HANDLING HEALTH DATA IN A GIS ENVIRONMENT:
Geo-referencing and Analysis

The case study of Dar es Salaam, Tanzania

Joseph Stephen Mayunga
February 2002

HANDLING HEALTH DATA IN A GIS ENVIRONMENT:
Geo-referencing and Analysis

The case study of Dar es Salaam, Tanzania
by

Joseph Stephen Mayunga

Thesis submitted to the International Institute for Geo-Information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Urban Planning and Land Administration

Degree Assessment Board

Chairman: Prof. Dr. F.I. Masser
External examiner: Prof. H. Ottens
First supervisor: Drs. S. Amer
Second Supervisor: R.V. Sliuzas, MSc

ITC

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-Information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Dedicated
to

my mother Rozaria and my father Mzee Stephen

Abstract

This research thesis proposes a methodology for geo-referencing health related data collected at a household level. It deals specifically with surveillance data collected by the Adult Morbidity and Mortality Project (AMMP) in Dar es Salaam city in Tanzania. The research explores the capabilities of Geographical Information System (GIS) and remote sensing in developing the methodology.

In Tanzania currently no geo-referencing method exists that could be used accurately to geo-reference health data collected at household level. The research examines various international examples of geo-referencing methods and builds a detailed evaluation of factors to be considered in developing the methodology in the context of the Tanzanian environment. Factors such as lack of up-to-date data especially in informal settlements and the use of Small Format Aerial Photography (SFAP) mosaic as a tool for data updating are discussed. The proposed methodology is tested and evaluated using two demographic surveillance sites. The strengths, weaknesses and recommendations on how to improve the methodology are discussed.

After geo-referencing the data, an attempt was made to analyse the data at ten-cell unit level. The analysis specifically examines mortality conditions of the two surveillance sites. It further tries to establish if mortality conditions in the two sites are related to unemployment and overcrowding conditions, as health determinant factors.

Acknowledgement

Table of contents

List of figures

List of Maps

List of Tables

ACRONYMS

AMMP	Adult Morbidity and Mortality Project.
CDO	Community Development Officer
DEM	Digital Elevation Model
DFID	United Kingdom Department For International Development
DSS	Demographic Surveillance System
ED	Enumeration District
GIS	Geographic Information Systems
ID	Identity Number
MOH	Ministry of Health
NBS	National Bureau of Statistics
NGO	Non Governmental Organization
NHC	National Housing Corporation
SFA	Small Format Aerial Photograph
TIGER	Topologically Integrated Geographic Encoding and Referencing
UK	United Kingdom
US	United States of America
VA	Verbal Autopsy
ZIP	Zone Improvement Program

Glossary

Balozi	Ten-cell leader
Kijiji	Village
Kitogoji	Hamlet
Mtaa	Sub ward
Shina	Ten-cell unit
Tawi	Branch

1. Introduction and general approach

1.1.   Background

The term geo-referencing in its broader meaning refers to defining or assigning a geographical location on space to an object or feature. Positions of geographic features can be defined in space by a set of coordinates. de By (2000) states that data is geo-referenced when coordinates from geographic space have been associated with it. That is, geo-reference or spatial reference, tell us where the object represented by the data is or was or will be. In the light of this statement, geo-referencing could be defined as that process of associating data with its geographical coordinates.

Objects are located on the global by using a plane coordinates system, which is defined as a set of rules, which specifies how the coordinates are assigned to points (de By, 2000). For arbitrary location on the earth surface, the Cartesian coordinate system and the geographical system of longitude and latitude of the earth are examples of coordinates system based upon Euclidean geometry (Figure1.1). A map projection, a spheroid of reference and a datum usually define a coordinate system. Geographical coordinates (latitudes and longitudes) are the universal system for defining spatial position. A set of geographic coordinates on a datum is complete and unique, worldwide.
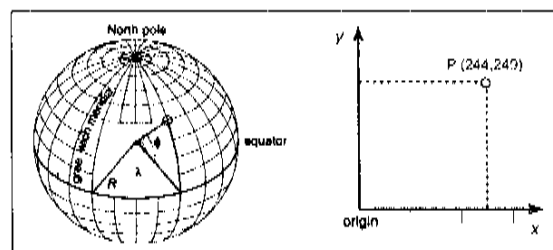
Figure 1:1:Geographical coordinate and Cartesian coordinate systems

Source: de By, 2000

To represent parts of the surfaces of the earth on a flat piece of paper or on a computer screen, normally it involves a mathematical projection method that transforms the location of features on the earth's surface to location on two-dimensional surface. Because the earth is three-dimensional some methods must be used to depict the map in two-dimension. There are a variety of map projections, but all are generally of three basic types; Azimuthal, Conical and Cylindrical projections. For instance a Transverse Mercator is a variant of cylindrical projection (de By, 2000).

On the other hand, for small scale, a position of geographical feature can also be defined in space by the use of indirect methods such as, its place name, address or property. Selection of which method to use depends to a large extent on the purposes and use of the data to be geo-referenced.

Though names could be easy to use than longitude and latitude, yet they are not convenient for many purposes. Their main disadvantage is that the chances of duplicate names to occur are high. Therefore

for small areas, the alternative method of spatial or location reference, which is convenient for many purposes, is the use of an address of the object or feature. Cowen (1997) defines such type of geo-referencing or geocoding as the process of assigning geographical location (e.g. latitude and longitude) to a geographical feature on basis of its address.

When we talk of geo-referencing health related data, we directly refer to geo-referencing information about people. It is quite clear that information about presence and characteristics of people could only be conventionally geo-referenced indirectly. This is because; it is difficult to define a location of an identifiable individual. Individuals are always mobile and cannot be tied to any specific location. Thus geo-referencing health related data could only be done through individual's place of residence, which depends on some form of geographical code, which can be applied to an address or property (Martin and Higgs, 1995).

Most of these methods are said to be indirect methods in the sense that such methods provide just a key or index, which can then be used with a table to determine the real world coordinates. Cowen (1997) states that a ZIP code is a good example of such geo-referencing methods, which locates an object or a feature based on its address. Instead of providing coordinates for a place directly, it gives a unique number, which has to be checked in a base map and again associate with it. The term geo-referencing in this study context therefore will explicitly refer to that process of assigning health events to a geographic location on basis of its address.

1.2.   General overview

There is a great need for improving the health status of the people in Tanzania. The big challenge re-mains to be how health delivery systems could be improved and ensure equitable allocation of limited health resources to all people in the country.

In recent past the government has recognized that, health planning process requires research in order to obtain sufficient and reliable information. Despite its main sources of information such as hospitals, registrar of births and deaths and the National Bureau of Statistics (NBS), efforts have been made, where by a number of international, private and government organizations, institutions, NGOs are now conducting a wide range of research on various health issues. However most of the research are con-ducted either at national, regional or district level, which is quite evident that these levels are too gen-eral to represent reality of the target population group (the poor) at a grass root level. This is because aggregation of data at this level is normally difficult to show clearly the inequalities between the poor and the rich. On the other hand, those projects, which attempt to go down to the grass root, are not able to analyse their data spatially, and therefore remain tied to traditional tabular analysis. As a result the potential of spatial aspect, which is an important element in health planning, remains untapped. Indeed, the data that is collected by different organizations cannot be linked and its use to other appli-cations also becomes limited.

Spatial analysis and linking of such vital information have not been possible because the data is not geo-referenced at a smaller scale. Geo-referencing of health data at a very fine resolution such as a

household in a country like Tanzania where address system does not exist, poses a critical challenge. The focus of this research therefore is to find out the possibility of using GIS to geo-reference such data at household level and further examine how and to what extent geo-referenced data at this level could be useful in health planning and policy making.

The following sections of this chapter, describe the research design in which significance of the research, research problem and objectives will be defined. Finally, the research methodology is illustrated.

1.3.   Significance of the research and problem definition

The Adult Morbidity and Mortality Project (AMMP) is one of the main reliable sources of health information for the Ministry of Health in the country. It is the project, which collects data at a household level. Currently, the AMMP data provides a wealth of information on health in the country. Hitherto, the data could only be analysed statistically, spatial analysis could not be performed, mainly because the data is not geo-referenced at household level. As a result, data in this form may not provide possibilities for health officials to adequately identify, analyse, and monitor spatial health problems at small-scale level.

Incorporation of spatial component in AMMP data acquisition, data processing and management may help to investigate the morbidity and mortality causes spatially. This is because geo-referencing allows health data to be examined in relation to its spatial distribution. Therefore, increases understanding of spatially related diseases and deaths in certain locations. This will possibly reveal new insights and new strategies for improving health planning and policy making in the country.

1.4.   Research objectives

The main objective of this research is to **develop a methodology for geo-referencing** of the AMMP health related data. The specific objectives of the study are:
1. To assess various methods which could be used for geo-referencing the AMMP health data.
2. To examine the usefulness of Small Format Aerial Photographs (SFAP) in geo-referencing the AMMP health data.
3. To evaluate the usefulness of geo-referenced data at a small scale for AMMP project and health planning.

1.5.   Research methodology

As stated in section 1.4, the main objective of this study is to develop a methodology for geo-referencing the AMMP data. It is believed that geo-referencing of such data, at such a fine resolution needs understanding of the method to be used, a clear knowledge of the type and/or characteristics of the data to be geo-referenced. In order to achieve this objective an empirical methodological approach was developed. This methodology consists of five key steps, literature review, method selection and development, field survey, data entry, evaluation of the method, and data analysis and exploration. Literature review, examines various existing geo-referencing methods. The main concern however is the way in which geo-referencing is carried out, its limitations and what information geo-referenced data at small-scale level could provide.

After a detailed evaluation of various geo-referencing methods, which are commonly used in different countries, the next step was to select a method, which could best be applied for geo-referencing the AMMP data in the context of Tanzanian situation. Selection of such a method could not be an easy task, as it entails understanding of various aspects with regard to the AMMP data collection methods, data requirements and general current situation of digital data availability in Dar es Salaam city. In order to be able to develop a method, which is feasible (the word feasible here refers to the method which is easy and simple to use and yet accurate enough), some criteria were defined based on these aspects.

The methodology was developed based on the two AMMP data collection methods, the census survey and the verbal autopsy method. The use of small format aerial photography was chosen as a method for up-dating the data. After selection and development of the geo-referencing method, the method was tested in two AMMP demographic surveillance sites, using two approaches; the use of enumerators and the use of ten cell leaders. Following the completion of the field survey was the data entry task and evaluation of the method. The result of the field survey and data entry was the geo-referenced data. Having geo-referenced the data, then the data was analysed and visualised. However the analysis focuses on exploration of the mortality condition, the relationship between mortality and unemployment and overcrowding conditions as health determinant factors in the two sites. Figure 1.2 shows a schematic diagram of the methodological approach that was followed in this research.

1.6. Dar es salaam city as the case study area

1.6.1. General overview

Dar es Salaam city is the administrative, industrial, commercial, and governmental centre of Tanzania. Its population is approximately 3 million, with an estimated growth rate of 8%, one of the highest in sub-Saharan Africa. Rapid growth has provided the urban population with many economic opportunities, but has also led to the rapid deterioration of environmental conditions. This has limited national and city economic development, and has adversely affected the health and welfare of the city's residents (Rory et al., 2000). It is estimated that 70% of its population live in unplanned areas or informal settlements (Sliuzas et al., 1999). Poor sanitation, poor drainage system, poor solid waste management and marginal access to piped water characterize these unplanned settlements. All these problems are a threat to human health.
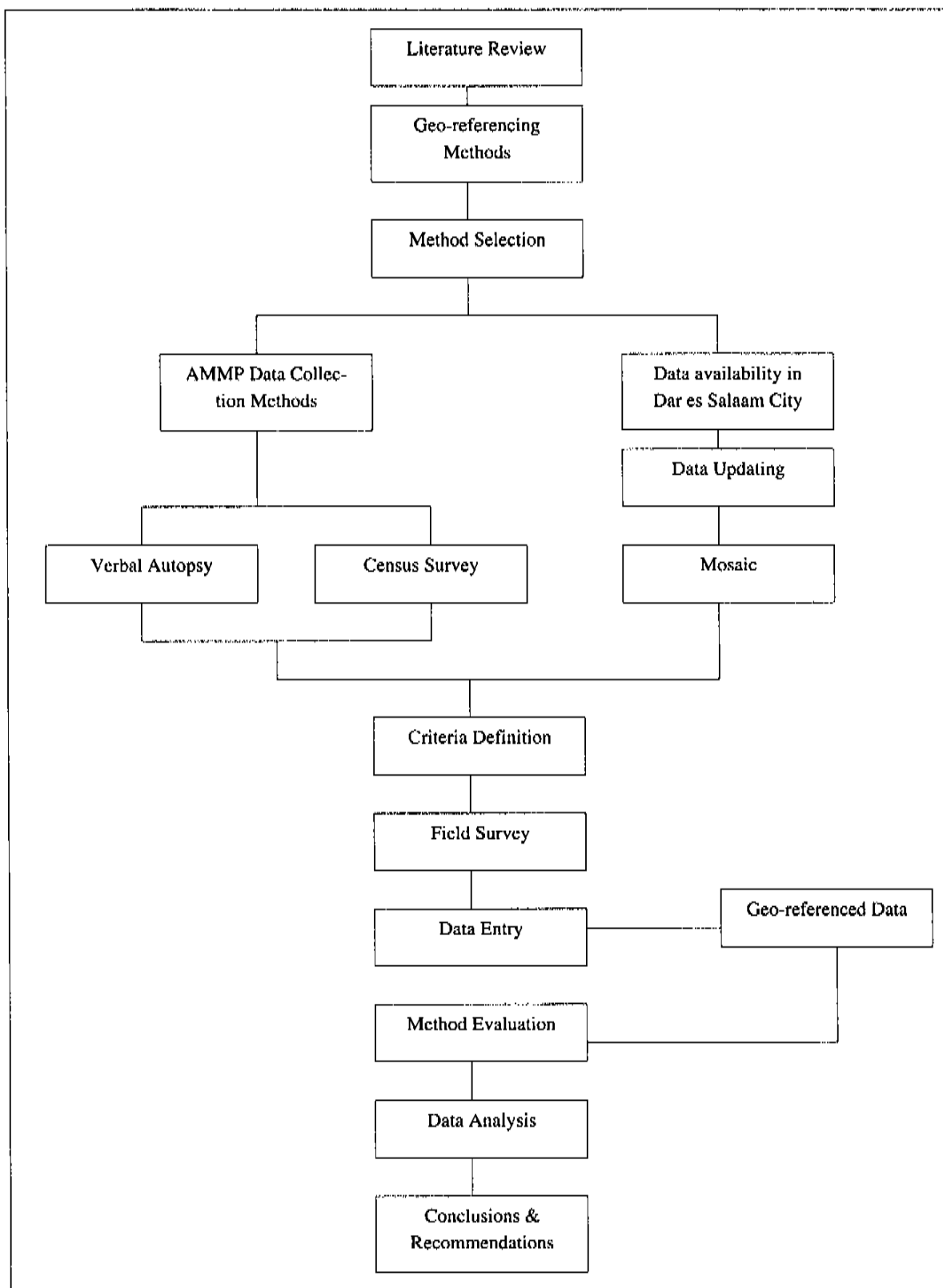
Figure 1:2 Research Methodology

1.6.2.   AMMP project and the study area

The bulk of the data used in this study mainly stemmed out of the AMMP project. This project is based on three Demographic Surveillance System (DSS) sites in Tanzania mainland. Two sites are located in rural area, these include, Morogoro rural district in Morogoro region and Hai district in Kilimanjaro region. The urban surveillance area is in Dar es Salaam city (Map 1.1). The AMMP demographic surveillance sites consist of approximately 73,000 households, with a total population of about 308,000. Hai district area, which consists of 51 villages, has approximately 32,000 households. Morogoro rural district consists of 61 villages and has approximately 25,000 households. Dar es sa-laam surveillance area covers three branches, with approximately 16,000 households (AMMP, 1997). This research is focusing only on the AMMP surveillance areas in Dar es Salaam city. In chapter two more details on Dar es Salaam surveillance area will be given.

According to the AMMP project, Demographic Surveillance System (DSS) is defined as a geographi-cally defined population under continuous demographic monitoring with timely production of data on all births, deaths, and migrations. The AMMP demographic surveillance system has three main com-ponents:

- Enumeration of denominator population by repeated household visits at regular intervals
- Continuous reporting of critical events, especially deaths, by key informants
- Cause of death determined by Verbal Autopsy (VA)

AMMP is a project of the Tanzanian Ministry of Health (MOH). It is funded by United Kingdom De-partment For International Development (DFID) and implemented with the technical assistance and management of the University of Newcastle upon Tyne (UK). The first phase of AMMP Project started in 1992 and ended in 1999. The second phase commenced in 2000 and is designed to run until 2004.

The origin of the AMMP project goes back as early as 1980s,when a number of physicians showed an interest to study the prevalence of diabetes in Tanzania. From diabetes the interest extended to the as-pects of other non-communicable diseases such as hypertension stroke, cardiovascular diseases and their associated risk factors. Today the interest of AMMP project has expanded to an investigation of major causes of morbidity and mortality for adults in the country (AMMP, 1997).

One of the main objectives of the project is to improve the sources of data on Adult Morbidity and Mortality causes. This is due to the fact that, the ministry of health of the Tanzania government, lacks sufficient and reliable data for essential health planning and policymaking. Due to this weakness, the Adult Morbidity and Mortality Project realized the importance of collecting such information. The second phase of the AMMP project is specifically concerned with encouraging the equitable allocation of health resources to reduce diseases and conditions most likely to cause suffering and disadvantage to poor people (AMMP, 1997).

1.7.   Structure of the report

This report consists of five chapters; chapter one outlines the introduction part and the general approach of the study, which covers the problem statement and significance of the research, main objective of the research, research methodology and description of the case study area.

Chapter two covers the literature review which is focusing on reviewing various geo-referencing methods, and it further evaluates the methods in the context of the Tanzanian situation.

Chapter three describes the methodology used to geo-reference the data. It also outlines problems and limitations of the methodology.

Chapter four describes how the analysis and exploration of geo-referenced data is carried out, and summarizes some of the results.

Chapter five draws some conclusions and recommendations based on the proposed methodology and usefulness of the result obtained from the analysis of the data.

Map 1:1 AMMP surveillance sites in Tanzania
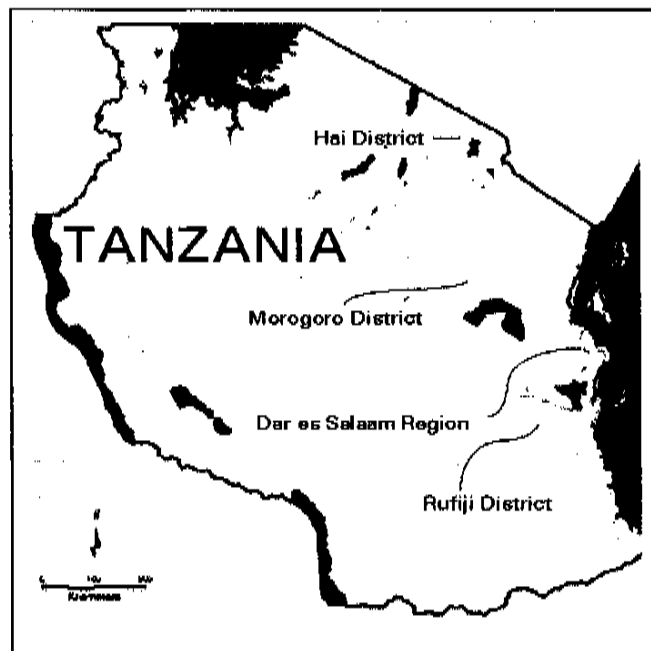
Source: AMMP, 1997

2. Geo-referencing methods

2.1. General overview

This chapter gives a general overview of the application of GIS technology in the science of health. It discusses various geo-referencing methods from different countries. It further describes a set of criteria, which is used to develop what is termed as "feasible" method for geo-referencing the AMMP health data.

