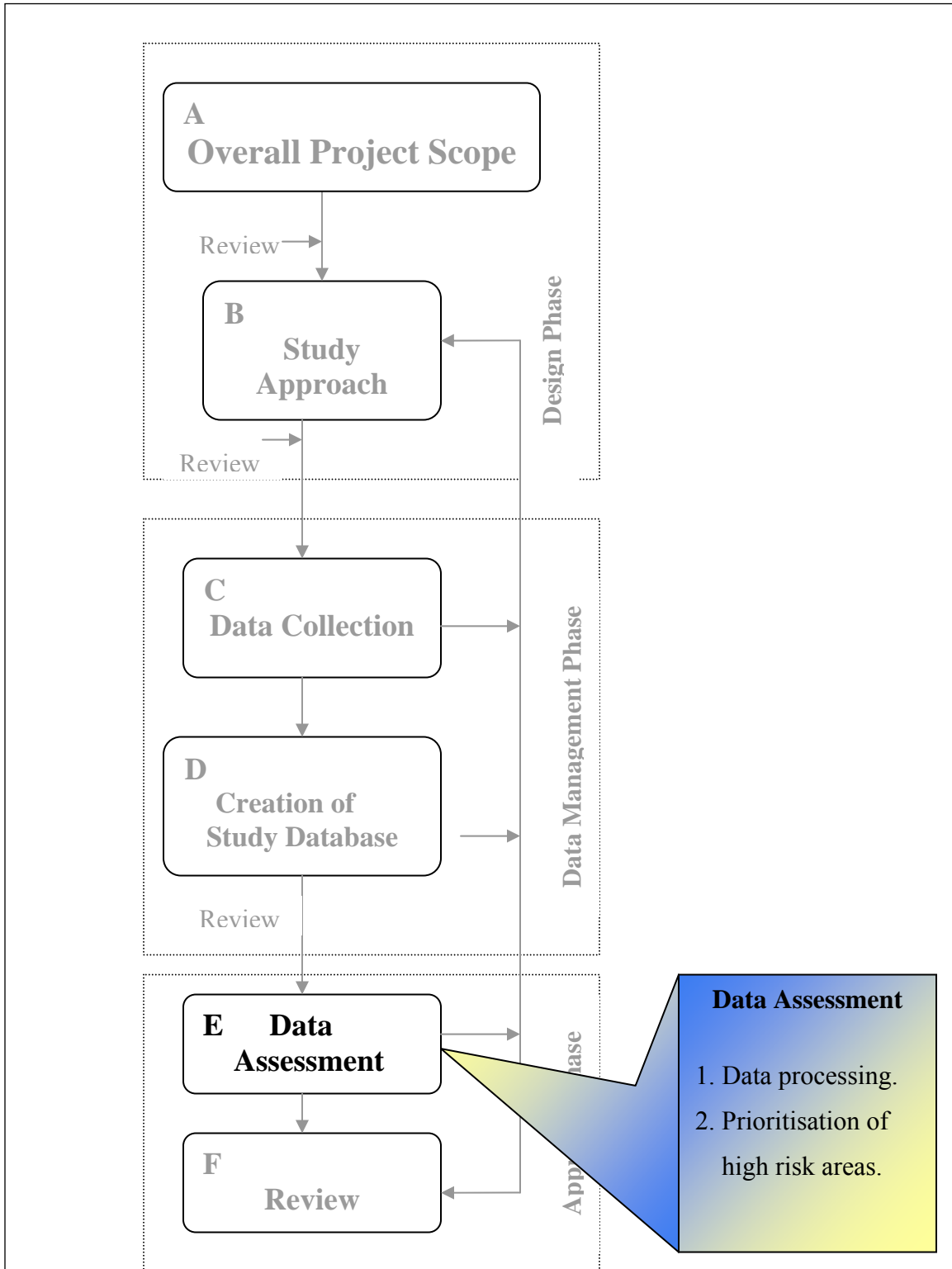| Description	Summer (Oct – Mar)	Winter (Apr - Sep)
Very hot	>35 ^{o}C	>35 ^{o}C
Hot	30 - 34 ºC	27 - 34 ºC
Warm	26 - 29 ºC	23 - 26 ºC
Cool	20 - 25 ºC	-
Mild	-	17 - 22 ºC
Cold	15 - 19 ºC	12 - 16 ºC
Very Cold	<15 ^{o}C	12 ^{o}C

Figure 3.8 Schematic illustration of the tasks assigned to the overall assessment of the data.