Mapping health data is not a new concept. There is a long and rich tradition of mapping health related phenomena stretching back at least as far as the 19th century. One of the first uses of the spatial analysis of epidemiological data was during the 1854 outbreak of cholera in London by Dr. John Snow. Working on his hypothesis that cholera was transmitted by unclean water, he was able to trace the source of the cholera outbreak to the Broad Street water pump in the area of Soho in London (Savigny and Wijeyaratne, 1994). Since then, the use of cartographic methods and spatial statistical analysis of epidemiological data has become a powerful tool used by health professionals to investigate causes of disease as well as its site of origin. Today, the techniques used by various professionals to map and analyse health related phenomena have become much more advanced than the techniques used during the era of Dr. Snow. Geographical Information Systems (GIS) and remote sensing techniques are now widely used to identify hot spots of regular features of diseases and deaths.

Mapping can be used as an exploratory tool for generating hypothesis such as establishing correlations with environmental factors. Diseases mapping may reveal the existence of spatial disparities and lead to the delimitation of risk area, it can be applied in health care planning and public health policy formulation (Verhassel, 1993).

Gatrell et al (1998) argue that experience with applications of geographic information systems (GIS) to health issues has increased considerably over the last decade. GIS and remote sensing techniques are now frequently used for health planning and policy formulation in various places. Despite the considerable use of GIS to address health questions, there had been limitations in use of GIS to facilitate health planning and policymaking. The limitation of application of GIS in this manner is due to the problems encountered in identifying, acquiring and integrating a wide range of geo-referenced data relevant to population health.

2.2. Review on existing geo-referencing methods

2.2.1. Introduction

There are several indirect geo-referencing methods that are commonly used in different countries worldwide. This section will make an attempt to discuss some few selected methods. The methods will be evaluated in relation to the Tanzanian situation and some remarks on each method will be made.

2.2.2. Small administrative units method

This method refers to the use of small geographical administrative units in geo-referencing the data. However the units differ in size from country to country. The most common small administrative units, which are used for geo-referencing are those based on population census surveys. Census method is one of the major sources of information on the number, socio economic characteristics and location of population in most countries in the world. Census survey systems again differ from country to country in various ways in which they are organised, managed and even the way they are conducted. In developed countries the census systems are far more advanced than that of the developing countries. In developed countries the census systems are now developed using GIS technology. In the context of this technological advancement, management and handling of census data becomes easy, indeed growth of use of census information has increased considerably.

In many countries, data is collected and reported using a system of several different types of reporting units; this could be a political or administrative unit such as province, county, municipal, city or ward. The data also can be reported based on units defined to easy data collection such as blocks or Enumeration Districts (ED). The data sometimes is reported using units defined to easy analysis of the data such as census tracts. In the UK, for instance, the most common unit used in census survey systems is the Enumeration District (ED), of which digital data files of boundaries and centroids can be obtained from the Census Bureau (Raper *et al.*, 1992) and could also be used for geo-referencing purposes. EDs can be linked with other data such as health data, and it becomes possible to know which health event is located or associated with a particular ED. However the most critical point here is the size of the area of the ED in which the census data is collected and the level at which the health data needs to be geo-referenced. In the UK, the 1991 EDs contained 400 individuals in 200 households. Enumeration districts can also be linked with e.g. postcodes, which are much smaller spatial units than EDs (Martin and Higgs, 1995).

Looking on Dar es Salaam administrative boundary and the population census system in the country in general, the use of small administrative unit as a geo-referencing method is not possible. The complexity is attributed to the fact that enumeration districts in the country are considered to be ad hoc and varied only for the purpose of enumeration in order to ensure consistence during census survey. Their sizes also vary considerably. Additionally, enumeration district boundary data are currently not available in digital form. As for spatial unit upon which this study is focussing (the household level), the use of small administrative units such as enumeration districts are not ideal.

2.2.3. Grid squares/hexagonal method

The use of grid square/hexagonal seems to be a useful geo-referencing method especially where small administrative units are not effective. The grid squares/hexagonal could be designed and created and then overlaid on the base map with known coordinates, which could then be used as spatial unit. Mohan and Rhind (1983) in their work (mapping mortality/morbidity data) revealed that the use of grid squares helps to avoid the difficult task of linking the data manually to a certain spatial unit such as enumeration districts especially when the digital boundaries of EDs do not exist.

Amer and Thorborg (1996) used hexagonal grids to capture spatial interaction behaviour of patients' mobility in Dar es Salaam city. In their work they used hexagonal grid, in which one cell unit repre-

sented an area of 250 x 250m on the ground. They argue that this level of spatial resolution enables rapid and accurate spatial data encoding. However a grid of 250 x 250 m may contain dozens of households. Although the use of grid squares/hexagonal seems to be a useful method where small administrative units are not effective, the method is seen to be much more useful for geo-referencing data at a large spatial unit than a household. Mohan and Rhind (1983) criticise the method by arguing that the use of grid squares/hexagonal breaks up the structure of the streets within which interactions and communication take place. In some cases grid squares/hexagonal decant the populations of different generic, socio economic and demographic characteristics into the same unit.

The criticism raised here however does not reflect the usefulness of the method in geo-referencing process but rather it is based on the implication of the final result after the analysis process.

2.2.4.   Postcode method

A postcode system is primarily meant for mail sorting. However it contains a number of attributes, which could be used for other applications. It is geographically referenced to quite a high degree of resolution and can be imported into various GIS software (Cowen, 1997). Postcode systems as a geo-referencing method are currently used in many countries of the world particularly in developed countries, however their forms and structure also differ from country to country.

Postcode method matches an address to a base map file on the basis of its postal zone information. The actual street address is not utilized, only the postal codes. It can be used in countries with postal code system. For better illustration of the postcode method, below an example of the US postcode system is discussed.

The US postcode system is probably the earliest system of all. It is known as ZIP code; Zone Improvement Program. The first digit of ZIP code divides the US into national areas, 10 large groups of states numbered from 0 in the North East to 9 in the west. Within these large areas, each state is divided into an average of 10 geographical areas, which are identified by the 2nd and 3rd digits of the ZIP code. The 4th and 5th digits identify a post office or a local delivery area (Raper et al., 1992). Figure 2.1 illustrates a simplified example of the US post code system. In the US, ZIP code files were developed in the late 1960s by the Bureau of the Census, and are now being replaced by more comprehensive TIGER files; Topologically Integrated Geographic Encoding and Referencing (Cowen, 1997).

The TIGER/Line files are a digital database of geographical features, such as roads, railroads, rivers, lakes, political boundaries and census statistical boundaries, covering the entire United States. The data base contains information about these features such as their location in latitude and longitude, the name, the type of feature, address ranges for most streets, the geographic relationship to other features, and other related information.

The TIGER/Line data can be imported into many Geographic Information System (GIS) software. With the appropriate software a user can produce maps ranging in detail from a neighbourhood street map to a map of the United States (Cowen, 1997).
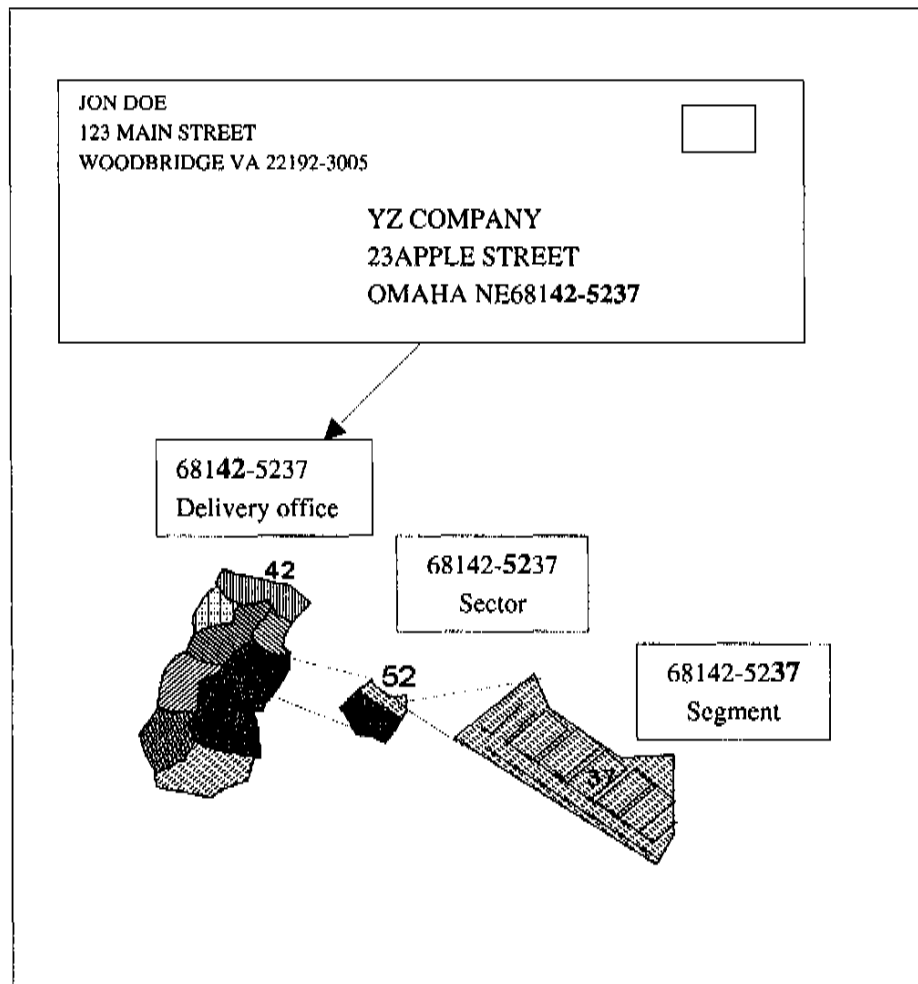
Figure 2:1 Postcode system in the US

Source: Modified after Raper et al., 1992

Despite the great potential in geo-referencing, which could be harnessed from the postcode methods, there are limitations in using these methods for geo-referencing purposes. In the first place, these methods have not been designed for geo-referencing purposes; they are in principle designed for internal use to the postal system. It is therefore difficult to ensure stability with time because post office does not have a mandate to maintain these systems for geo-referencing purposes, therefore postcode zones may change without notice. Cowen (1997) claims that, the postal zones sometimes have overlapping and fragmented boundaries.

With regard to the Dar es Salaam city context, although postcode method seems to be one of the reliable methods for geo-referencing data at smaller scale, applicability of this method in Dar es Salaam city is extremely difficult if not impossible. This is due to the fact that, from the planning point of view, Dar es Salaam city, as it has been indicated earlier, has more than 70 percent of its population living in unplanned settlements or informal settlements. Informal settlements more often contain neither street address nor postcodes. Indeed, what has been defined as postcode system in the US does not exist in Dar es Salaam city or in Tanzania in general.

2.2.5.   Street address matching method

Street address matching is the method used to match address to streets, it is commonly known as Street address matching. Databases to support address-matching method exist in most industrialized countries. This method differs from the postcode in the way that it deals with an individual address file in which each address has a unique identifier. It provides possibilities of geo-referencing the data at smaller scale. It is done by first identifying the block, which contains the address. Then the position of houses is estimated by using the coordinates of the end point of the block. For instance if 551 B Street lies in the block running from 501 to 599, then the exact position of the house is estimated using linear interpolation. That is, 551 is roughly half way down the block (Cowen, 1997). However the main disadvantage of such estimation is that in some countries addresses are not sequential along the street and indeed such estimates are crude.

Address matching method, has some problems that might make it difficult to obtain a one hundred percent match rate. Such problems may include, for instance, spelling and punctuation variations. This makes the matching to be more difficult e.g., Ave. or Avenue. Cowen (1997) asserts that a failure rate of 10% is regarded as good. The precision of addresses matching as geo-referencing method varies. It tends to be higher for houses in cities and it is lowest for rural area where the address may indicate only that the place is somewhere in the area served by the post office without street names.

2.2.6.   Parcel Based method

Parcel based system as a geo-referencing method is commonly used in the US. It is the basis for land surveys and legal land description. Unlike the previous methods, it is designed to reference land parcels. Since it is a comprehensive, systematic approach, it has been possible to use it as a geo-reference method. There are software packages that exist, which can be used to convert Parcel Based System descriptions to latitude/longitude.

More often accurate parcel level files are created from legal description of the property on deeds using GIS software. A good example of Parcel based system is that of Columbia in the US, where by a parcel centroids file that include streets, addresses, and nine digit ZIP codes has been created (Cowen, 1997). That means other attribute data such as health and social economic could accurately be linked to the file and used in GIS database.

Looking on the cadastre system of Tanzania, Dar es Salaam in particular, if more than 70% of the urban residents are accommodated in informal settlements, definitely the cadastre system is still very weak. Indeed this might be one of the main reasons for the higher rate of informal settlements in the city. As it has been stressed that the parcel based system operates well where cadastre system is strong, effective and well structured, with such a poor cadastre system in Dar es Salaam city application of parcel based system as geo-referencing method especially in informal settlements is not feasible at this moment.

2.2.7.   Ground survey method

This method is used to determine the coordinates of an object or feature by physically visiting the site. The location of an object or feature could be established by one, calculating the location of a property through conventional surveying methods, two by the use of Global Positioning Systems (GPS) and three by using address from the digital version of boundary of properties. It is quite an effective

method, however GPS device could be extremely useful if the data has to be geo-referenced at a much higher spatial unit such as a village, ward, district or region. In such a situation where the use of GPS is not possible therefore it requires other strategies of exploring the method to produce another meaningful method to geo-reference the data.

2.3. AMMP data collection methods and digital data availability.

2.3.1. Introduction

After an evaluation of various geo-referencing methods, the next step was to look on the issue of lack of up-to-date data, the AMMP data collection method and data requirements. These three factors are basis for criteria used to develop and implement the method. In this section these factors will be discussed.

2.3.2. AMMP data collection methods

The first factor to be considered was how the AMMP project is collecting its data. Before starting to discuss the data collection methods, an introduction on AMMP surveillance areas in Dar es Salaam is outlined.

The AMMP project in Dar es Salaam city covers three areas, Ilala and Keko, which are located in Ilala municipality, and Mtoni area, which is located in Temeke municipality. Each area is divided into smaller areas called branches, and these three areas together consist of 8 branches. Ilala has been divided into two branches called Ilala Ilala and Shauri Moyo. Keko is divided into three branches, called Keko Keko, Keko Toroli and Chang'ombe. Mtoni area is divided into three braches, which are Mtoni Mtoni, Mtoni Relini and Mtoni Saba Saba (Map 2.1).

Although AMMP surveillance sites in Dar es Salaam were chosen to be the case study area for this study, because of time constraint and resource limitations it was not possible to work on all the eight braches of the AMMP project in Dar es Salaam city. In this case, only two branches were selected; these are Ilala Ilala branch and Mtoni Relini branch. These two branches were chosen mainly because of their clear distinctive social and economic characteristics. Mtoni Relini is the unplanned area characterized by poor housing condition, while Ilala Ilala branch is an old planned area, which contains houses built by the National Housing Cooperation (NHC) in 1960s and comprises both middle and high-income level groups. These two branches could also represent an acceptable wide range between being small enough to allow detailed variation to be identified and yet large enough to contain sufficient number of cases of deaths.

Map 2:1 AMMP Dar es Salaam surveillance areas
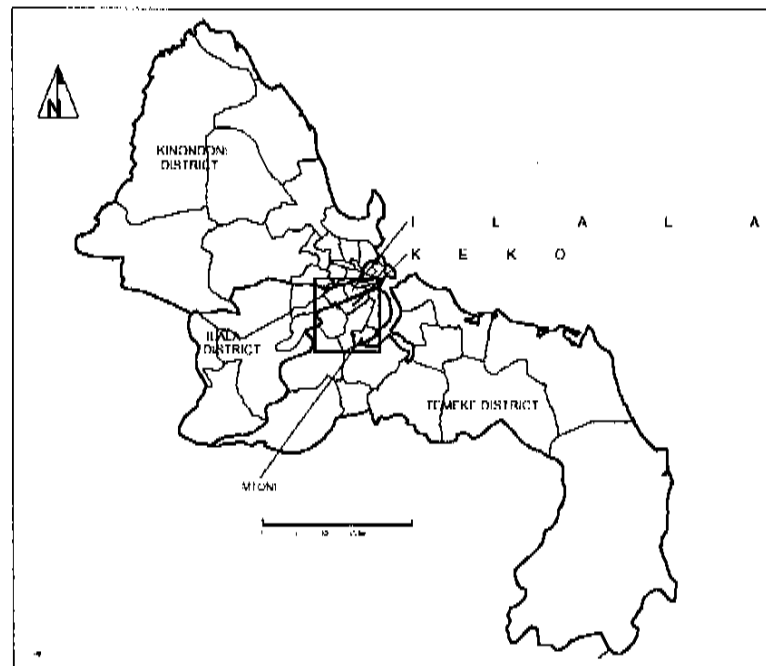
Source: AMMP, 1997

More often data collection, no matter what type of data, is always not an easy task. The complexity of data collection mainly depends on the purpose of the data to be collected, and to a large extent the degree of accuracy required. Collection and handling of health related data seem to be much more demanding and complex than one would think. This is because of its confidentiality and sensitivity. The AMMP project uses two methods in its routine data collection system, which are, census method and verbal autopsy method.

(i) Census method

This method provides the data on the total number of people in the surveillance area and their social and economic status. In Dar es Salaam, the census interview is conducted twice in a year.

When the project started in 1992, a baseline study was conducted to determine all the residents in the selected surveillance areas. Information of each individual household was entered in a single census form. The data was then entered in the computer. Thereafter, subsequent census rounds have been made twice in a year, one in June and the other in October.

Before each census round starts the information on each household is printed out on a new form. In the field, existing data is verified and in some cases updated by the enumerators. New households that migrate into the area are registered in a new form. Through this way information of each household is monitored, controlled, updated and maintained.

It is also clearly stated that this surveillance system does not attempt to track individuals as they migrate into and out of the surveillance areas repeatedly all the time. The census system determines only an individual who is present at that particular date. Thus it is not possible to track an individual who migrate out and then come back into the area. The data variables that are recorded for an individual by the enumerator during the household interview include; tawi, shina, household code, balozi, name,

age, sex, marital status, drinking habit, smoking habit, date of entry in the household, mode of entry, date of exit, mode of exit, orphan status[1] (See appendix 1).

(ii) Verbal Autopsy method

This method provides the data on number and causes of death. Verbal Autopsy is a tool used to determine causes of death in area where vital registration system are unable to produce such information. The questionnaire form is used to collect data during verbal Autopsy. The interview is carried out by trained health personnel, who interview the relative of the deceased or any other person who was taking care of the deceased (AMMP, 1997). Unlike census survey method, detection of death and use of verbal Autopsy is a continuous process throughout the year. The questionnaire form of verbal autopsy contains sections which identify the respondent, a section which identifies the deceased, an open ended history section, a checklist of previous diagnosed conditions, a checklist of health services sought in the period leading to death, a residential history and summary of any confirmation evidence (See appendix 2).

2.3.3.   An overview of Census and Verbal Autopsy methods

Based on the census methods described above, two important aspects were noted which could directly influence the strategies on how the proposed method had to be developed. One, enumerators were identified as key persons who undertake the survey in the field; therefore they need to understand clearly the method. This indeed necessitated understanding of their level of education in developing the method. Two, how the AMMP census method is carried out provided a better understanding of what the proposed method should reflect

It was also found that these two methods, (Verbal Autopsy and Census method) are carried out separately by different people at different periods of time. Linking of the data from the two methods is not possible, as there is no common variable that can be used accurately to link the two data sets. The only option variable, which could be used to link the data, is the names. However using names as it has been clarified by the AMMP is not a reliable method because names do change with time and sometime different people could have exactly the same name, a phenomenon, which is very common in Tanzania. This is one of the main limitations of this study.

There is a need for the AMMP project to find out a way, which may enable the two data sets to be linked together. As linking of the data sets may provide more opportunities, since many more applications may become possible when the data are linked together than when they are handled separately.

2.3.4.   AMMP data requirements

Another factor, which was considered to be very important in developing the method, is the AMMP project data requirements, in terms of accuracy, and spatial unit. The accurate location of cases of death was considered as an important factor towards developing the methodology. The main concern here was the level of accuracy at which a case can be located to its correct geographical location. Geo-referencing process of small area requires knowledge of the administrative set up of the city. In this section, the administrative structure of the city is briefly described.

---

[1]Tawi = Branch, Shina = Ten-cell unit, Balozi = Ten-cell Leader.