The temperature values in the Attribute database represent the average monthly minimum, maximum and average temperatures. Though the values used to determine the possible relationships between temperature and the incidence of cholera was that of maximum temperature.

ii. Rainfall

Rainfall can be defined as more or less continuous precipitation measured in millimetres. Precipitation includes all of the forms of water particles, whether liquid or solid, that falls from the atmosphere and reach the ground. Values representing average monthly rainfall were the ones incorporated in the Attribute database, thus used throughout the study while investigating the possible relationships between rainfall and the incidence of cholera. Averages of monthly rainfall were calculated by adding all the rainfall values recorded for the month, divided by the number of days the average rainfall readings were taken for that month.

iii. Humidity (Relative humidity)

This is the ratio of the amount of moisture in the air to the amount that the air could hold at the same temperature and pressure if it were saturated, usually expressed as a percent. The humidity values incorporated in the Attribute database represent average monthly humidity. Averages of monthly humidity were calculated by adding all the humidity values recorded for the month, divided by the number of days the humidity readings were taken for that month.

Geographical data

Geographical parameters were not grouped into specific categories. Their categorisation only came in during data processing and analyses, i.e. whether the data is evaluated according to DCs (District Councils), MDs (Magisterial Districts) or individual place-names.

3.3.4.2   *Compilation of the Attribute database (D2)*

The creation of the Attribute database was made possible once all the variables described thus far, were collected, verified and in some cases categorised. This means the three databases of

cholera cases, climatic, demographic and socio-economic variables required to be reconciled before being merged into a single Attribute/study database. The success of the creation of the Attribute database exercise required that each place-name be linked with a full set of variables across all the three databases mentioned above. The magnitude of the study database can be describes as follows: There were 2 741 place-names listed. Each place-name was accompanied by 135 variables related to the cholera case profiles, i.e. cholera by place name, by age groups and by gender. There were 34 variables related to demographic profiling for each place name, i.e. the size of the population and the associated age groups and gender. A total of 68 types of socio-economic variables including those related to water, sanitation, refuse, income and energy options were available for each place name together with 11 types of geographic variables of topology and natural water bodies and 216 monthly climatic variables that included rainfall, humidity and temperature data. This meant that each place-name had a total of 464 variables geo-spatially linked to it. Thus in its entirety, the Attribute Database was made of a spreadsheet containing 2741 rows and 464 columns of data entries, therefore a total of 1 271 824 cells with data.

3.3.5   Data assessment (E)

The fact that the study is of a multi factorial nature, each variable was assessed on its own as well as in relation to other variables. Data analyses were performed in terms of how the epidemic correlated with different environmental factors; e.g. temperature; rainfall; humidity versus cholera incidence; and which socio-economic factors like water; sanitation; housing; refuse collection and income may have facilitated the spread of the outbreak.

3.3.5.1   *Data processing (E1)*

Data was processed using Microsoft Excel 2000, ArcView GIS 3.3 and SAS statistical computer software. The study made use of these research tools to analyse and interpret the Cholera Database as well as the Attribute Database. The initial set of analyses served to study the general trend of the 2000-2004 cholera epidemic. These preliminary analyses used spreadsheets to explore the Cholera Database, whereby results were presented in the form of tables and graphs.

Later on, once the creation of the Attribute Database was complete, more involved statistical analyses and spatial representations were performed. As such the available data was looked at from three different perspectives, vis-à-vis the general, the statistical and the spatial approaches.

The general approach

The following spreadsheet analyses using MS Excel (2000) were performed. The results from this exercise are presented in the form of line graphs and bar charts in Chapter 5.

1. Cholera according to DCs.
2. Cholera according to MDs.
3. Cholera trend from 2000-2004.
4. Cholera cases and gender.
5. Cholera according to gender and age group.
6. Cholera according to DCs and %CIR of cholera

Statistical approach

Once the Attribute Database was created, it was then considered as unprocessed data, from which derived values were obtained. Derived values were obtained through statistical data processing using SAS® (version 8.2). The following is a list of derived values calculated from the Attribute Database based on the DC totals of KZN.

1. Population proportions of DCs for the total population of KZN.
2. Cholera case proportions according to DCs.
3. Cholera death proportions according to DCs.
4. Age group category proportions according to DCs.
5. Cholera age group category proportions according to DCs.
6. Mean monthly humidity according to DCs.
7. Mean monthly minimum temperature according to DCs.
8. Mean monthly maximum temperature according to DCs.
9. Mean monthly rainfall according to DCs.