Dar es salaam city consists of five administrative levels: City, Municipal, Division, Ward and Sub-ward. The highest level is the city followed by the municipal. The city consists of three municipals, which are, Kinondoni, Ilala and Temeke (Map 2:1). The third administrative level is the division; four divisions are in Kinondoni municipal, three in Temeke, and three in Ilala. Below the division, there are wards and subwards (*Mtaa*). In some areas, especially in the rural areas, there are also villages (*Vijiji*) and hamlets (*Vitongoji*)(Table 2.1).

Table 2:1 Dar es Salaam administrative set up

Municipal	Divisions	Wards	Sub wards	Villages	Hamlets	Population est. 1998/99 (million)
Kinondoni	4	27	113	14	14	1.2
Temeke	3	24	97	15	62	1
Ilala	3	22	65	9	27	0.8
Total	**10**	**73**	**275**	**38**	**103**	**3**

Source: Rory et al., 2000

Table 2.2 provides the summary of spatial units, which indicates the city administrative set up, the AMMP project and that of the research approach. It was learnt that the AMMP project is still using branch unit, which is no longer part of the city administrative set up. Branch is a unit, which was established during a single party system. The reason for this was claimed to be that, the sub ward, which is the new replacement of the branch unit, covers too large a geographical area to suit the project purposes. However it is also worth mentioning that with city council administrative system, the units below the ward have no clear geographical boundaries. For the purpose of this study, block as a spatial unit, which is smaller than a branch, is introduced.

2.3.5.   Digital data availability in Dar es Salaam

The overall goal is to build a GIS database, from a set of geo-referenced health related data. The possibility to achieve this goal depends mainly on the reference or base maps that are available for the case study area. Looking on the situation of digital data availability in Dar es Salaam city, it was found that, the only data available is the base map which dates 1992 at a scale of 1:2500.This base map covers only the main, built up area of the city. Another data source is the Digital Elevation Model (DEM), which also covers most of the built up area of the city.

Table 2:2. Summary of urban Spatial Units

CITY COUNCIL	AMMP approach	RESEARCH approach
City	City	City
Municipal	Municipal	Municipal
Division	Division	Division
Ward	Ward	Ward
Sub ward (Mtaa)	**Branch**	Branch
	Ten-cell unit	**Block**
		Ten-cell unit

The advantage of this base map with such a scale is that, it is detailed enough to enable identification of important small features such as buildings and street networks. However the main draw back of this base map is that, it is out of date, as up until now (2001) many changes might have occurred especially in unplanned areas. Therefore updating of the base map becomes inevitable. The use of small format aerial photograph mosaic was chosen as method to be used for updating the data. This method will be discussed in chapter three.

2.3.6. Criteria for developing a geo-referencing method

Based on the three factors, described above the following criteria were defined for developing the method. (1) With such a method updating of the base map should be easy and cost effective. This is because most of the AMMP surveillance areas are located in unplanned areas where densification is the order of the day. (2) The method should be simple and clear. This is because enumerators, with low levels of education, carry out the AMMP data collection. (3) The method should be able to geo-reference the data at household level. This is because the aim of the AMMP project is to collect the data at household level and (4) The method should be able to ensure quality control of the data (accuracy) and individual confidentiality. This is because health related data are sensitive and confidential.

2.4. Summary

This chapter has reviewed some geo-referencing methods, which are used in most developed countries. The goal however is to establish replicability of such methods in developing countries like Tanzania. After evaluation of the methods, with regard to the Dar es salaam situation, it is now suffice to say that direct painstakingly on ground survey method is currently the best possible method that could be used to geo-reference the AMMP data. There are various factors favouring the use of direct on ground method in Dar es Salaam city context, which can be summarised as follows;

(i) Dar es Salaam city has neither postcode system nor proper street address system.
(ii) The level of cadastre system in Dar es Salaam city is in its early stages and indeed still very poor.
(iii) The on ground survey method is flexible and therefore could be modified anyhow to suit specified requirements.

Despite its advantages, ground survey method has also some drawbacks, which require much attention in its application. It is labour intensive method and quite expensive in terms of time and money. Additionally, the method might be useful only for small project area.

3. Ground survey method and general approach

3.1. Ground survey method in the planned area

3.1.1. Introduction

After selection of a ground survey method, the next step was to develop it, according to the predefined criteria and within the framework of the AMMP survey method. This chapter explains how the method was developed prior to its implementation. As it has been already mentioned that the methodology was tested in two areas using two different approaches, the discussion in this chapter therefore is based on the two approaches adopted. In the planned area the approach used has four main parts, geo numbering, training of the enumerators, pilot survey and the survey operation. In this section the four factors are discussed and in section 3.2 the approach used in the unplanned area will be discussed.

3.1.2. Geo numbering

In the planned area updating of the base map was not given priority because it was thought that the area is relatively stable and therefore no many changes were expected. By using ArcView GIS software the base map was divided arbitrarily into four smaller working blocks and named A, B, C and D. The blocks were generated in order to enable the enumerators to ensure consistency during the survey process. The shape and the size of the blocks differ considerably. Each building on the map was assigned an ID number, which was also generated automatically from ArcView GIS software. A map, which shows four blocks, was produced and printed in A3 paper size at a scale of 1:2500. An individual map of each block was also produced and printed at a larger scale of 1:1250. The maps were produced at larger scale and printed at A3 paper size in order to enable the enumerators in the field to see the building features and their ID numbers more clearly (Map 3.1).

3.1.3. Training of the enumerators

It has been mentioned before, that in Dar es Salaam city, census interview is carried out twice in a year. Before every census round starts, enumerators have to undergo training one week before the census starts. This training is designed specifically to describe how the survey is carried out in the field, the possible problems that may arise in the field and how to handle them. Special attention is given to newly recruited enumerators. During the training programme the proposed method was properly described and demonstrated to enumerators on how it should be conducted. Further more, some basic ideas of geo-referencing were highlighted. Indeed the importance of geo-referencing the AMMP data was discussed.

A team of 16 enumerators and 4 fieldwork supervisors carried out the survey operations. All enumerators are Community Development Officers (CDOs), with college education level. They have good experiences in working with communities, however no one had ever used maps before. This created some difficulties in using maps during the survey. The task of the fieldwork supervisor is to assist the enumerators in the field whenever they encounter problems, to make close follow-ups to every enu-

merator and ensure that the census forms are completed correctly. Two fieldwork supervisors arc of university education level while the other two, one is a clinical officer with special training courses and the other one is the community development officer. It was noticed that the fieldwork supervisors had not used maps before either.

In order for the proposed method to be effectively implemented, training of the enumerators together with the supervisors was seen as an important stage towards implementing the proposed method. This is because it acquaints the enumerators with full knowledge on how the method should be carried out and indeed creates awareness on some possible problems and difficulties, which might arise during survey process and how to handle them. Clear understanding of the method to the enumerators, reduces the chances of generating unnecessary errors while conducting the survey.

3.1.4. The pilot survey

Since using maps in the field was completely a new approach to the AMMP project, the plan was to conduct a pilot survey prior to the practical implementation of the method. This would provide a general picture of whether the proposed method is feasible and understood, especially in assessing the ability of the enumerators in using maps after the training and use this experience to review the methodological approach. However, it was not possible to carry out such a survey because the AMMP had already pre-selected an area for pilot survey, which was not among the branches, which were selected as part of the case study area. This was another limitation of the research.

A different approach of pilot survey method was however designed and tested, whereby four ten-cell leaders were randomly selected, and used to assess the maps in a planned area. It should be noted that the aim of this pilot survey was not to evaluate the feasibility or validity of the methodology as it was planned before but rather to assess whether the maps could be understood by none GIS experts, or people who are not familiar with maps. Each ten-cell leader was given a map and asked to identify his/her house on the map. Three of them were able to locate quickly their houses and correctly, however one was not able to identify the house. After some guidance, and explanation of street names and unique buildings such as mosques, it was possible to finally locate the house. It was concluded that even a layman could also understand the base maps, of course this was a too general conclusion, but provided a flavour of how difficult /easy it is to read the maps. It should also be noted that the ten-cell leaders live in the area and they might be very familiar with the area while the enumerators might not be familiar with the area at all.

Conducting a pilot survey is again an important yardstick in order to test whether the proposed method is feasible. A pilot survey provides understating of the problems which might crop up during the real survey, and therefore provide opportunity to review some aspects /steps which seem to be problematic. In short a pilot project provides crude indications of areas prone to problems during the implementation process, and hence be able to take appropriate measures prior to commencement of the real survey.

Map 3:1 The planned area

(a) All Blocks

(b) Individual Block
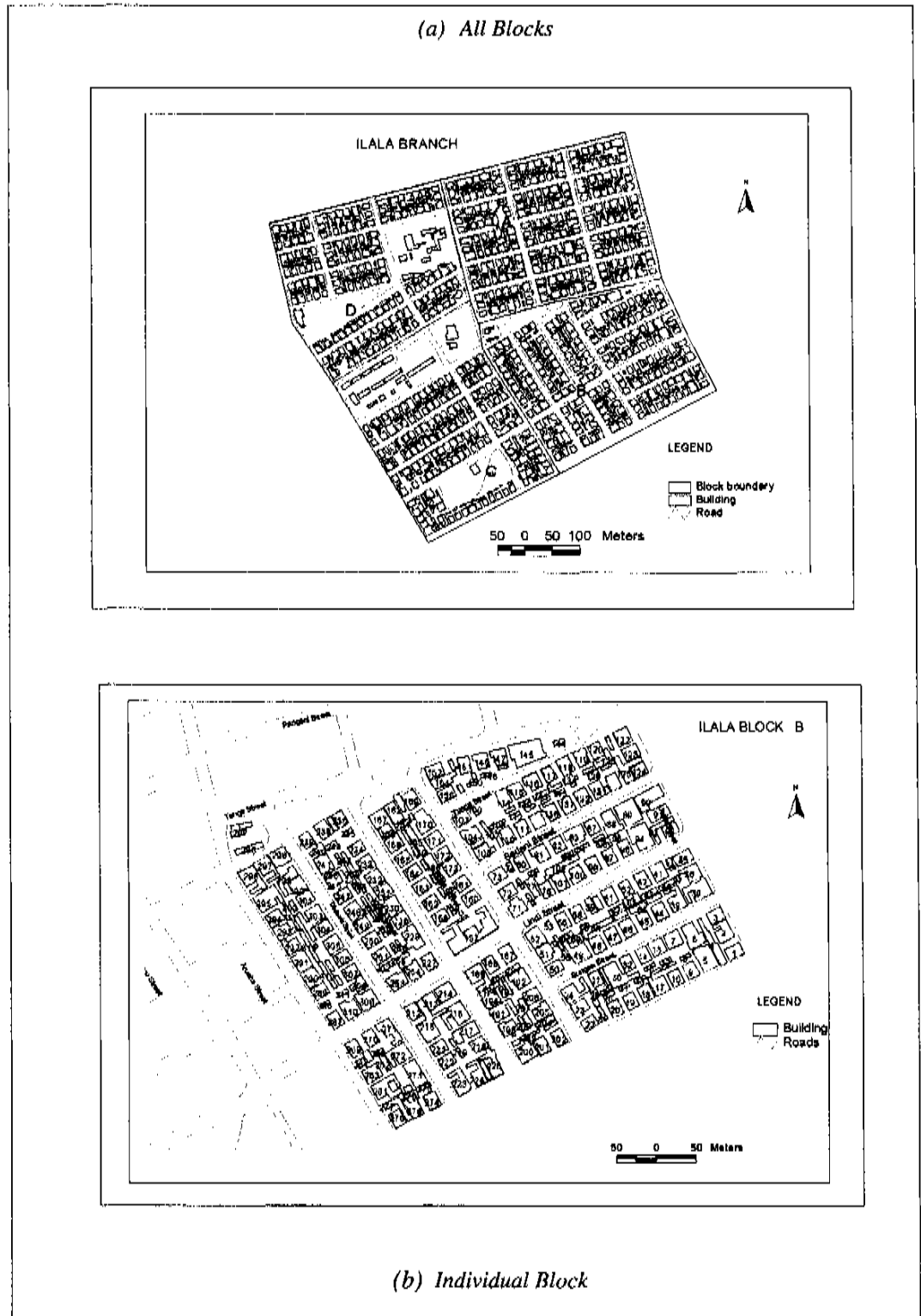
This small pilot survey revealed some potential weaknesses of the base maps, which required substantial improvements. The main improvements were made based on street names, institutional and public buildings. It was learnt that street names, public and institution building such as schools, mosques and churches which exist in the area are very important features to non GIS experts to understand maps and therefore needed to be more highlighted.

3.1.5.   The survey operation

This section explains how the survey was carried out in the area. Every enumerator was given a census form, a social economic questionnaire form, and a new form for new households, which have migrated in the area and a newsletter, which every household had to be given a copy. Together with these forms each enumerator was also provided with a set of Ilala branch maps (Maps 3.1).

During the interview every enumerator was required to identify on the map the building of the household, he /she is interviewing and record its ID number on a census form. The ID number or the geo number as commonly used in this study was recorded on top right corner of the census form below the printed date. The geo number was recorded and enclosed in a circle. This was done in order to differentiate the geo number from other numbers, which were also recorded on the form, which might create confusion. After recording properly the geo number on the census form the enumerators were required to mark the identified building on the map, with X mark. This helps to ensure consistency and avoid duplications and chances of generating unnecessary errors.

3.1.6.   Summary of survey problems

While carrying out the survey several problems arose. These problems can be summarized as follows.

  a) Digitizing problem

It was noted that some buildings were mistakenly joined together during digitizing process. The joined buildings by default were also assigned one geo number (See Figure 3.1). That means all households found in the joined buildings could be registered under one building. Such problem needs considerable attention as it might have greater implications on the analysis part such as, calculating occupancy rate or overcrowding.
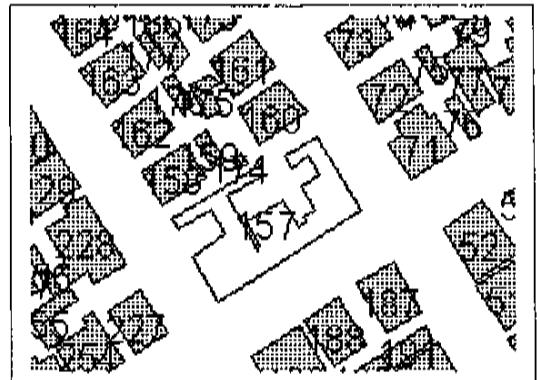
Figure 3:1. Joined buildings

It might also have direct effect on spatial pattern analysis of mortality data. The problem was resolved by requesting the enumerators to draw a line to separate the joined building into two parts as *a* and *b*, and record the geo number as e.g. *157a* and *157b* or *157c* depending on the number of the combined buildings.

b) Updating problem

Ilala branch map was not updated, as the area was considered to be planned and relatively stable, that is no many changes had occurred. However cases were noted where two or more buildings had been demolished and instead one big multi-storey building has been erected. Such a situation caused a bit of confusion during survey. Again in order to handle such problems only one of the geo numbers of those buildings was recoded on the census form, and a line drawn to enclose the two buildings as one. This situation suggests that there is still a need to update the base map of the area.

c) House numbering problem

The base maps, which were used in the survey, did not contain any house IDs that are provided by the city council. The method used to assign geo numbers on buildings did not base on property IDs either. Geo numbers were generated automatically using Arc View GIS software. However it was noted that it would be much more useful if geo numbers would be assigned based on house IDs provided by the city council, as this might help to ensure consistence and maintain accuracy during survey. This is because the IDs are clearly labelled on each house. This weakness was observed even before the survey was conducted; however identification and the use of house IDs provided by the city council is beyond the scope of this study.

d) Spatial unit problem

According to the research approach, block as a spatial unit was introduced. However it was learnt that the AMMP census survey is carried out based on the ten-cell units. That means, during survey each enumerator is assigned to work in a particular ten-cell unit(s). This approach created problems in a sense that blocks were just divided arbitrarily without considering ten-cell units boundaries, therefore there was an overlap between blocks and ten-cell units (Figure 3.2). This was seen as a significant problem because the boundaries of the ten-cell unit do not exist and therefore were not indicated on the base map either. This problem may cause enumerators to generate errors because in such a situation consistence control becomes difficult. This is one of the aspects of the proposed method, which need some more attention in the future. The best possible solution to solve such problem is, the blocks should be defined based on the ten-cell units so as to maintain the ten-cell unit structure.
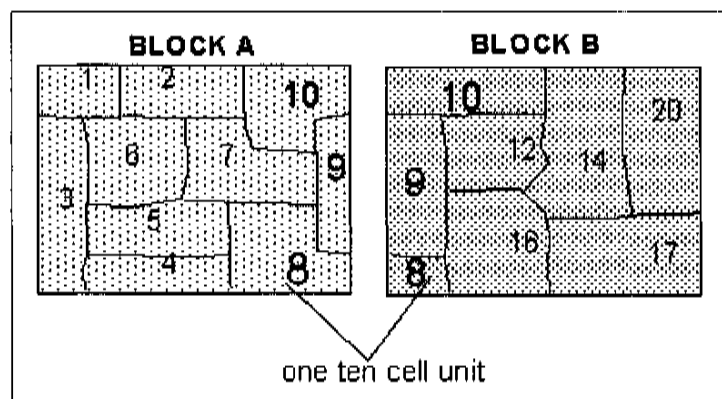
Figure 3:2. Spatial unit problem

e) Visualization problem

From the base maps it was noticed that some geo numbers were not clear and some were too small to read. This weakness suggests improvement on numbering system.

3.1.7. Accuracy

It is one thing to be able to avoid errors but it is another matter to entirely be able to correct them. Any data collection method has its strengths and weaknesses; indeed selection of a method of data collection mainly depends on its degree of accuracy. There is no method, which is error free. For the purpose of accuracy control an error check up mechanism was developed in order to evaluate the degree of accuracy attained during the survey process.

The data shows that there are 1919 households in Ilala branch, the number that is equal to the number of census forms. In order to assess the degree of accuracy in recording the geo numbers on the census form, 10 % of the census forms were randomly selected and verified in the field by the help of the key informant person and the local residents who are very familiar with the area. A house number was used to assess the level of accuracy of the method. Each census form has a variable called house number, which refers to the house number written on top of the door of each house in the area. Where there is no house number the name of the owner of the house is written on the form. The map was used to cross check with the completed form if the geo number, which is recorded on the form, is the same as that on the map. Like wise the house number on the form was also compared with that on the door.

A ten percent sample might be too small to reflect reality, but it was used mainly because, verification, had to be done by a researcher alone and not the enumerators. Hence it was small enough to be manageable. A total number of 192 forms were selected. After careful examination of the 192 forms, and verification in the field, it was found that 185 forms, which is 96.3 %, were correctly completed, while 7 forms were not correct. This indicates that, in recording the geo numbers, there was an error of 3.7 %. This means, there is a probability that 71 forms out of 1919 might be recoded wrongly. If we refer to Cowen (1997) who argues that a failure rate of address matching of 10% is regarded as good, this result was and will remain to be invaluable.

3.1.8. Final out put

After the field survey the data with geo number was entered in Microsoft excel, and the database generated in Microsoft excel was then exported in the ArcView GIS software where the data was linked with spatial database. This is an operation, which was carried out using ArcView GIS software to establish the spatial database. Figure 3.3 shows part of the final output of the methodology in the planned area. In general this is how the methodology was developed and tested in the planned area.

Figure 3:3 Spatial database.

3.2. Ground survey method in the unplanned area

3.2.1. Introduction

Section 3.2 discussed the methodology approach, which was used in the planned area and also depicts some potential survey problems, which require some improvements. In this section an approach, which was used in the unplanned area, will be discussed. Generally in the unplanned area the approach consists of two main parts, the problem of lack of up-to-date data and the survey process.

The section starts by making a general introduction on Small Format Aerial Photographs (SFAP). Then it discusses the problem of lack of up-to-date data in the area, which requires updating of the existing base map. It provides a step-by-step description of how the base map was updated using the aerial photograph mosaic. It explains the process of creating a mosaic and then how a mosaic was used as a tool to update a base maps.

The current aerial photographs taken in July 2001, by Drs. Paul Hofstee of ITC were used to update the base map. However because of the quality of these aerial photos, most of which were too oblique, it was not possible to create a mosaic, which covers the whole of the unplanned area. Due to this problem a mosaic of only part of the area was generated. In updating the data, the main interest is to monitor the changes, which have taken place in the area between the period of 1992-2001 by assessing and identifying the total number of new buildings, which have been constructed between these periods. This could indicate the rate at which the area is growing spatially.

3.2.2.   Lack of up-to-date data

(i) Small Format Arial Photograph (SFAP)

Aerial photographic professionals have made a broad distinction of aerial photographs between large scale format and small scale format aerial photographs.

Large-scale format or vertical aerial photography provides a vertical view and systematic stereo coverage. The scale is uniform all over the photo as on the map. The aircraft usually has to be modified to install the camera system. In most mapping applications large-scale format aerial photographs are used (Graham and Real, 1986).

On the other hand small format or oblique aerial photograph provides a bird's eye view. It shows perspective and gives a better idea of vertical dimensions than a single large-scale format vertical aerial photograph. Oblique photographs are obtained when the axis of the camera is not vertical. No modification to the aircraft is needed and nearly all light air planes can be used to take this kind of photos.

There are number of reasons favouring the use of small format aerial photographs verses the large scale format, these reasons include (Hofstee, 1990;Graham and Real, 1986):

(a) The cost of such small format aerial photographs is very low compared to the large-scale format.

(b) Small format aerial photography flights can easily be organized and implemented without specialized equipment and organization.

(c) People not familiar with aerial photos or maps very easily understand small format aerial photographs

(d) The technique of small format aerial photography is very simple if it has to be compared with the large-scale format aerial photography. Photos are taken with hand–held camera, while the airplane is circling the area to be photographed.

Although SFAPs have many advantages to large-scale format aerial photos, they have some limitations in their use in that (Hofstee, 1990;Graham and Real, 1986):

(a) Systematic coverage of the area cannot be realized

(b) For precise mapping purposes small format aerial photos are not optimal, because scale is not uniform all over the photo as it is in large-scale format aerial photos.

(c) The perspective image is not easily transformed into orthogonal map projection but it can be done quite well when accuracy requirements are not very high.

Apart from its advantages and disadvantages, an important aspect regarding taking the oblique photographs is a pre flight planning in order to ensure correct coverage of the project area as well as economic operation. Graham and Real (1986) asserted that in order to obtain the good quality small format aerial photos, the most essential variables, which must be carefully observed in the mission of making this type of photographs, are altitude, camera angle and offset distance from the picture centre. Graham and Real added that the most difficult variable however is to control the angle of obliquity. Scale and coverage of a single image are always estimated.

Small Format Aerial Photographs have proved to be useful tool for many urban applications. The widely recognized field of application of SFAP is in monitoring changes in unplanned urban areas (Hofstee, 1990). Densification can easily be recognized by comparing two photos of different years, or by using the most common methods of updating the existing topographic maps using a mosaic.

The use of SFAP for monitoring rapid changes in informal areas have been applies in many places. For instance, the SFAP has been used in Bangkok metropolitan in Thailand to analyse housing and land use problems (Chamber, 2001). Small format aerial photographs have been used also in Tanzania for monitoring and upgrading of Manzese and Hanna Nassif squatter settlements (Bruijin, 1987;Sliuzas, 1988).
Unlike maps, which represent the physical terrain with generalised symbols and colours small format, aerial photographs reveal the terrain, as it exists in real life. Therefore small format aerial photography could be extremely useful for evaluation and analysis of urban areas development.

(ii) SFAP mosaic and updating of the base map
There are basically two main steps, which need to be followed when one intends to update a base map by using small format aerial photography. These steps include creation of a mosaic and digitising of features of interest. However it is worthy mentioning that this method is useful only for small project areas, with specific mission, as small format aerial photographs do not often cover large areas. The data required for updating process include, current Oblique aerial photographs that best covers the case study area, the topographical map of the area (digital or analogue) and the Digital Elevation Model (DEM). The use of DEM is sometimes optional, depends on the method one intends to use.

(iii) Creation of a small format aerial photography mosaic
Creating a mosaic or Mosaicking is a process of assembling individual rectified photographs and systematically fitting them together to form a complete view of an area covered by photographs. Generating a mosaic of either vertical or oblique aerial photography is not a straightforward process. It involves long and sometimes very complex process. As it needs to bring together all the individual photographs in a common scale and rectify the camera tilt through control points whose position are determined from an existing topographical map or field survey. Below is an outline of the procedure, which has been used in creating a small format aerial photography mosaic for part of the unplanned area (Relini).

(a) Scanning of the photography
A set of five recent small format oblique aerial photographs, taken in July 2001, was used to create a mosaic. The photographs were first scanned and saved in a TIF file format before they were imported to ILWIS, the software that was used to generate the mosaic.

(b) Geo-referencing of the photography
The second step was to geo-reference each individual photograph. The geo-referencing referred here, of course is associating aerial photograph with the real world coordinates. Since the geo-referencing method relies on ground control points, the control points were determined using the existing digital topographic base map of 1992. In ILWIS, geo-referencing of small format aerial photography can be done by first creating geo-reference direct linear. The geo-reference direct linear is always applicable when the photographs are taken with normal camera and without fiducial marks, when the terrain covered by the photograph has clear height differences and therefore there is a need to correct for tilt and relief displacement and when the Digital Elevation Model (DEM) of the area is available. In this particular case, geo-reference direct linear was used.

(c) Image matching

After creating the geo-reference direct linear and display the photograph, ILWIS allows directly to do screen digitizing on the displayed non-rectified photograph. Common points from both the photo and the topographical map, in this case the corner of the buildings were used, as matching points to obtain XY real world coordinates and the DEM was used to obtain the height (Z) value. The DEM was densified to a pixel size of 0.2 meters. A geo-reference direct linear in ILWIS requires at least 6 tie points, also called control points before it starts showing the margin of error (sigma). Indeed, in order to have good result it is advisable to have as many tie points as possible and to make sure that they are well distributed on the image. The Digital Elevation Model of the area is included in Appendix 3.

(d) Resampling of geo-referenced images

After geo-referencing the image, the result is that, the image gets the real world coordinates, but it is not north-south oriented, thus the orientation of geo-referenced image and the topographical map to be updated are different. Interpolation is therefore required to resample the image, so as to bring it into a common north-south orientation. There are various resampling algorithms available in ILWIS software. The main methods are nearest neighbour, bilinear interpolation and cubic convolution. However the choice of the resampling algorithms depends on among others the ratio between input and output pixel size and the purpose of the resampled data. For the purposes of this study, nearest neighbour was chosen. The photographs were resampled to a new geo-reference with resolution of pixel size of 0.2 meters. This resolution allows for detail analysis of smallest objects on the map.

(e) Gluing multiple images

For final creation of mosaic a glue operation was used. In ILWIS the glue raster maps operation glues or merges two or more geo-referenced input raster maps into one output raster map. The output map then comprises the total area of all input maps. The same operation is used to glue the rectified aerial photos, and produce a mosaic. Figure 3.4 shows the final mosaic of part of the unplanned area.
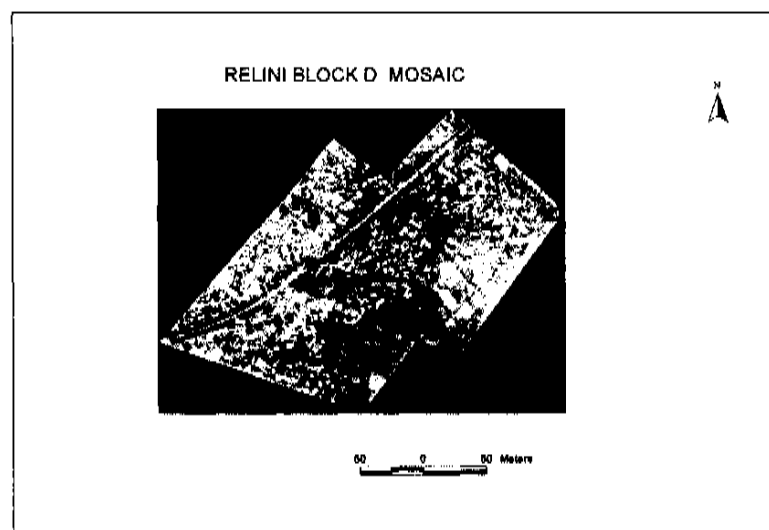
Figure 3:4 An aerial photography mosaic of part of the unplanned area

(iv) Updating topographical map

ILWIS software allows for on screen digitizing. Compared to manual digitizing, the main advantage of on screen digitizing is that features are drawn on top of the raster image as they are recorded. In this way the control of digitizing is improved and number of digitizing errors reduced. The software enables digitizing errors to be graphically easily traced on the screen and corrective editing can then be carried out. Therefore the accuracy of digitizing can be directly inspected, omissions are easily detected and duplications avoided.

Finally the base map, with buildings that existed in 1992, was overlaid with the photo mosaic. All buildings on the photo mosaic, which existed in 1992,overlaped with the buildings on the map. The new buildings, which were constructed after 1992 had no overlap on the mosaic and each building, were therefore digitized, to produce a map of new buildings constructed between the periods of 1992-2001.

The original building map (base map) of 1992 and that of the new buildings were then overlaid to form one up-to-date map. Map 3.3 shows the final output map.

Map 3:2 The updated part of the unplanned area (Relini block D)
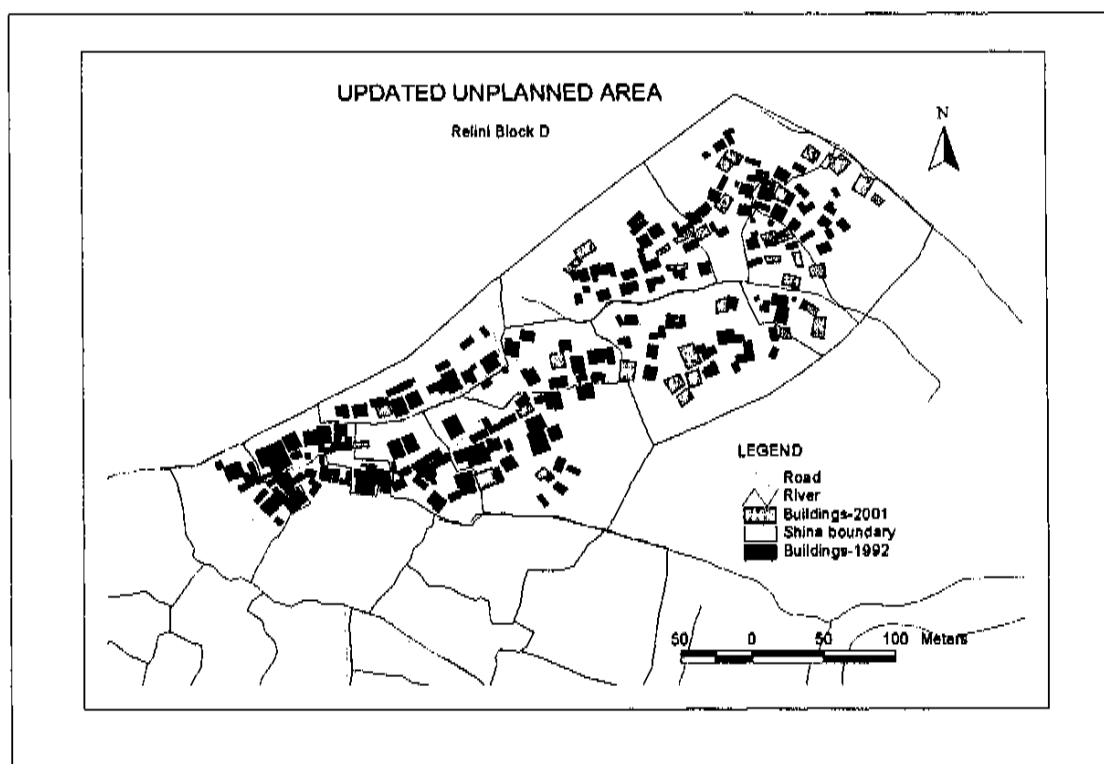
(v) Geo numbering

Both the updated area and the area, which was not updated, were again divided into four smaller working blocks. Each building in each block was also assigned a unique ID number. A map, which shows four blocks, was produced and printed in A3 paper size at a scale of 1:2500. An individual map of each block was also produced and printed in the same A3 paper size at a larger scale of 1:1250 (Map 3.3)

Map 3:3 The unplanned area

(a) All Blocks

(b) Individual Block
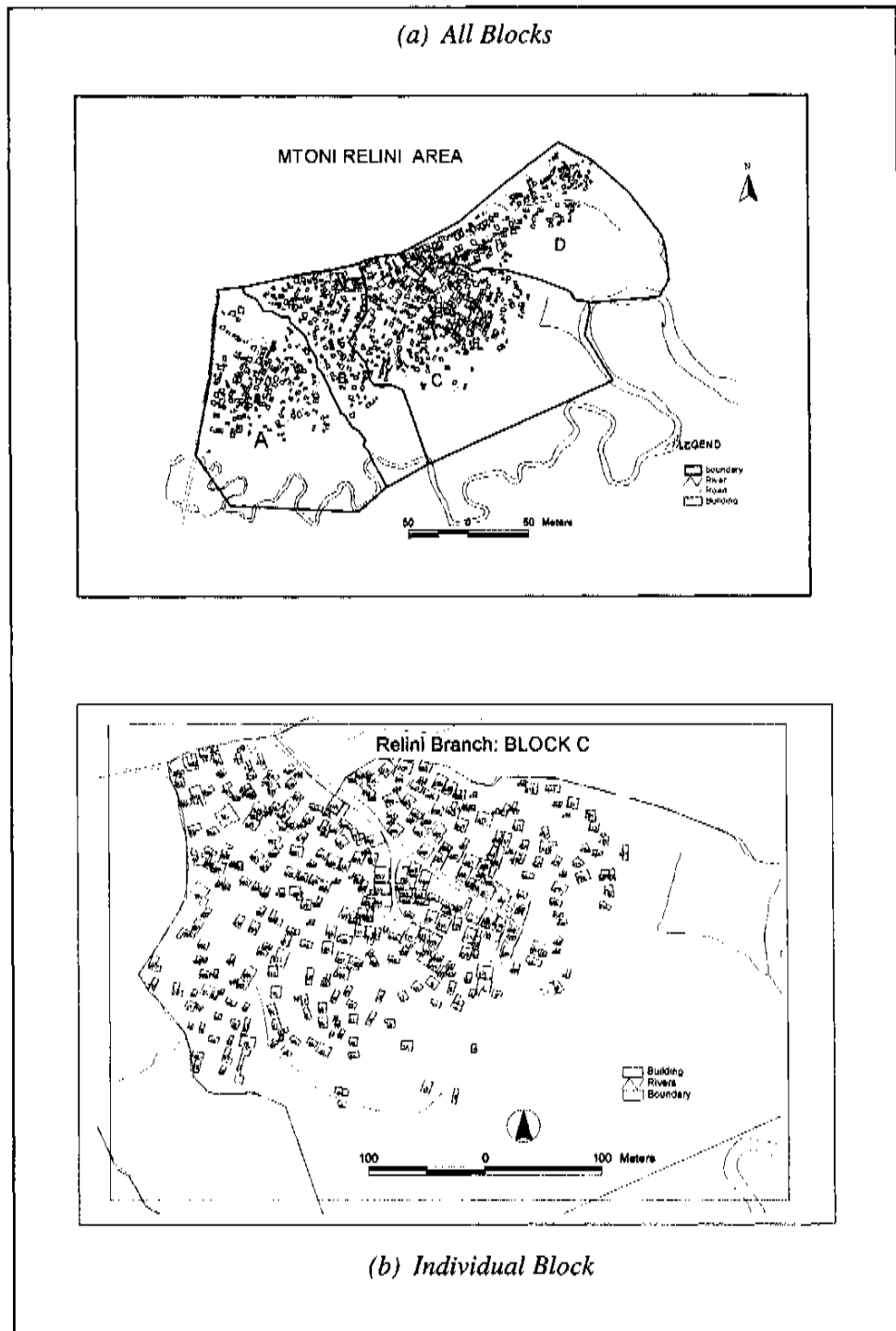
3.2.3. The survey operation

In the unplanned area the data was geo referenced at the ten-cell unit level. It was revealed that the area consist of 60 ten cell units. With the help of the key informant person and the local residents, every ten-cell leader in the area was visited and his/her house was identified on the map and its geo number was recorded. The ten-cell leaders were further asked to indicate their area of jurisdictions. The identified boundary was then plotted on the map (See appendix 4). The task of identifying the boundaries seemed to be difficult as no geographical boundary exists at this level. Therefore it should be noted that there was some element of "guess work" involved here, because delineation of boundaries of the units mainly was based on number of houses under each ten-cell leader. Hence no answer could be given to questions such as why a certain unit bears a particular shape.

Supported with an aerial photography mosaic, locating houses of the ten-cell leaders in block D was found to be much more easy, compared to other blocks. It was found that with the help of mosaic the task of identifying and locating houses in unplanned area becomes easy and fast. With this perspective aerial photography mosaic was seen as a very useful tool in geo-referencing process because it provides a real life view of the area at very high resolution. Thus there is a need to generate a mosaic not only for unplanned areas but also for planned areas.

In general this is how the methodology was developed and implemented in the unplanned area.

3.2.4. The Survey Problems

There is one problem, which was cited during survey in the unplanned area, which might be seen as trivial but need to be considered if such approach has to be adopted.

As it was indicated before that no boundary exists at this level, thus units were defined based on the area covered by houses under the ten-cell leader. It was observed that some ten-cell leaders were not able to define their geographical area coverage. That means they do not know exactly how many households /houses belong to their units. It was also interesting to note that even some residents do not know their ten-cell leaders. It was noticed that such situation occurs mainly because of two reasons, first, the unit is no longer active. This is because; the unit was established when the government was under one party system and now the government has changed to multiparty party system. However it was also interesting to note that all ten-cell leaders in the area are members of ruling party (CCM). Secondly, due to various reasons, ten- cell leaders change frequently. It was observed that from June 2001 until October 2001, 6 ten-cell leaders have left their job because of various reasons such as deaths and migration out side the surveillance site. It was reported that no election was planned to take place recently to fill the places. No further investigation was made to identify how the election of ten cell leaders is conducted.

In general, ambiguity in ten-cell unit boundaries makes the possibility of utilizing the ten cell, as a spatial unit for geo-referencing and analysis much more difficult. Further investigation need to be done in order to find out ways to improve such problem. This is because; boundaries of spatial unit for geo-referencing and analysis need to be clear and never overlap.

3.3. Summary

This chapter has made an attempt to demonstrate how the methodology was developed and implemented in the two areas. The direct on ground survey method seems to be a useful method. The indispensable results obtained from both areas show great promise of the method and could be used for further improvement of the method. However the approach used in an unplanned area may encounter significant limitations in terms of cost e.g. in obtaining up-to-date small format aerial photography.

An aerial photography mosaic proves to be a useful tool in the process of geo-referencing data in the unplanned area. However it should be noted that by itself can not be used as a proxy to other tools like maps, rather it could be used as a complementary tool.

The approach used in the planned area points that the method could be extremely useful in geo-referencing the AMMP data, nevertheless there is a possibility that the methodology could be integrated in the AMMP traditional data collection method. Though the result suggests that integration of the method would require some "minor" improvements as discussed in section 3.1.6, there are several factors favouring the integration of the method as they have been supported by the AMMP team. With the use of maps in the field the survey process becomes easier and takes less time, maps help to ensure consistence and accuracy during survey and therefore chances of missing a household are reduced and it is easy to locate where you are, while in the field.

4. Data analysis and exploration

4.1. Introduction

After geo-referencing the data, the intention of this chapter is to examine the usefulness of the geo-referenced data at household level. However, since mortality data is not geo-referenced at the same level (household) with the social economic data, the analysis and exploration of the data will therefore base on the ten-cell unit level. The chapter has three main components, the general conceptual framework, the analysis and discussion of the result, and the concluding remarks.

4.1.2 Conceptual Framework

This section provides an overview of the approach, on how the analysis and exploration of the data is done. It first discusses some health determinant factors. Secondly, it sets out a framework on how the two selected determinant factors are analysed and associated with the general mortality conditions in the two areas.

Literature suggests that there are various factors, which contribute to ill heath of the people. To what extent or in what way these factors impact health varies, however impacts of some factors are not clear and still under debates. In this section several factors will be discussed. Figure 4.1 illustrates a summary of factors, which may increase the risk of mortality and morbidity.