10. Mean seasonal humidity by DC.

11. Mean seasonal minimum temperature by DC.

12. Mean seasonal maximum temperature by DC.

13. Mean seasonal rainfall by DC.

14. Household type category proportions by DC.

15. Household number proportions by DC.

16. Mean number of people per household per DC.

17. Water services proportions by DC.

18. Sanitation services proportions by DC.

19. Health facility number proportions by DC.

20. Mean number of people per health facility per DC.

21. Health facility category proportions by DC.

22. Natural water source category proportions by DC.

23. Natural water source category by number by DC.

24. Cholera cases per month per year proportions by DC.

25. Cholera cases per month per season proportions by DC.

26. Energy source proportions by DC.

27. Refuse services proportions by DC.

28. Income category proportions by DC.

29. Monthly mean % CIR according to DC.

30. Mean % CIR according to age group categories of DCs.

The following is a list of derived values calculated from the Attribute Database based on the MD totals of KZN:

1. Population proportions according to MD

2. Cholera case proportions according to MD.

3. Cholera death proportions according to MD.

4. Household type category proportions according to MD.

5. Household number proportions according to MD.

6. Mean number of people according to household according to MD.

7. Water services proportions according to MD.

8. Sanitation services proportions according to MD.

9. Health facility number proportions according to MD.

10. Mean number of people according to health facility according to MD.

11. Health facility category proportions according to MD.

12. Energy source proportions according to MD.

13. Refuse source proportions according to MD.

14. Income categories proportions according to MD.

15. Mean %CIR according to MD.

16. Mean % CIR according to age group categories according to MD.

These derived values formed the basis of the statistical correlations to investigate which variables statistically correlated with the incidence of cholera as outlined in 4.3.5.2.

Spatial approach

The outputs from both spreadsheet and statistical processing supported the production of GIS maps using GIS ArcView 3.3 Software. GIS maps presented a spatial picture of the different characteristics of the cholera epidemic. As such, the following spatial analyses were performed and the associated GIS maps generated.

- The creation of basic maps to spatially depict housing, income, population density, water and sanitation services in KZN and how they correlate with the incidence of cholera.

- The creation of maps to show how the climate of KZN spatially correlated with cholera incidence during the major peak of the epidemic from November '00 to January 01.

Statistical correlations

Partial Spearman's correlation was used in the statistical analysis of the data as all the variables in the different categories were used as proportions, which implies that the data was continuous. By performing partial correlations, a variable in a particular category was assessed while keeping

the effect of all the other variables in that category, constant. For example, in the household category, there are 5 household type variables. Thus, when performing a partial correlation of traditional houses with %CIR of cholera, the other 4 household types variables were kept constant, as a household type was a proportion of the entire category considered.

 Partial Spearman's correlation analyses were performed in two parts. The first part dealt with all the variables that had a once off, data value. This refers to the data presented as a single value representing a particular collective time period e.g. number of households with piped water services during the epidemic period in question. Thus, such a value was statistically assessed versus the percent cumulative incidence rate (%CIR) of cholera in the context of the entire epidemic period. The second part statistically assessed the %CIR of cholera with variables whose data was collected in a continuum, as was the case with the climatic variables that were collected on a daily basis. As already explained in 4.3.4.1, climatic data was extrapolated to give monthly and seasonal averages. This meant that the monthly %CIR of cholera in an area (e.g. in a MD) could be correlated with the climatic variables on a monthly or seasonal basis.

The preliminary assessment of the Cholera Database did point out the fact that although the cholera case reports were collected from August'00 to March'04, the majority of the cases were experienced in the year 2001. Thus to insure that the cholera cases reported in the Cholera Database prior to and after 2001 would not introduce bias in the statistical correlations, a decision was made to propose a minimum cut off value for the monthly %CIR of cholera in an area (place-name). That meant that for a place name to be included in the monthly calculations of %CIR of cholera under its MD, it had to qualify by having the proposed minimum cut off value for the monthly %CIR of cholera for every individual month of the epidemic under study. Thus a place-name will only be included in its MDs monthly %CIR calculations for the months it individually qualifies with the minimum monthly %CIR cut off value, and left out in the calculations in the months that it did not qualify.