(i) Behavioural factors [life style]

Gatrell (2002) states four aspects of behavior which might contribute to poor health; these include diet, alcohol consumption, smoking, and sexual conduct. He points out that excess consumption of fat, salt and unrefined sugar may lead to heart diseases. Smoking is well established as a risk factor for more chronic diseases such as lung cancer. Alcoholism is one of the biggest cause of social and health problem in the population worldwide (AMMP, 1997). Other behaviours like use of illegal drugs and number of sexual partners, may lead to high risk of mortality. These factors vary with socio economic status, however more often people with low income or low socio economic status are more likely to engage on health damaging behaviours.

Family background may also influences education opportunity and attainment, which in turn is a good indicator of employment and income, which have direct impact on health (Gatrell, 2002).

(ii) Overcrowding

Overcrowding can affect health in many ways. For instance, overcrowding increases the risk of multiple infections. This is because the number of potential transmitters increases. Overcrowding can also increase the risk of disease transmission e.g. for some infections living in a small room or sleeping in the same bed the risk of transmitting the diseases increases. Bradley *et al.*(quoted in Clauson-Kaas *et al.*, 1996) point that in the recent past a number of studies have shown that diseases such as whooping, cough, polio, diarrhoea, malaria, meningitis, acute respiratory infection and influenza may be related to overcrowding. However they caution that the mechanism explaining how overcrowding influences the spread of and severity of these diseases and many others is still unclear.

(iii)   Working conditions

The environments within which people work has major impacts on health. Many occupations in which people with low socio status are employed can be quite hazardous. The risk could be from the exposure to dangerous machinery or chemical or poor work place practices, which as a result may increase the risk of mortality (Gatrell, 2002).

(iv)   Social and community influence

Social exclusion and lack of community involvement may increase the risk of mortality. A study carried out by Durkheim (quoted in Gatrell, 2002) shows that suicide was more common among socially isolated individuals.
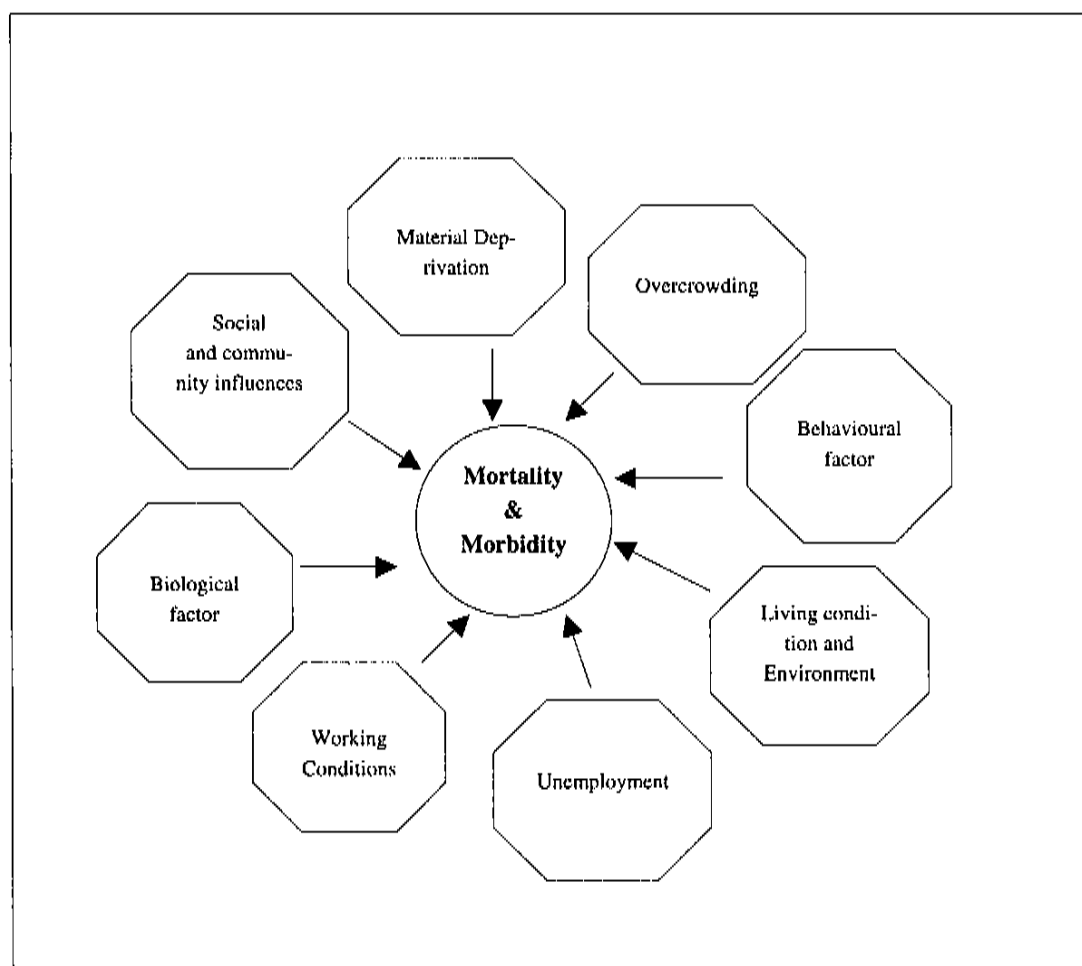
Figure 4:1 Health determant facfors

(v) Material deprivation

Lack of wealth and income, poor quality housing, lack of services, amenities and poor physical environment lead to material deprivation. There is a well recognized link between material deprivation and poor health. This is demonstrated by a study carried out by Villerme (quoted in Gatrell, 2002), which showed strong relationship between mortality rate and poverty.

(vi) Biological factor

Notwithstanding many criticisms of whether diseases such diabetes and chronic bronchitis and cardiovascular diseases are biologically formed in the early stage in the womb or in infancy, Scientific evidence has been provided to support this theory (Gatrell, 2002) that such diseases are biological in nature.

(vii) Unemployment

Like employment with adequate wages contributes to positive health, unemployment has a strong, negative impact on health. Various studies examining the association between unemployment and a variety of health outcomes concluded that the evidence suggests a strong association between unemployment and many adverse health outcomes. Unemployed people have higher mortality rates than employed. Becoming jobless and prolonged redundancy have negative effects on health and increase the risk of premature death. The effects of unemployment are mainly assumed to be mediated through increased psychosocial stress, tobacco and alcohol consumption, as well as loss of income and material deprivation (Martikainen, 1990)

(viii) Living condition and the environment

People living near potential sources of pollution are always at high risk of mortality. Gatrell (2002) argue that more often the burden of environmental pollution is among the low-income groups. A study carried out by Schell and Czerwinski (quoted in Gatrell, 2002) show that the poorest groups in USA and New Zealand have the highest exposure to cadmium and lead. The study also revealed most of Black and Hispanic Americans live in communities closer to landfill sites that receive hazardous waste and therefore increases their risk of mortality/morbidity.

This study is interested to look on mainly two aspects, which are considered to be among the health determinant factors in the two areas. These aspects include unemployment and overcrowding conditions. The reason for choice of these two aspects is mainly due the fact that it is widely believed that people with low income and people, who live in overcrowded homes, are always at high risk of mortality/morbidity (Clauson-Kaas et al, 1996;Gatrell, 2002).

Thus the study intends to investigate the relationship between the two aspects and the mortality condition in the two areas. Four expectations have been formulated, which will act as a benchmark towards achieving this objective.
1. It is expected that in the unplanned area there is higher mortality rate than in the planned area.
2. Spatial distribution patterns of cases of death exist in the two areas.
3. It is expected that where unemployment rate is high, mortality rate is also high.
4. Where overcrowding condition is high, mortality rate tends also to be high.

The analysis is carried out in line with these four expectations. However it is important to note that in this study context these four aspects are considered as expectations rather than hypothesis. The discussion of the result is focusing on testing whether the result reflects reality, in other words whether the result matches the expectations. Based on the assessment of the result and comparison, conclusions about general mortality condition in the two areas will be drawn. Figure 4.2 illustrates the principal elements of the conceptual framework.
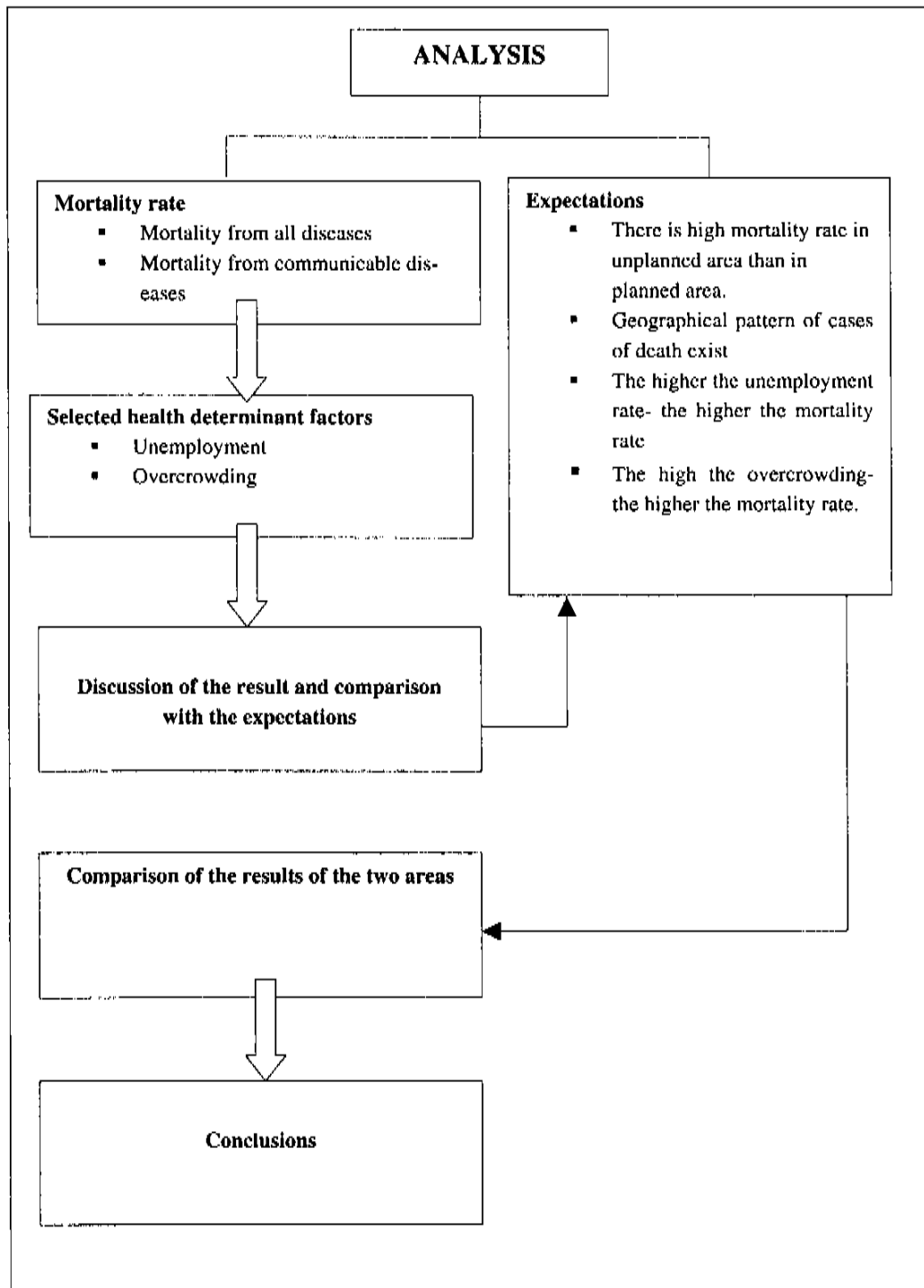
Figure 4:2 Conceptual framework for data analysis and exploration

4.2.   Analysis procedure

The analysis was carried out using SPSS and ArcView GIS software. The two files, (mortality and socio economic) were first exported to SPSS statistical package where the data were aggregated at a ten-cell unit level. The mortality rate of each unit was then calculated and classified into appropriate classes. The data were then imported into ArcView GIS software where thematic maps were generated. Basically, the SPSS software was used to calculate the number of deaths in each ten-cell unit by cause, age and sex[2]. It was also used to rank and classify the variables, calculate simple regression models and spearman's rank correlation coefficient. ArcView GIS was used to link the two data files with spatial database and, calculate density and to produce thematic maps. Figure 4.3 shows the analytical procedure, which is followed.
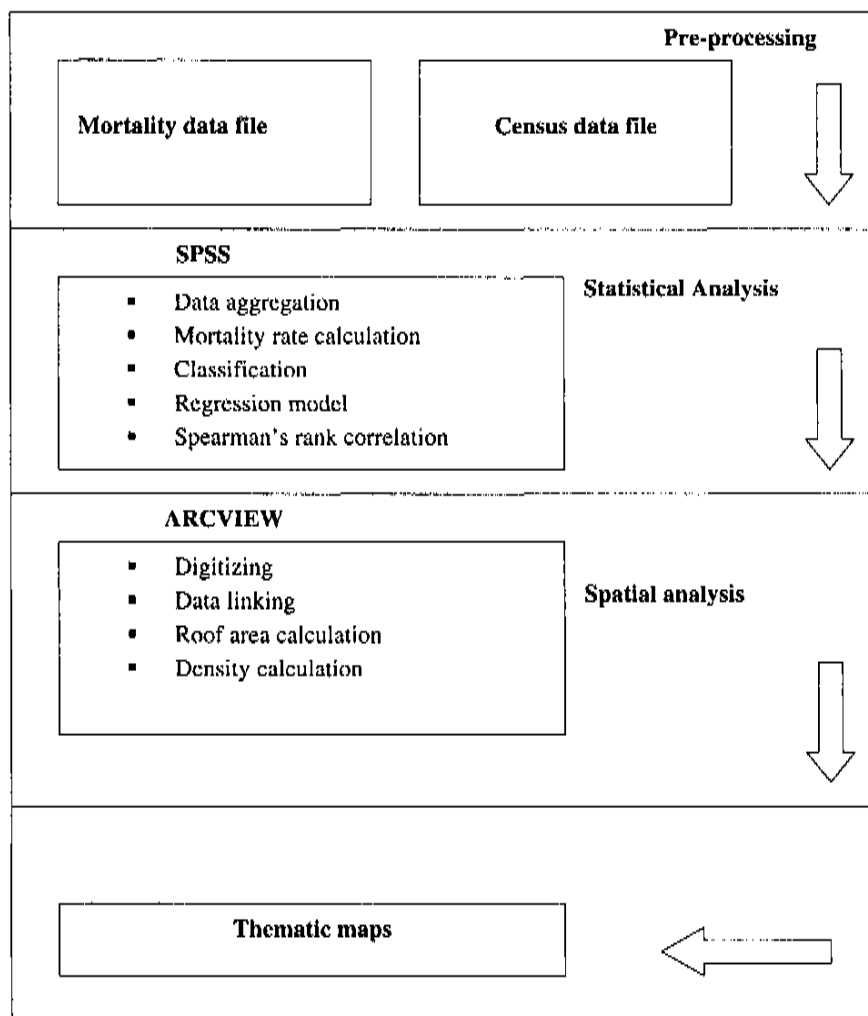
Figure 4:3 Data analysis procedure

---

[2] Note that because of few cases being observed, the analysis does not consider sex and age classes.

4.2.1 Mortality condition

4.2.1.1 Introduction

The Cambridge English dictionary defines mortality as the number of deaths in a given time or a given community, or the proportion of deaths to population, or to a specific number of the population.

The first expectation is that, there is higher mortality rate in the unplanned area than in the planned area. In order to examine this expectation the mortality rates of the two areas were calculated using the following formula.

Mortality Rate=d/p*1000/t

Where

$d=$ *number of death*

$p=$ *total population*

$t=$ *number of years*

The AMMP project classifies causes of death in three levels hierarchical system (See appendix 5). The highest level, which is level 3, has 4 classes. (1) Communicable diseases (2) Non communicable diseases (3) Direct maternal causes and (4) External causes. The mortality status of the two areas is examined based on a broader classification level three. The main reason for selecting a higher level of classification is that, few cases of mortality incidence at levels two and one were noted.

The mortality data shows that, a total number of 194 cases of death were recorded in the planned area while a total number of 89 cases were recorded in the unplanned area, for a period of two years, 1999 and 2000.The two areas have different size of population and different geographical coverages. The planned area has a population of about 7913 people with an area of approximately 37 hectares while the unplanned area has a population of about 6307 people with area coverage of about 35 hectares.

The mortality rate (death per 1000 population) in the planned area was found to be 12.3 while in the unplanned area is only 7.1. This result shows that there is high mortality rate in the planned area than in the unplanned area, which is opposite to our expectation. Since the calculation was based on a two years period only, further verification of the result was necessary. As a justification for this situation, mortality rates for 7 years period were calculated (1994–2000) to see whether the trend is similar. The result shows that in most years mortality rate is higher in the unplanned area except for 1999. The data shows that in 1999, the planned area had more deaths than in the unplanned area and even higher than in any other year before (Figure 4.4). This exceptionally high number of deaths of 1999 affected the result in one way or another. On average it can be said that, in these two areas mortality rate is high in the unplanned area than in the planned one. However if we closely examine the graph, the trend shows that from 1995–2000,the mortality rate in the planned area has been increasing while in the unplanned area it is gradually decreasing. If the conclusion is to base only on this data this graph shows that there is health improvement in the unplanned area. However further investigation is required to verify this result. This is because in order to make firm conclusion on this situation more information is required.

On the other hand in the past seven years, the population size in the planned area has been decreasing considerably, while the population size of the unplanned area is increasing (Figure 4.4). Even if there is no any gentrification study carried out, this decrease of population is an indication that gentrification is taking place in the planned area[3].
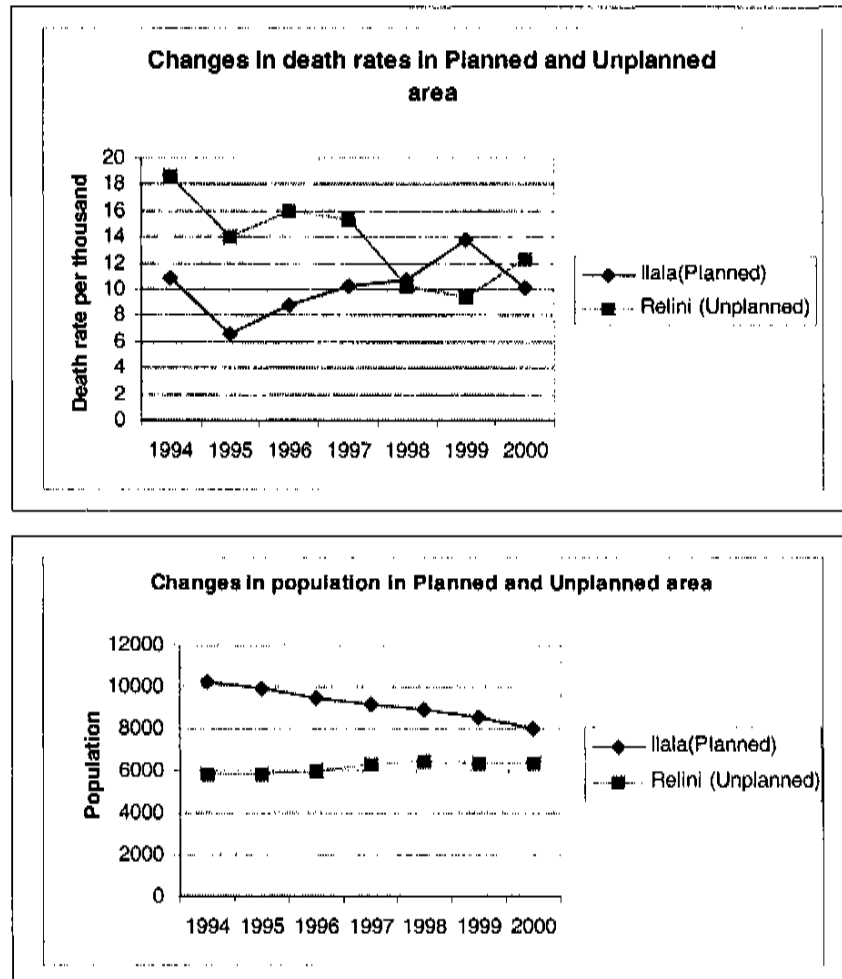
Figure 4:4 Changes in death rate and population in the planned and the unplanned area

From the analysis it is clear that communicable disease is the main cause of deaths in both areas (Table 4.1). The result indicates that in both areas, deaths caused by communicable disease account for more than 50% of the total deaths, which occurred in two year periods. In both areas non-communicable disease is the second cause of death, which account for 23.2% and 16.9% in the planned and the un-planned area respectively. In the unplanned area direct maternal cause accounts for large percentage of death, of 5.6% than in the planned area, which is only 1.0 %. This situation could be explained, as women in the unplanned area have less access to health facilities than in planned area mainly because of poverty. It has been observed that in the planned area deaths caused by external causes, which the data indicates that these are mainly due to traffic accidents, account for 4.1% of the total deaths in the area. This proportion is higher, compared to, that found in the unplanned area. In the unplanned area external causes account for only 2.2% of the total number of death in the area. This shows that in

---

[3] Source: Personal communication with local residents.

planned areas we find more traffic problems and therefore more deaths likely to be caused by traffic accidents than in unplanned areas.