60

Table 3.3 lists the monthly %CIR for each of the 52 MDs of KZN. The %CIR value represents the average of the monthly %CIR of cholera of that particular MD; which was calculated from the %CIR of cholera of each place-name (town, village etc) under that particular MD, divided by the total number of place-names under that MD. The value 0.082 was the lowest monthly %CIR of cholera value (Paulpietersburg) of all the MDs. Thus, the lowest %CIR of cholera value of 0.082 was proposed as a cut off value that the places that make up the MDs had to have. This meant that the monthly %CIR of cholera of each place name must be 0.08 or above to qualify for inclusion in the monthly climatic (rainfall, maximum temperature and humidity) correlations of the MDs to the %CIR of cholera. This cut off point was justified considering %CIR 0.082 of cholera translates to 0.82 person or approximately one person in a population of 1000 will get cholera, which is a high enough risk for transmission of cholera in a community. Thus as a prerequisite, there should be at least 1 case of cholera in a community (place-name) in a particular month, for the place to be included in the monthly correlation analyses of the monthly MD climatic variables against the %CIR of cholera (Figure 3.9).
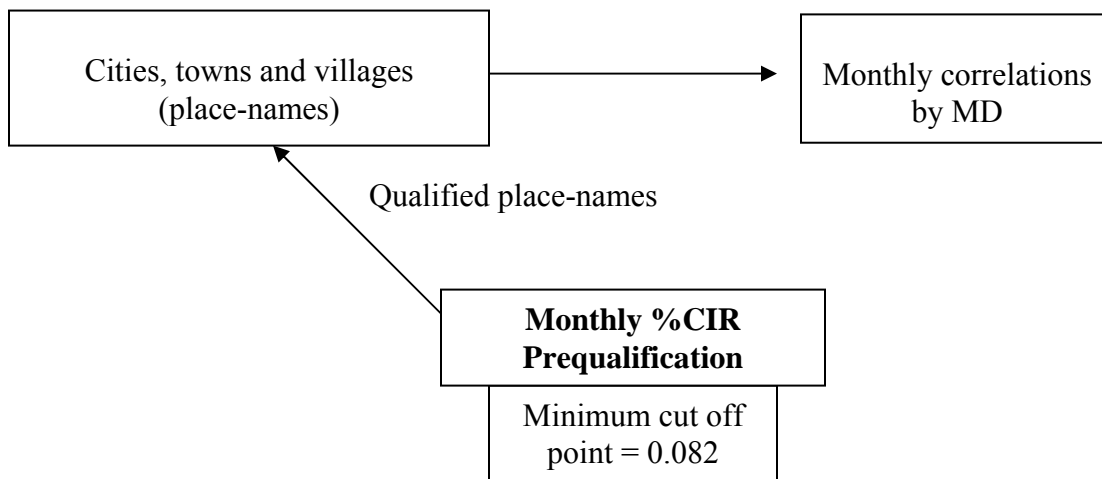
Figure 3.9: Flow diagram to show the selection criterion for place-names to qualify
for the monthly climatic correlations variables-cholera correlations.

GIS output

Parameters, which were statistically proven to have a positive correlation to the incidence of cholera were, selected to portray the possible spatial relationships. Thus the GIS maps of the selected parameters were either reconciled and merged with the cholera cases reported over the entire epidemic period or the cholera cases within a specified reporting period, as was the case with the climatic variables (Maps 5-26).

Table 3.3: The mean % CIR of cholera of the Magisterial Districts in KwaZulu-Natal.