Table 4:1 Number of cases of death recorded in 1999/2000

	Causes of death	Planned area		Unplanned area	
		Total	%	Total	%
1	**Communicable disease**	**109**	**56.2**	**51**	**57.3**
2	Non-communicable disease	45	23.2	15	16.9
3	Direct maternal cause	2	1.0	5	5.6
4	External causes	8	4.1	2	2.2
5	Unclassified[4]	30	15.5	16	18.0
	TOTAL	**194**	**100**	**89**	**100**

In sub sections 4.2.1.2 and 4.2.1.3, deaths caused by all diseases and deaths caused by communicable diseases only will be examined. The aim is to find out if there is any geographical pattern of deaths in the two settlements and to identify areas (units) with higher mortality rates. This will quantify our second expectation that: There are geographical patterns of cases of death in the two areas.

4.2.1.2 Mortality from all diseases

Maps, which show absolute number of deaths from all diseases, in both areas, were generated using the procedure indicated in figure 4.3. Though in both areas the pattern is not clear enough, in the planned area, more deaths could be observed in the northern and southern part of the area. In these two parts, units with more cases are located. However in the unplanned area deaths are more or less concentrated on the central part of the area and few cases can be observed in the rest of the area. Significant difference of number of deaths among individual ten cell units can also be noted in both areas. In planned area the number of deaths in a unit ranges from 1 to 10 cases while in unplanned area it ranges from 1 to 8 cases (Maps 4.1 and 4.2).

---

[4] This class includes all cases which for various reasons their causes were not classified.

Map 4.3 and Map 4.4 show mortality rate from all diseases per ten-cell unit in the planned and the unplanned areas respectively. In both areas there is considerably variations in death per 100 population among ten cell units, although hardly any clear geographical pattern again could be observed. However the result gives us a clear indication of internal variations of mortality rate among units. For instance in unplanned area two units have the highest mortality rate. In section 4.2.2, the mortality rates of both areas will be associated with unemployment rates to see whether there is any relationship.

4.2.1.3 Mortality from communicable disease

Maps 4.5 and 4.6 show the absolute numbers of deaths caused by communicable diseases in the planned and the unplanned area respectively. In the planned area more deaths caused by communicable diseases are observed in the upper and lower parts of the area. The central part is characterized by few cases of deaths. On the other hand in the unplanned area more cases are located in the central part of the area. The number of cases in the unplanned area ranges from 1- 4 and in the planned area ranges from 1-9. In general it is observed that there is very little indication of spatial pattern of deaths caused by communicable diseases and all causes in both areas. Even though there is no clear spatial distribution pattern of deaths yet we could see where more cases are located in both areas. Many factors could influence these variations among units. However this study is interested to find out if these variations are associated with overcrowding condition in the area. This will be discussed in section 4.2.3.

4.2.2 Unemployment Condition

4.2.2.1 Introduction

In this study context, the AMMP definition of the term unemployment will be used (See appendix 4). Basically, the AMMP classifies occupational status of the population in the area into two main groups, working group and non-working group. Under working group there are two subgroups, one, group of people who work and earn income (income earners group) and two, group of people who work but do not earn income (non income earners group) e.g. house wife. Under non-working group there are three sub groups, which are (1) unemployed group, this is a group of people who are able to work but currently are not employed, (2) Group of people who cannot work, which includes, disabled, old people aged above 60, blind people, children under age 7 and people who are classified as permanently sick and (3) Student group. In this study, unemployment will refer to that group of people who are currently not employed but are seeking to work. It will exclude group of people who cannot work and students group.

From the socio economic data, it is found that in the planned area, out of 7913 people, unemployed group consists of 570 people, which is 7.3% of the total population in the area. While in the unplanned area, out of 6307 people, the unemployed population group consists of 447 people, which is 7.2% of the total population in the area. However it is important to note that, due to the problem of definition of occupation codes, in unplanned area a total number 103 records were excluded in the calculation, while in unplanned area a total number of 76 records were excluded. Table 4.2 shows employment classification in the two areas.

Table 4:2 Employment classifications

GROUP	SUB-GROUP	Planned area [Ilala]		Unplanned area [Relini]	
		Total	%	Total	%
Working Group	Income earners	2668	34.0	1961	31.6
	Non-income earners	1266	16.2	1083	17.5
Non-working Group	**Unemployed**	**570**	**7.3**	**447**	**7.2**
	Not able to work	1601	20.4	1185	19.1
	Students	1732	22.1	1528	24
Total		**7837**	**100**	**6204**	**100**

It is revealed that the employment status of the people between the two areas is very similar as it can be seen from the table. Indeed, very surprisingly, in both areas the groups of students and people who cannot work account for large number of percentages. This points out that both areas have large group of dependants, which is an indication of poverty. Map 4.7 and 4.8 show unemployment status of the two areas.

4.2.2.2 Relationship between unemployment rate and mortality rate
(i) Spatial analysis

Here the expectation is that: The higher the unemployment rate, the higher the mortality rate. The mortality rate maps, which were generated in section 4.2.1.2, are associated with the unemployment rate in the two areas (Maps 4.9 & 4.10). It is important again to stress that because of the problem of linking of the two files (mortality and census) it was not possible to calculate the number of people who died in each class of occupational status. Therefore the comparison made here is between the number of deaths caused by all diseases in each ten-cell unit, which includes all ages and all sex. Unemployment rate is based on the total number of unemployed people from the total working age population.

- **The planned area**

The map shows that, there are few numbers of units, which denote an indication of relationship between unemployment rate and mortality rate in the area. That is, there are units, which show a situation where unemployment rate is high; the mortality rate also tends to be high. However in many units it is noted that where unemployment rate is high, mortality rate is low. This means that unemployment rate is not the only factor, which influences mortality in the area. There are other factors, which need to be considered.

- **The unplanned area**

In the unplanned area the number of units where both unemployment rate and mortality rate are high becomes less than that found in the planned area. Most of the units indicate that where unemployment rate is high mortality rate is low. This situation again signifies little indication of relationship between unemployment rate and mortality rate in the unplanned area.

(ii) Statistical analysis

▪ The planned area

In order to get a better understanding of the relationship between mortality rate and unemployment rate in the area, a statistical test was also carried out. First a scatter diagram of the relationship was generated. The relationship is summarised by the line of best fit. The result shows a wider scatter about regression line. This result indicates that there is very weak relationship. The calculated r^2 value is 0.0211 (See figure 4.5). Strength of the relationship between the two variables was measured using a spearman's rank correlation coefficient (Table 4.3). The relationship between unemployment rate and mortality rate in the area is positive (0.372**) and is significant at 0.01 levels, indicating that the relationship is very weak.

▪ The unplanned area

In the unplanned area the statistical test, indicates that there is a negative weak relationship. The calculated r2 value is 0.0248 (See figures 4.6)

A spearman rank correlation coefficient was again used to measure the strength of the relationship between unemployment and mortality in the unplanned area. The correlation between unemployment and mortality is negative (-0.327*), which is statistically significant at 0.05 level indicating a negative very weak relationship (Table 4.4).
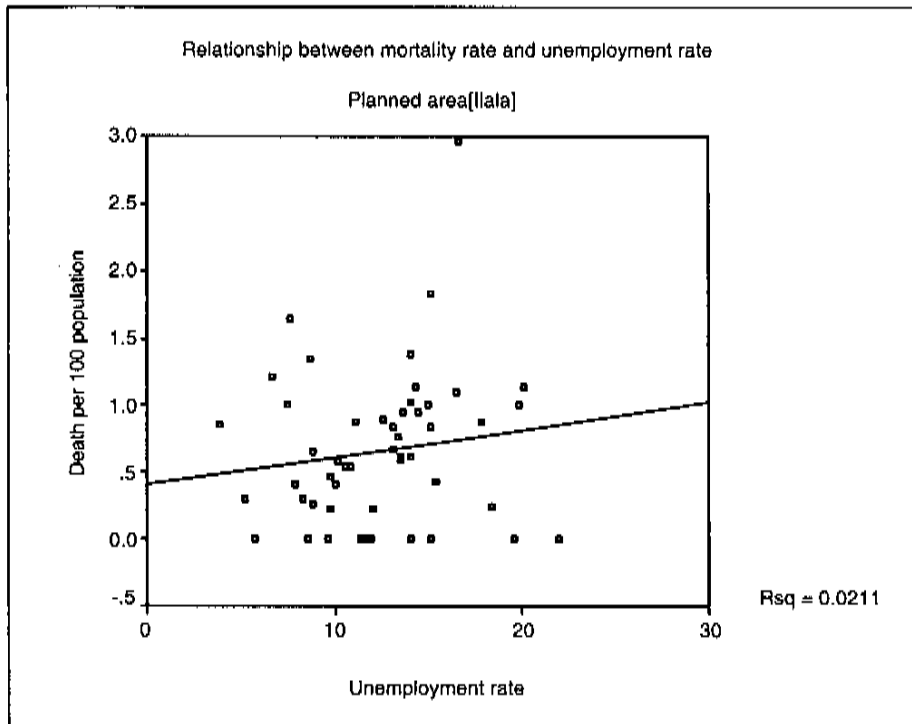
Figure 4:5 Relationship between unemployment rate and mortality rate in the planned area
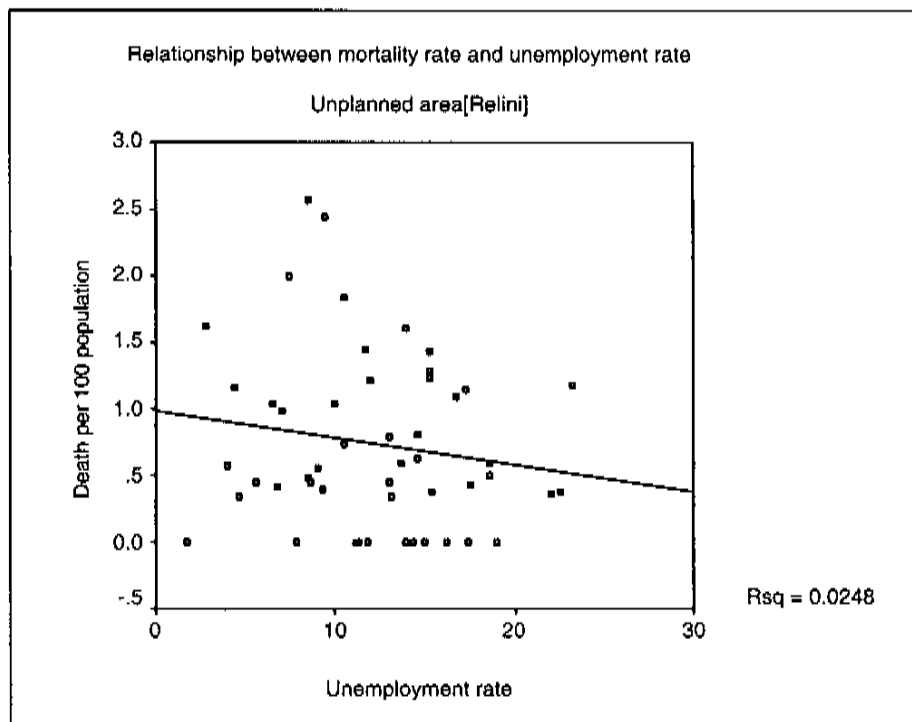
Figure 4:6 Relationship between unemployment rate and mortality rate in the unplanned area

Table 4:3 Spearman's rank correlation coefficient for unemployment rate and mortality rate from all diseases in the planned area

Correlations

			Mortality rate	Unemploym ent rate
Spearman's rho	Mortality rate	Correlation Coefficient	1.000	.372**
		Sig. (1-tailed)	.	.003
		N	52	52
	Unemploym ent rate	Correlation Coefficient	.372**	1.000
		Sig. (1-tailed)	.003	.
		N	52	52

**. Correlation is significant at the .01 level (1-tailed).

Table 4:4 Spearman's rank correlation coefficient for unemployment rate and mortality rate from all diseases in the unplanned area

Correlations

			Mortality rate	Unemploym ent rate
Spearman's rho	Mortality rate	Correlation Coefficient	1.000	-.327*
		Sig. (1-tailed)	.	.010
		N	50	50
	Unemploym ent rate	Correlation Coefficient	-.327*	1.000
		Sig. (1-tailed)	.010	.
		N	50	50

*. Correlation is significant at the .05 level (1-tailed).

4.2.3 Overcrowding condition

4.2.3.1 Introduction

In section 4.2.2 the issue of unemployment rate in the two areas has been discussed and associated with mortality rate. In this section, overcrowding as one of the important determinant factors of health status will be analysed and examined in relation to mortality caused by communicable diseases. To start with a definition of overcrowding and some indicators, which are used to measure overcrowding, are discussed.

Clauson-Kaas et al.(1996) define crowding or overcrowding as too many people for available space and facilities. Today, overcrowding in human settlements, especially in the unplanned areas, is a critical issue in many developing countries. In Tanzania for example, a study carried out by Lugalla (n.d) on environmental crisis and health, in Vingunguti and Hanna Nassifu settlements, in Dar es Salaam city revealed that in some cases more than four people live, cook and sleep in one room. Implication of such overcrowding is obvious, high risk of transmission of diseases.

The definition of overcrowding seems to be simple, however the complexity lies in its concept and how to measure it. Clauson-Kaas et al.(1996) argue that there are two bases for the concept of overcrowding, the number of persons occupying a residential area e.g. people per hectare and the living space a person has within the home. They further argue that overcrowding could also be associated with social and cultural factors. That is, how people themselves perceive overcrowding. For instance a slum dweller who depends on many residents of his /her home for economic support, childcare, and house guiding may see high number of people as socio support and health protection.

Literature suggests number of way on how overcrowding can be measured. Grove et al., Ruback and Pandey (quoted in Clauson-Kaas et al., 1996) measure over crowding using number of persons per room and number of persons per household, respectively. Turkstra (1998) in his work points out that in Colombia, overcrowding is measured by number of people per bedroom, and the situation is considered critical when the number exceeds three people per bedroom. Though all these indicators provide a crude indication on the state of overcrowding yet how they are calculated and how their variables are defined could still be criticized especial when they have to be related with health. For instance calculation based on number of people per room may create doubts as room sizes may vary considerably.

The interest of this study is to measure overcrowding using two methods, (1) roof area per person (2) roof area density per hectare as indicators of overcrowding. These two methods are proposed mainly because in developing cities, particularly in informal settlements, data for other commonly used indicators such as number of person per room, number of persons per bed, or number of children under five per room is not easily available or it is expensive to collect. Thus the study intends to apply an alternative method of measuring overcrowding by using GIS and remote sensing techniques. The application or the use of roof area per person is not a new method; Dangol (1998) used roof area per person to estimate the number of people in informal settlements in Dar es Salaam city. He concluded that the method is useful and effective in estimating population where common census method is difficult to apply. Roof area per person can also be related to a common indicator of floor space per person, which is commonly used to measure the pressure on habitable space in homes.

(i) Roof area per person

In this study context roof area is defined as a total area of a building as seen from an aerial photograph. Roof area per person therefore is the ratio between total roof area of the building and the number of people living in that particular building (Dangol, 1998). The formula can be summarized as;

$$\textit{Roof area per person = Total building area /Total number of people}$$

Since the aim is to measure overcrowding, in the two areas the question then is how many square meters per person could be considered as a sufficient space for human health? How many meter squares per person could be considered as a threat to human health? What is the standard unit? What are the criteria for defining the unit?

To answer these questions a number of factors were considered in both areas. In the first place, no standard unit for measuring overcrowding using roof area per person was found that could be used as a guide to justify that a certain size of roof area for a person in homes may have an impact on health. Therefore the standard unit was estimated by considering various previous studies done on housing as discussed below.

Looking on the aspect of size of buildings especially in informal settlements in Dar es Salaam city, it was found that, various studies conducted on housing in informal settlements (Sliuzas, 1998; Kyessi, 1990; Dangol, 1998) classify type of houses in unplanned area mainly in three classes. One, large Swahili type buildings, which consist of 6 bedrooms and comprise of an average size area of ± 8 x 12m and an occupancy rate of 2.4 people per room. Two, medium size buildings with an average size of ± 8 x 8m,with about 4 bedrooms, and an occupancy rate of 2 people per room. The third class is that of small buildings with an average size area of about ± 4 x 3,with 2 bedrooms and an approximately occupancy rate of ± 1.9 people per room.

In the planned area most of the sizes of the buildings are considerably bigger compared to that found in unplanned area thus in order to evaluate this difference the average roof area of the buildings in both areas was computed and compared. However in planned area special attention was given to multi-storey buildings, in which in order to calculate their areas, the number of floors was considered. In unplanned area, based on the results obtained from the updated part of the area (Chapter 3), it was revealed that the rate of densification is high and therefore many new buildings have been constructed, others being demolished or collapsed. This created doubts of whether it is reasonable to calculate the roof area per person using the population of year 2000 and the roof area of 1992. Definitely the significant changes, which have been observed in the area, could affect the result; therefore in this area the calculation is based only on Block D, that part of the area, which is updated.

It is found that in the planned area the average roof area per person is 12.4 square meters. While in the unplanned area the average roof area per person is 10.0 square meter (Figures 4.7 & 4.8). This means the average space in living homes in planned area is considerably bigger than in the unplanned area. The average roof area per building in both areas was also computed, and the result shows that in planned area the average roof area per building is 143.3 square meter while the average roof area per

building in the unplanned area is 59.1 square meters (Figures 4.7& 4.8). This result indicates that most of the buildings in planned area are twice as bigger as that of unplanned area.
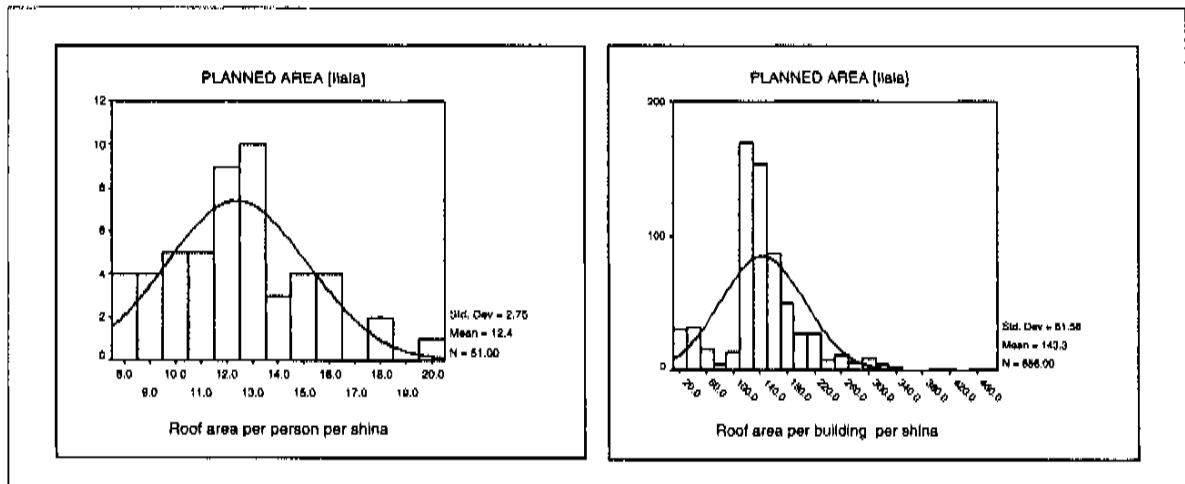
Figure 4:7 Histograms of average roof area per person and roof area per building in the planned area
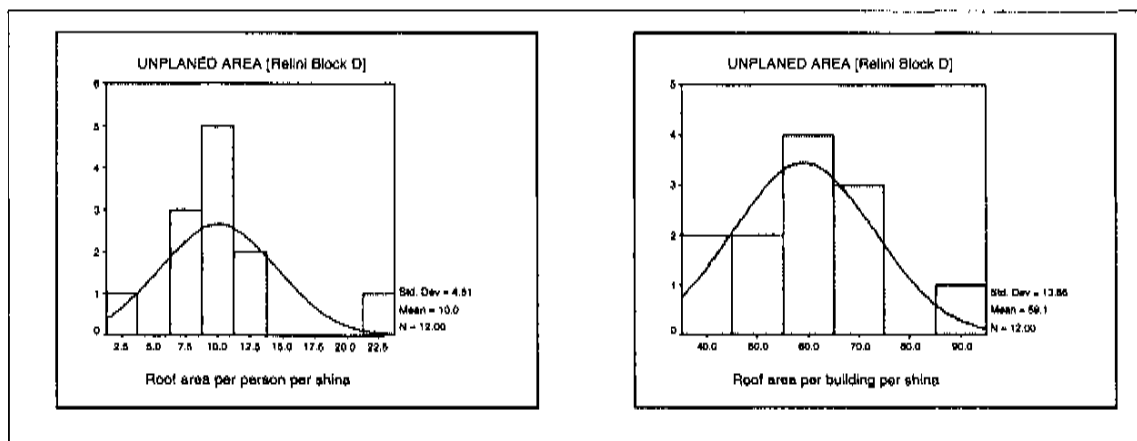
Figure 4:8 Histograms of average roof area per person and roof area per building in the unplanned area

This result was then compared with the result obtained by Dangol (1998), in his work conducted in three different informal settlements in Dar es Salaam city (Table 4.5). From the table it can be concluded that there is no significant difference of roof area per person in informal settlements. That means the type and the size of houses and the occupancy rates in informal settlements are almost similar. However there is a slight difference of roof area per person and roof area per building between the unplanned and the planned settlements as can be seen from the table. It should be noted that the planned area includes also multi-storey buildings, thus comparison of roof area per building may not give clear picture of the differences in average size of the buildings.