MD	N	MPCIRN	STD	STDERR	VAR	MIN	MAX	MEDIAN
Alfred	69	0.307	0.812	0.098	0.659	0.000	4.865	0.053
Babanango	25	0.117	0.150	0.030	0.023	0.000	0.615	0.055
Bergville	37	0.442	0.733	0.121	0.537	0.000	3.494	0.163
Camperdown	31	1.086	2.752	0.494	7.572	0.000	13.333	0.114
Chatsworth	18	0.124	0.253	0.060	0.064	0.000	0.962	0.012
Dannhauser	33	0.232	0.448	0.078	0.201	0.000	2.300	0.081
Dundee	31	0.162	0.287	0.051	0.082	0.000	1.389	0.057
Durban	96	1.223	5.820	0.594	33.867	0.000	54.546	0.028
Escourt	44	0.699	2.522	0.380	6.361	0.000	14.127	0.086
Eshowe	97	0.179	0.603	0.061	0.363	0.000	5.314	0.000
Glencoe	7	0.148	0.207	0.078	0.043	0.000	0.559	0.090
Hlabisa	52	0.175	0.340	0.047	0.116	0.000	2.100	0.083
Impendle	28	0.566	1.604	0.303	2.574	0.000	7.143	0.000
Inanda	145	0.610	1.766	0.147	3.120	0.000	14.400	0.024
Ingwavuma	85	0.394	2.004	0.217	4.018	0.000	17.122	0.000
Ixopo	61	0.194	0.365	0.047	0.133	0.000	1.640	0.000
Klipriver	32	0.111	0.214	0.038	0.046	0.000	1.136	0.042
Kranskop	28	0.664	2.443	0.462	5.966	0.000	12.935	0.007
Lions River	5	0.574	0.711	0.318	0.505	0.000	1.469	0.188
Lower Tugela	26	0.674	1.619	0.317	2.621	0.000	7.607	0.037
Lower Umfolozi	86	0.299	1.173	0.126	1.376	0.000	9.203	0.000
Mahlabathini	78	0.103	0.181	0.020	0.033	0.000	0.880	0.000
Mapumulo	76	0.196	0.429	0.049	0.184	0.000	2.387	0.000
Mooi River	4	1.166	1.439	0.720	2.072	0.018	3.099	0.775
Mount Currie	9	0.171	0.310	0.103	0.096	0.000	0.927	0.000
Msinga	61	0.133	0.198	0.025	0.039	0.000	0.706	0.043
Mthonjaneni	50	0.243	0.728	0.103	0.530	0.000	4.024	0.000
Mtunzini	71	0.130	0.213	0.025	0.046	0.000	1.017	0.000
Ndwendwe	51	0.185	0.332	0.047	0.111	0.000	1.704	0.062

Table 3.3: cont'd

MD	N	MPCIRN	STD	STDERR	VAR	MIN	MAX	MEDIAN
Newcastle	6	0.279	0.248	0.101	0.062	0.044	0.705	0.193
New Hanover	18	0.326	0.907	0.214	0.823	0.000	3.855	0.027
Ngotshe	4	0.153	0.187	0.093	0.035	0.000	0.420	0.095
Nkandla	114	0.154	0.342	0.032	0.117	0.000	1.798	0.000
Nongoma	126	0.218	0.518	0.046	0.268	0.000	3.968	0.000
Nqutu	94	0.120	0.222	0.023	0.049	0.000	1.579	0.000
Paulpietersburg	38	0.082	0.182	0.029	0.033	0.000	0.761	0.000
Pietermaritzburg	100	0.650	2.046	0.205	4.187	0.000	15.416	0.096
Pinetown	77	0.299	0.510	0.058	0.260	0.000	3.233	0.136
Polela	74	0.197	0.359	0.042	0.129	0.000	1.942	0.000
Port Shepstone	81	0.656	3.448	0.383	11.891	0.000	30.435	0.056
Qutu	1	0.000	.	.	.	0.000	0.000	0.000
Richmond	18	0.645	1.413	0.333	1.997	0.000	5.951	0.155
Simdlangentsha	24	0.259	0.672	0.137	0.452	0.000	3.274	0.089
Ubombo	64	0.224	0.585	0.073	0.343	0.000	3.644	0.000
Umbumbulu	48	0.262	0.622	0.090	0.387	0.000	3.580	0.106
Umlazi	26	0.942	0.922	0.181	0.850	0.000	3.720	0.730
Umvoti	24	0.300	0.452	0.092	0.204	0.000	1.376	0.121
Umzinto	97	0.456	2.935	0.298	8.613	0.000	28.800	0.086
Underberg	10	0.729	2.009	0.635	4.034	0.000	6.423	0.000
Utrecht	3	1.382	2.205	1.273	4.863	0.101	3.929	0.117
Vryheid	8	0.598	1.171	0.414	1.370	0.000	3.372	0.102
Weenen	9	0.095	0.155	0.052	0.024	0.000	0.459	0.000

Abbreviations:

MD = Magisterial District   N = Total place-names in the MD   MIN = Minimum
MPCIR = Mean % CIR   STD = Standard deviation   MEDIAN = Median
STDERR = Standard error   VAR = Variance   MAX = Maximum

3.3.5.2 *Ranking of areas with a high risk for cholera (E2)*

The purpose of this exercise was to determine why some areas were more affected than others. In light of the fact that no positive correlations could be established between any of the climatic variables and the incidence of cholera (Chapter 6: 6.1), it was decided to pick 30 of the most

affected place-names for a comprehensive examination into their climatic variables. Therefore, 15 of the most affected places during the major peak and 15 of the most affected places during the minor peak of the epidemic were selected. The main criterion of selecting the 30 most affected places was that they showed an exponential increment of cases from November to January of each peak period in question. Thereafter, GIS ArcView 3.3 was used to calculate the monthly averages of the climatic variables i.e. monthly averages of humidity, rainfall, minimum, maximum and average temperatures of November to January of the two peaks. For the purpose of this exercise, the average monthly climatic variables, refers to the monthly climatic variables averages of the 15 chosen place-names of the respective epidemic peaks. The results generated from this evaluation were used to develop a spatial model for the creation of a vulnerability map using ArcView GIS 3.3. Thus the selected climatic and socio-economic data of the 30 selected place-names were compared to the rest of KZN during the same time period of November, December and January (NDJ) over the 4 years of the epidemic period. Patterns to compare the average climatic and socio-economic profiles of the selected place-names to the averages of the rest of the place-names in the Attribute Database can be distinguished from the graphs portrayed in Chapter 5: Figures 5.22-5.25). At the same time, a spatial model was built using spatial queries in the GIS system to depict place-names with similar environments, natural and human made, which were most susceptible to cholera outbreaks. Due to the significant humidity variations between the two datasets of the major peak and the minor peak, 2 models were proposed, depending on the height above sea level, i.e. one for the lowlands and one for the highlands.

3.4 Spatial Modelling

Spatial models are site selection or suitability models that attempt to find optimum locations, in this case, for outbreaks of cholera. The purpose is to locate the most probable place that is vulnerable to cholera. As the two models are explained below, its worth keeping in mind that the start of the cholera epidemic was in August 2000. It reached peak proportions during the spring and summer months of November and December 2000 and January and February 2001. The

second peak in the following year of 2002 recorded a similar trend during the same afore-mentioned months.

3.4.1   Lowlands Model

According to NDJ 2000/01 major peak, a place-name has to have the following characteristics for it to fit into the lowland model. This means that if a place-name situated in the lowland areas satisfies the following criteria, the possibility of a cholera outbreak is high. Thus:

- The humidity should be $\geq 70.7\%$.
- The height above sea level should be below 900m.
- The minimum temperature should be $\geq 20.8\ {}^{0}\mathrm{C}$.
- The maximum temperature should be $\geq 28\ {}^{0}\mathrm{C}$.
- The monthly rainfall should be ≥ 222 mm.
- The population density should be at least 170 persons/sq km.

Additional criteria based on the statistical correlations supporting the probability that socio-economic factors also have an influence, were also included (refer to Chapter 5: 5.3.1). A flow chart portraying all the characteristics required to be fulfil the pre-requisites for inclusion in the lowlands model is given in Chapter 5: Figure 5.26.

3.4.2   Highlands Model

According to NDJ 2001/02 minor peak, a place has to have the following characteristics for it to fit into the highland model. This means that if a place-name situated in a highland area satisfies the following criteria, the possibility of a cholera outbreak is high. Thus:

- The humidity should be $\geq 58.7\%$
- The height above sea level should be > 900m.
- The minimum temperature should be $> 15\ {}^{0}\mathrm{C}$.
- The maximum temperature should be $> 26.3\ {}^{0}\mathrm{C}$.
- The monthly rainfall should be 130.4 mm.
- The population density should be at least 324 persons/sq km.

In addition, as socio-economic factors (all related to water and sanitation) could also have an influence, selected criteria from these data types were also included. A flow chart portraying all the characteristics required to be fulfil the pre-requisites for inclusion in the highlands model is given in Chapter 5: Figure 5.27.

The results presented in the following Chapter 4 are part of the outcome of the methodology described in this chapter. In particular, chapter 4 gives an overview of the study area of KZN, highlighting the province of KZN, its geographical characteristics; demography, socio-economic status and the history of cholera going back two decades prior to the epidemic in question. The description of the study area brings forth awareness of the existence of cholera in the province of KwaZulu-Natal over the last two decades. This, together with the crisis brought about by the epidemic that started in August 2000, rationalised the interest in undertaking the study.