Table 4:5 Roof area per person & Roof area per building

	Area	Average Roof area/person (m^2)	Average Roof area/building (m^2)
1	Ilala (planned	12.4	143.3
2	Relini (unplanned)	10.0	59.1
3	Tabata	9.3	65.1
4	Hanna Nassifu	8.2	100.9
5	Manzese	8.5	122.2

In Jakarta, Indonesia various studies conducted on overcrowding by measuring habitable area per person indicate an average floor space per person of between 10.1 and 13 square meters (Clauson-Kaas *et al*, 1996). This result is comparable to the result obtained from the study in Jakarta, though the indicators used are different but the definition of the space to measure is the same.

Based on the average roof area per person and the average roof area per building calculated in the two areas and those from other studies, and the average size of houses and occupancy rate in the informal settlements; the average roof area of 8.0 m^2 per person was estimated and set to be used a standard for measuring overcrowding in both areas. Four classes were formulated, a roof area of less than 8.0m^2 per person in a living home in the two areas was considered as bad, and a roof area of between 8-13m^2 is classified as moderate, while a roof area of 14-16m^2 is considered as good and roof area greater than 16m^2 as very good.

The use of roof area per person as a measure of over crowding condition cannot be used without assumptions. The indicator was used under three fundamental assumptions. (1) All buildings considered in the calculations are for residential purposes (2) With the exceptional of multi-storey buildings in the planned area all buildings in the study area do not contain kitchen and toilets inside the building, that is cooking is done in a separate building and the household use pit latrine. (3) All rooms in the buildings are occupied and therefore no vacant room is available, and no room is used for any other purposes such as storage, and all rooms have windows for ventilation.

Map 4.11 illustrates the situation of over crowding in the planned area. The result shows that most part of the area has moderate level of overcrowding of roof area between 8-13 square meters per person. However there are units with overcrowding problem in the area located in the upper and lower parts. These parts are characterized by a roof area of less than 8 square meters per person which is, in this study context, is considered as a problem. The population living in these parts of the area is considered to be vulnerable to high risk of communicable diseases. In section 4.2.3.2 the relationship between overcrowding and mortality rate from communicable disease will be examined.

4.2.3.2 Relationship between roof area per person and mortality rate from communicable diseases

- **Spatial analysis**

By comparing the analysis results of overcrowding and mortality rate from communicable diseases (Map 4.12) in the planned area[5], it can be said that there is a little indication of an association between overcrowding and mortality rate from communicable diseases in the area.

- **Statistical analysis**

A statistical test was also carried where a scatter plot was used to examine the characteristic of the relationship. As it can be observed from the graph (Figure 4.8) there is a wider scatter about the regression line. The result shows that mortality rate decreases as roof area per person increases indicating a negative relationship. Spearman's rank correlation coefficient was also used to measure the strength of the relationship between overcrowding and mortality from communicable disease in the area. The result shows that there is a weak negative relationship (-0.120). Table 4.6 illustrates the results. It is difficult to draw a firm conclusion about this relationship because it is not clear whether the slope truly represents the association between overcrowding and mortality rate or whether it is biased by the effect of the data quality.
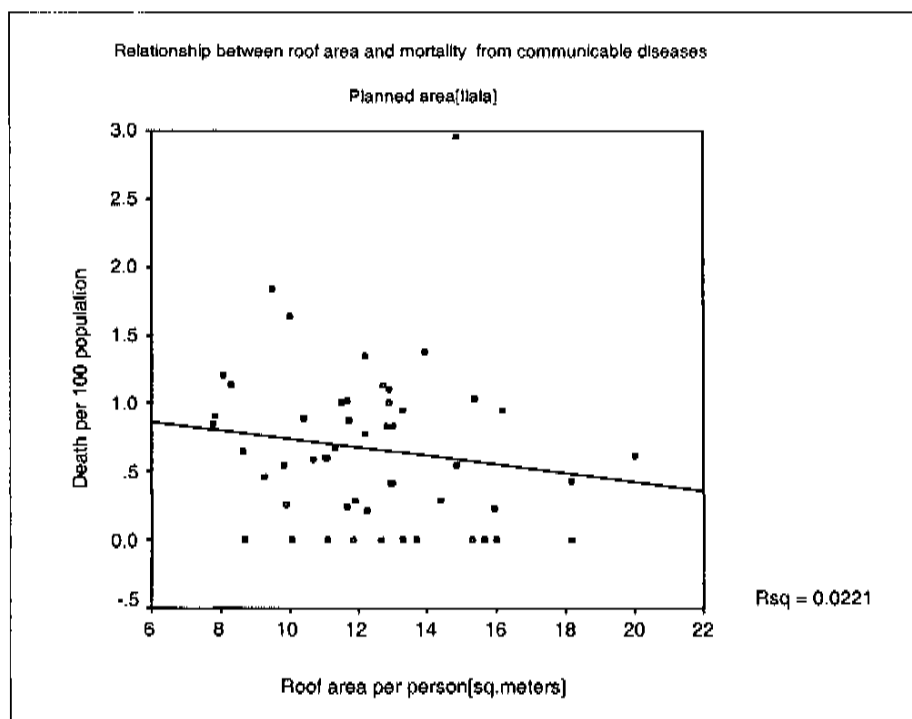
Figure 4:9 Relationship between roof area per person and mortality caused by communicable diseases in the planned area

---

[5] Since the map for unplanned area covers only (Block D), no comparison is made between overcrowding based on roof area per person and mortality rate from communicable diseases in the area

Table 4:6 Spearman's rank correlation coefficient for roof area per person and mortality rate from communicable diseases in the planned area

Correlations

			Mortality rate	roof area/person
Spearman's rho	Mortality rate	Correlation Coefficient	1.000	-.120
		Sig. (1-tailed)	.	.199
		N	52	52
	Roof area/person	Correlation Coefficient	-.120	1.000
		Sig. (1-tailed)	.199	.
		N	52	52

(ii) Roof density per hectare

Clauson-Kaas *et al* (1996) classified overcrowding indicators into four levels (1) Area level (2) Building level (3) Housing unit level and (4) Household level. Thus the roof area per person is an indicator of overcrowding based at building level. Under this sub section an overcrowding indicator based on density of building is discussed. Here the aim is to measure the building density in the two areas, using roof area density per hectare. Again in section 4.2.3.3 the result will be associated with the mortality rate caused by communicable diseases to see whether there is any relationship.

The roof area density per hectare was calculated using ArcView GIS software, where by a grid cell size of 2 meters and a search radius of 56 meters were used. Two types method for calculating density in ArcView GIS are available, simple and kernel methods. In this study context a simple method was used. However it should be noted that there is no much difference between the two methods in how the density is calculated. With simple method, density is calculated for each cell by summing the value found in an input map for each point found in the search radius and dividing by the area of the circle in area units. While with kernel method, density is calculated in the same way, as with the Simple method except the value found in an input map is distributed out from each point.

In the planned area (Map 4.13), high roof area density could be observed in block A and some parts of block B and C. In the unplanned area, the result points out that block C, which is the central part of the area have high percentage ratio of roof area density per hectare (Map 4.14). However the rate decreases as one moves closer to the river, that means this part of the area is characterized by few scattered buildings. In comparison, block A and D seem to be of low building density in the area and hence the area is expected to be of low risk of mortality from communicable diseases.

4.3 Summary

This chapter has described an attempt to analyse and explore the geo-referenced data. Though the analysis is not exhaustive, yet the data shows some valuable information. There is an indication of room for improvement and therefore room for better results.

With regard to the expectations, below is the summary of the main findings.

Expectation 1.*There is higher mortality rate in unplanned area than in planned area.*

Generally, over the period 1994-2000 there was a higher mortality rate in the unplanned area than in the planned area. However, in the past seven years the mortality rate in the planned area has been increasing while in the unplanned area it has been decreasing, to an extent that, in this study, the rate in the planned area was higher. This may be, in part attributable to the relatively small population and mortalities in each area, that contributes to small changes in the number of deaths having relatively large consequences in mortality rates. It is expected that the 2000 situation is not typical but it would require additional information to confirm if these are trends or more annomolies.

Expectation 2.*There is geographical pattern of cases of death in the two areas.*

Little indication of spatial pattern distribution of cases of death has been observed. This can be attributed due to small number of cases observed with the available data for a two year period. It is possible that if more years would be included, that more evidence of spatial patterns may be found. Further, the inability to analyse the data at house level because of privacy considerations and other constraints related to the availablity of spatial data does not allow the variation in socio-economic and environmental conditions that may exist within 10-cell units to be included in the analysis.

Expectation 3.*The higher the unemployment rates the higher the mortality rate*

Although the unemployment rate of the 2 areas is similar, there can be considerable differences in the income levels within the two areas. The statistical analysis at 10-cell level, has shown that there is a weak positive relationship between unemployment and mortality rates in the planned area and a weak negative relationship in the unplanned area. However, it must be borne in mind that in both cases there are relatively small numbers of deaths in each 10-cell unit and a considerable number in which no deaths occurred. Here too small changes in the number of deaths in a few 10-cell areas could have substantial effects on the results of statistical analysis.

Expectation 4.*The higher the overcrowding the higher the mortality rate*

The results show that there is no relationship between overcrowding and mortality rate from communicable diseases in the planned area. There were however several problems related to the measurement of overcrowding, such as the use of out dated base map of 1992 for calculating roof area in planned area.

Socio economic and environmental factors like unemployment and overcrowding condition if analysed spatially could also provide a better understanding of where the vulnerable groups to high risk of mortality are located. However in order to understand the effect of these factors on health status their definition need to be free from ambiguity and their method on how to measure them need to be

clear. Clauson-Kaas *et al* (1996) caution that there is little understanding of the impact and extent to which physical and environmental factors such as overcrowding, affect health, since most forms of diseases and illness are caused or contributed to by multiple factors. Causes of mortality are therefore diverse and complex. Currently the AMMP is involved in defining a set of key variable for indentifying poverty, and if such data could be used at house or household level for targetting concentrations of impoversihed households it may be possible more useful spatial outputs.

Although health related maps can be useful tools on conveying complex health situation in a simple way, their interpretation may sometimes be difficult and may lead to misinterpretation. In general visualization is one of the most important stages in investigating spatial patterns of health data. However, Westlake, (quoted in Gatrell, 1998), cautions that visualization could be the most dangerous part of the whole process because scale and colour are not neutral and are easy to manipulate. Interpretation and understanding of maps is both perceptual and cognitive process. Nevertheless, there are many issues of both cartographic and statistical nature, which pose difficulties in health mapping. Problems such as choice of number of classes, class interval and colour schemer or colour shading all these affect how the map can be read or perceived (Gatrell, 2002).

5. Conclusions and Recommendations

5.1 Introduction

The main goal of this research was to develop a methodology for geo-referencing the health related data in the context of the AMMP data collection methods. Developing such a methodology required systematic approach, clear description of what the methodology intends to achieve and how it could be operationalised not only by AMMP project but also other projects in the country, which collect (health) data at household level and intend to incorporate GIS. In order to achieve this goal, various geo-referencing methods were examined and critically evaluated in the context of Tanzanian situation. Understanding only the existing methods, how they work, their strengths and weaknesses would not be enough to achieve this goal. Therefore the problem of lack of up-to-date data, particularly in fast growing informal settlements is also highlighted. In line with the above-mentioned factors the methodology has been developed and tested in two surveillance sites and evaluated.

This chapter will attempt to draw some conclusions based on the validity and applicability of the proposed methodology for geo-referencing health data. It will also depict some specific recommendations on how the methodology could further be improved in order to provide better results.

5.2 Conclusions

Basically, the methodology is developed based on two different survey approaches, which were implemented in two areas. The conclusions drawn from each approach have been illustrated below in two main categories, which reflect the two main parts of the research, i.e. methodological part and the data analysis and exploration part.

5.2.1 Methodological aspects

(i) There are various methods, which are currently used to geo-reference health related data, at a very fine resolution from different sources such as statistics, registries of births and deaths, hospital records etc. These methods include, the use of small administrative units, grid squares/hexagonal method, post code systems, street address systems, parcel based systems and ground survey method, to mention a few. It is revealed that although the application of these systems started even before the use of computer or GIS technology, most of these methods rely on computer or /and GIS based systems and are mostly used in developed countries where computer or /and GIS technology is more advanced. Hence replicability of such methods in developing countries like Tanzania where computer technology, let alone GIS is still very low is extremely difficult.

(ii) The research has demonstrated that direct on ground survey method is an effective method in geo-referencing health data at a household level especially in countries where computer or/and

GIS technology is still low, or more specifically in countries where postcode, street address and census systems do not exist. It appears that in countries like Tanzania, the direct on ground survey as a method for geo-referencing health data will remain valuable for the foreseeable future. That is to say, projects, which are currently involved in collecting (health) data at a household level in the country and intend to incorporate GIS in their work, will likely continue to use the method for some years to come.

(iii) Direct on ground survey method is considerably cost effective method, in terms of human and physical resources. The use of small format aerial photographs for instance, which is an important tool for on ground survey method, is comparatively low on cost. However the on ground method if compared with other indirect method such as postcode systems could be expensive, because it is time consuming. Nonetheless, comparatively post code or street address systems are more expensive as they require national investment capital.

5.2.1.1 Ground survey method in the planned area

(i) Training of enumerators as one of the main components in AMMP data collection method, proves to be a very important stage in order for the methodology to be effectively implemented, as it provides the enumerators with full knowledge on how the methodology should be implemented. This reduces the chances of enumerators to generate errors during the survey process. Indeed it creates a sense of involvement to enumerators, not only in the survey process but also in the whole project.

(ii) Conducting a pilot survey seems to be a useful approach in order to assess the validity and applicability of the methodology prior to its implementation. It helps to understand the points of weaknesses of the proposed methodology, and hence improvements could be made.

(iii) Although it was thought that in planned areas densification is insignificant, the research revealed that there are still spatial development changes taking place in the area, which need to be examined properly before geo-referencing process takes place. Therefore the issue of updating should not only be considered in informal settlements but also in planned areas.

(iv) The method demonstrated by this research divides arbitrarily the area of study into small working blocks. It is found that small blocks are very useful to use in the process of geo-referencing because the area becomes easy to manage. However dividing the area arbitrarily may cause blocks to overlap with the ten cell unit boundaries, which creates difficulties to enumerators while locating households within the blocks (Chapter 3). Hence partition of the areas should be done based on the ten-cell units boundaries. This problem is not trivial as it may create confusion to enumerators during survey process and as a result, chances of generating error become high.

In general, concerning the approach used in the planned area it has been found that most of the weaknesses which have been experienced during the survey process could be avoided if more time were

dedicated in organizing and preparing all the necessary requirements such as updating of the base map and preparation of the photographs. Thus this approach could effectively work if more time will be invested in preparation before the survey takes place.

5.2.1.2 Ground survey method in the unplanned area

(i) The research demonstrate that small format aerial photograph is a useful tool in developing a methodology for geo referencing health data, especially in informal settlements where up-to-date-data is often not available and yet the rate at which spatial development (densification) takes place is high. The usefulness of small format aerial photography manifests itself in data capture and data updating process (Section 3.1). The use of small format aerial photography mosaic for instance, helps to understand complex spatial development of unplanned settlements as it provides both spatial and temporal information, which is useful information in developing a geo-referencing methodology.

(ii) The research demonstrates two different approaches while geo-referencing the data. The main difference of the two approaches is that, in planned area the data is geo-referenced at a house-hold level while in unplanned area the data is geo-referenced at a much higher level, a ten-cell unit level. The study proves that geo-referencing health data at household level is more useful than ten cell unit level because it allows more and detailed explorations of the data and hence more understanding of mortality conditions. Therefore this could be the most appropriate approach to focus in this type of work. However the level of confidentiality and privacy might limit the approach.

5.2.2 Usefulness of the geo-referenced data

The main objective of the data analysis and exploration is to assess the usefulness of the geo-referenced data. After carrying out the analysis the following conclusions have been drawn.

(i) A thematic map, as a method to visualize geographical patterns of mortality, seems to be useful. The main advantage of thematic map is that it clearly depicts the areas with high and low mortality rates by displaying the values using symbols with varying sizes, from small size to large size. The map-reader will automatically associate the large size symbol with high mortality rate and the small size symbol with low mortality rate. Hence thematic maps could be easy to understand by health planners or people who are not familiar with maps.

(ii) For this work mortality conditions of the two settlements was mapped and analysed with unemployment and overcrowding conditions. Although in this research it has not been possible to make use of all available data on health for the 2 case areas some examples of how such data can be use and represented are given. The concentration on mortality data, with a very small number of cases is a major constraint for this approach.

(iii) Roof area per person can be used as an indicator for measuring overcrowding, especially in informal settlements. However it has been noted that, it might be useful to

measure overcrowding using different indicators such as floor space per person so that the results could be compared and evaluated.

(iv) Though the research is proposing geo-referencing health data at a more fine resolution, yet such resolution poses a number of limitations particularly in terms of analysis. As Gatrell (1998) has pointed out, analysis of geo-referenced data at a very high resolution such as household creates challenges on ensuring individual confidentiality. The use of aggregated data avoids the problems of confidentiality but at the price of a loss of detail and possible problems related to ecological falacy.

(v) The research revealed that at smaller unit level such as household level the number of mortality events or frequency of events may become very low, and therefore it may become difficult to identify spatial variations of mortality rates with a high degree of confidence.

5.2 Recommendations

(i) There is a clear need of establishing a national standard address system which could be used not only for geo-referencing health data but also for other applications in the country at large. Since there is no any address system, which currently exists the recommended system, may cause no conflict with any other systems. Abbott (2000) in his work suggests a similar system to be established in informal settlement. He proposes that the community in informal settlement could create a numbering system and physically allocate the numbers on the shacks. Such a system in informal settlement may provide basis for community involvement towards establishing a GIS based system.

(ii) The analysis has only illustrated variation in mortality experience in the area based on two year period, 1999 and 2000. The two-year mortality data is not sufficient to examine variation. Therefore there was insufficient data to effectively evaluate the usefulness of geo-referenced data. It is recommended that in order to examine the usefulness of geo-referenced data more effectively in relation to morbidity problems where relationships to spatial and environmental factors may be more pronounced.

Bibliography

Abott, J., (2000). *The Use of Spatial data to Support the Integration of Informal Settlements into Formal City*, Department of Engineering, university of Cape Town.

Alpert, S. and Haynes, K., (1994). *Privacy and the Intersection of Geographical Information and Intelligent Transportation Systems*, Proceedings of the Conference on Law and Information Policy for Spatial Database, pp. 198-211. Temple, Arizona,October 28-29, 1994.

Amer, S. and Thorborg, H., (1996). Health Coverage in Urban Dar es Salaam: users interaction patterns, under-serviced areas. ITC, Enschede, The Netherlands.

Binala, J., (2000). Monitoring a low standard settlement in Dar Es salaam, A study Of Hanna Nasifu, IFA, ITC, Enschede, The Netherlands.

Bruijin, C. A., (1987). 'Monitoring a large squatter area in Dar es Salaam with sequential aerial photographs', *ITC journal*, Enschede, The Netherlands.

Bradley, D., Stephen, C., Harphan, T. and Caircross, S. (1992). *A review of Environmental Health Impact in Developing Country Cities*, Urban Management Programme Discussion Paper, World Bank.

Bryman, A. and Cramer., D., (1997). *Quantitative Data Analysis with SPSS for Windows: A Guide for Social Scientists*, Methuen, London, UK.

Chambers, K., (2001). Planning and Monitoring Informal Settlements Using Small Format Aerial Photographs: implications for parish council in Jamaica. IFA report, ITC, Enschede, The Netherlands.

Clauson-Kaas, J., Dzikus, A., Stephens, C., HØjlying, N. and Aaby, P.,(1996). 'Urban Health: Human Settlement Indicators of Crowding', *TWPR*, 18(3): 349-363.

Cowen, D., (1997). 'Discrete Geo-referencing: NCGIA Core Curriculum in GIS Science', University of South Carolina, Department of Geography, Columbia, USA, http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u29.html (Last visited in December, 2001).

Croner, C., Sperling J. and Broome, F., (1996). 'Geographic Information System (GIS): New Perspectives in Understanding Human Health and Environmental Relationships.' *Statistics in Medicine*, 15:1961-1977.

de By, R.A.(ed), (2000). *Principles of Geographical Information Systems, ITC educational textbook series*, ITC, Enschede, The Netherlands.

Gatrell, A. and Löytönen, M. (eds), (1998). *GIS and Health*, Taylor and Francis, London, UK.

Gatrell, A., (2002). *Geographies of Health: An introduction*, Blackwell, UK.

Graham, R. and Real, R.E., (1986). *Manual of aerial photography*, Butter worth &co. Ltd, London, United Kingdom.

Hoftsee, P., (1985). 'An introduction to small aerial photography for human settlement analysis', *Communications* (2) 121-126, ITC, Enschede, The Netherlands.

Hofstee. P., (1990). *Low-Cost Aerial Photography and Mapping for Urban Projects*: ISPRS Commission Symposium, 17-21 September, Victoria, Canada.

Kingham, S. P., Gatrell, A. C. and Rowlingson, B., (1995). 'Testing for clustering of health events within a geographical information system framework', *Environmental and planning A*, 27:809-821.

Kyessi, A., (1990). The Urbanization of Fringe Villages and Growth of Squatters, The case of Dar es Salaam, Tanzania, Unpublished Msc. Thesis, ITC, The Netherlands.

Liao, H., Lymon, P. and Shull, K., (1999). 'Vital Health and Census data Integration System in South Carolina Department of Health and Environmental Control (SCDHEC)', Columbia, US, http://www.esri.com/news/arcuser/0499/vital.html (Last visited in October, 2001).

Lugalla, J. L.P.,(n.d).Paper prepared for publication in a book on Social Change and Health in East Africa.

Macintyre, S., (1998). Social Inequalities and Health in the Contemporary world: Comparative Overview in Strickland, S. S. and Shetty, P.S. (eds). Human Biology and Social Inequalities, Cambridge University press, Cambridge.

Martin, D. and Higgs, G., (1993). 'Population geo referencing in England and Wales: basic spatial units reconsidered', *Environment and Planning A*, 1997. 29(2): 333-347.

Martikainen, P., (199 0). Does employment cause mortality?, Department of sociology, university of Helsinki, http://www.stat.fi/isi99/proceedings/arkisto/varasto/mart047, (Last visited in December,2001)

Marx, R.W., (1990). 'Cartography and Geographic Information Systems: The Census Bureau's TIGER System', *Journal of American Congress on Survey and Mapping*.17 (1): 7-120.

Ministry of Health and AMMP Team, (1997). The Policy Implications of Adult Morbidity and Mortality, End of Phase 1 Report, Dar es Salaam, United Republic of Tanzania.

Mohan, J. and Rhind, J., (1983). *Linking and Analysing Census and Medical Data: mortality experience in county Durham, 1969-1977*. University of Durham, UK.

Parson, T. and Knight, P., (1996). How to Do Your Dissertation in Geography and Related Disciplines, Chapman &Hall, London, UK.

Raper, J., Rhind, D. and Shepherd, J., (1992). *Postcodes: the new geography*, London, UK.

Rideout, T.W. (ed.), (1992). *Geographical Information Systems and Urban and Rural Planning*, The institute of British Geographers, UK.

Rory, R, Anotinette L. and Mtani A., (2000). Results of a City Victim Survey, Dar es salaam, Tanzania, http://www.iss.co.za/Pubs/Other/DaresSalaam/Tables.htmll (Last visited in September, 2001).

Savigny, D. and Wijeyaratne, P. (eds.), (1994). *GIS for Health and the Environment*, International Development Research Center, Ottawa, Canada.

Sawio, C.J., (1998). Managing Urban Agriculture in Dar es Salaam: Cities Feeding People Programme Initiative, International Development Research Centre, Canada.

Sliuzas, R.V., (1988). Problem in Monitoring the Growth of Squatter Settlements. The housing process in Manzese, Dar es Salaam, Tanzania, Unpublished Msc thesis, ITC, The Netherlands.

Turkstra, J., (1998). Urban Development and Geographical Information: spatial and temporal patterns of urban development and land values using integrated geo-data, Villavicencio, Colombia, ITC Publication No. 60, Enschede, The Netherlands.

University of Newcastle upon Tyne, The Adult Morbidity and Mortality Project, http://www.ncl.ac.uk/ammp/(Last visited in January, 2002).

Verhasselt, R., (1993). 'Geography of health: Some trend and perspectives'. *Social science medicine* 36(2) 119-123.

Walsh, J. and Simonet, M., (1995). 'Data analysis needs for health sector reform', *Health Policy*, 32 (1-3): 295-306.

Walter, D.S., (1993). 'Visual and statistical assessment of spatial clustering in mapped data', *Statistics in Medicine*, 12:1275-1291.

Warner, S., Graham, R. and Read, R., (1996). *Small format aerial photography*, Whittles, Caithness, Scotland.

Westlake, A. (1995). Strategies For the Use of Geography in Epidemiological Analysis.Page135-144 in de Lepper, M. J. C., Scholten, H.J. and Stern R.M. (eds), *The added Value of Geographical Information System in Public and Environmental Health*, Kluwer Academic Publishers, Donrdrect, 135-144.

Appendices

Appendix 1. A sample of census form

Appendix 2.A sample of Verbal Autopsy form

20.6 verbal autopsy form

| AREA | | SERNO | | IDNUM | | Date of interview | |

Village/Branch   Balozi   Head of Household

Respondent: Relation to deceased   Lived with deceased?   YES / NO

       Education   Occupation

Deceased: Name   Relation to head of household

       Age   Sex   Marital Status   Education

       Smoking   Drinking   Weeks gestation   Weeks postpartum

       Occupation   Place of death   Date of death

       Cause of death according to respondent

History of patient's illness including duration of illness

Previously *diagnosed* medical conditions

Illness	Duration (months)	Illness	Duration (months)	Illness	Duration (months)
1. Hypertension		5. Heart Disease		Leprosy	
2. Epilepsy		6. Cancer of		Other Chronic illnesses	
3. Asthma		7. Sickle cell disease			
4. Diabetes		8. Tuberculosis			

Symptoms	Duration (weeks/days)	Symptom	Duration (weeks/days)
1. Fever		22. Vaginal discharge	
2. Breathless on light exercise or work		23. Vaginal bleeding	
3. Breathless on lying flat		24. Haematuria	
4. Indrawing of intercostal spaces		25. Urethral discharge	
5. Palpitations		26. Dysuria	
6. Wheezing		27. Unable to pass urine	
7. Ankle/leg swelling		28. Polyuria	
8. Cough dry		29. Thirst	
9. Cough productive		30. Headache	
10. Haemoptysis		31. Neck pain	
11. Chest pain		32. Fits	
12. Poor appetite - unable to breast feed		33. Mental confusion	
13. Weight loss or failure to thrive		34. Loss of consciousness	
14. Mouth sores/painful swallowing		35. Paralysis	
15. Dysphagia		36. Failure to sit or stand	
16. Jaundice		37. Pins and needles in foot	
17. Vomiting		38. Skin disease	
18. Vomiting blood		39. Joint pains	
19. Abdominal pain		40. Joint swelling	
20. Diarrhoea		Others:	
21. Bloody diarrhoea or blood in stools			

Duration: 0 = no symptom or illness

99 = Unknown

APPENDIX 3. Digital Elevation Model [Unplanned area, Relini]

Digital Elevatiom Model:Relini
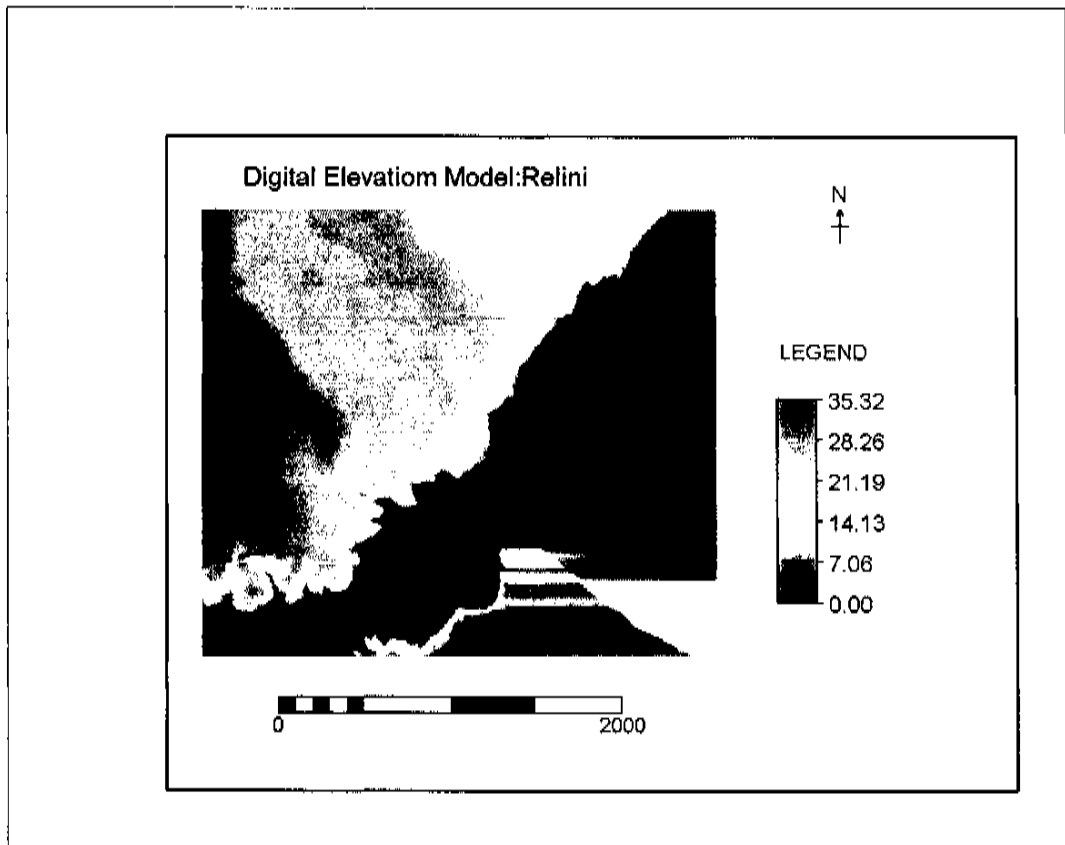
APPENDIX 4. Relini Ten-cell Boundary

Relini Ten Cell Boundary
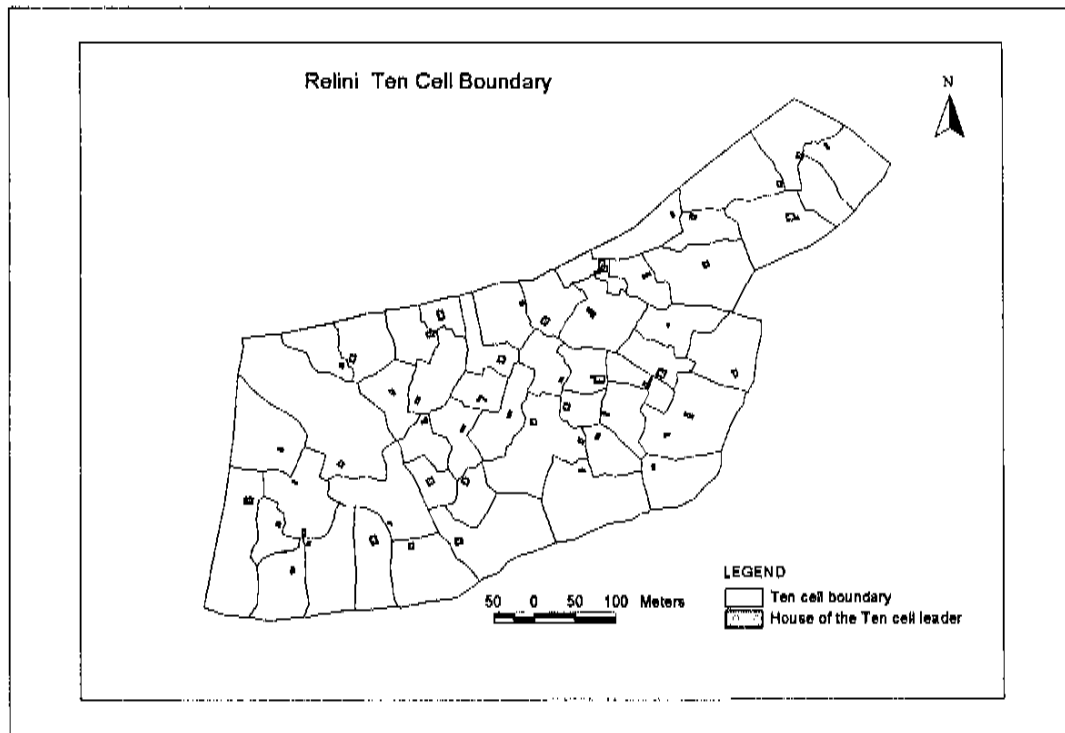
Appendix 4. Occupation code and classification

Working	110	Working and earning an income	111	Legislators & Officials, District Level and Above, Senior Army & Police Officers
			112	Professionals & Managers of Large Businesses
			113	Technicians, Small Business Managers, Teachers, Nurses
			114	Clerks, Service Workers, and Shop and Market Sales
			115	Farmers, Fishermen, Herders who sell products for cash
			116	Skilled fundis and Artists (not Apprentices or Casual Labour)
			117	Machinists, Drivers, Seamen, Assembly line workers, Army Regulars, Police
			118	Unskilled Manual Labour, Casual Labor, Apprentices, Petty Traders, Charcoal Makers
			119	Other
	120	Working, no income	122	Subsistence activities
			123	Unpaid household production
			124	Other
Not working	210	Not working but able to work	-	
	220	Not able to work or retired	221	Elder Age over 60 or Retired
			222	Ill/Disabled
			223	Child under 7 years
	230	Student		

Appendix 5. Classification of causes of death

COMMUNICABLE DISEASES

1.0.0 *Unspecified communicable diseases*

  Acute Febrile Illness

1.1.0 *Unspecified acute febrile illness*

  1.1.1 *Malaria*

  1.1.2 *Meningitis*

All other specified acute febrile illnesses

Acute Respiratory Infections

1.2.0 *Unspecified Acute respiratory infections*

1.2.1 *Pneumonia*

1.2.9 *All other specified acute respiratory infections*

  Hepatitis

1.3.0 *Hepatitis*

Tuberculosis/AIDS

1.4.0 *Unspecified TB/AIDS*

1.4.1 *Pulmonary Tuberculosis*

1.4.2 *AIDS*

1.4.3 *AIDS + Pulmonary Tuberculosis*

1.4.9 *All other forms of tuberculosis*

Diarrhoeal diseases

1.5.0 *Unspecified diarrhoeal diseases*

1.5.9 *All other Specified diarrhoeal diseases*

Tetanus

1.6.0 Tetanus

Rabies

1.7.0 *Rabies*

Other specified Communicable

1.9.0 all other specified communicable diseases

DIRECT MATERNAL CAUSES

2.0.0 *Unspecified direct maternal causes*

2.1.0 *Unspecified Abortion*

2.1.9 *Specified Abortion*

2.2.0 *Eclampsia*

2.3.0 *Unspecified Ante/postpartum haemorrhage*

2.3.9 *Specified Ante/postpartum haemorrhage*

2.4.0 *Obstructed labour*

2.5.0 *Puerperal sepsis*

2.9.0 *Other specified direct maternal causes*

NON-COMMUNICABLE DISEASES

3.0.0 Unspecified non-communicable causes

Cardiovascular Disorders

3.1.0 Unspecified cardiovascular disorders

3.1.1 Hypertension

3.1.2 Congestive Cardiac Failure

3.1.3 Ischaemic Heart Disease

3.1.4 Cerebrovascular Disease

3.1.9 All other specified cardiovascular disorders

Pulmonary diseases

3.2.0 Chronic Obstructive Pulmonary Disease

Liver diseases

3.3.0 Unspecified liver diseases

3.3.1 Liver cirrhosis

3.3.9 All other specified liver diseases

Acute abdominal conditions

3.4.0 Unspecified Acute abdominal conditions

3.4.9 All other specified acute abdominal conditions

Diabetes

3.5.0   Diabetes

Neoplasms

    3.6.0 Unspecified neoplasms

    3.6.1 Carcinoma breast

    3.6.2 Carcinoma cervix/uterus

    3.6.3 Hepatoma

    3.6.4 Carcinoma of gastrointestinal tract

    3.6.5 Carcinoma of the lung

    3.6.6 All other specified neoplasms

Renal Disorders

    3.7.8 Unspecified Renal disorders

    3.7.9 Specified Renal disorders

Central Nervous System disorders

    3.7.0   *Unspecified Central Nervous System disorders*

    3.8.1   *Epilepsy*

    3.8.9   *All other specified CNS disorders*

Other Non-communicable

    3.9.0   *All other specified non-communicable diseases*

Symptoms, signs, syndromes not elsewhere classified

    4.1.0   *Anaemia*

    4.9.0   *All other specified symptoms, signs and syndromes*

EXTERNAL CAUSES

    5.0.0   *Unspecified external causes*

    Unintentional Injuries

    5.1.0   *Unspecified unintentional injuries*

    5.1.1   *Road traffic accident*

    5.1.2   *Accidental poisoning*

    5.1.3   *All Other specified unintentional injuries*

Intentional Injuries

    5.2.1   *Homicidal injuries*

    5.2.2   *Suicidal injuries*

Undetermined / Uncertain cause of death

    6.0.0   *Undetermined*

Classification of causes of death (children)

Diarrhoeal diseases*

Injuries*

Undetermined*

Acute febrile illnesses*

Acute respiratory infections*

* As categorised above

10.   0.0   Still birth

11.   0.0   Birth injury and/or asphyxia

12.   0.0   Prematurity and/or low birth weight

13.   0.0   Congenital abnormalities

14.   0.0   All other perinatal causes

15.   0.0   Neonatal tetanus

16.   0.0

17.   0.0

18.   0.0   Malnutrition

19.   0.0   Measles

20.   0.0   All other specified diseases

APPENDIX 6. Questionnaire for data manger, supervisors and enumerators

QUESTIONS FOR DATA MANAGER

Name of the interviewee---
Address--
Date--
Occupation/position--
Name of interviewer--

SECTION A: BACKGROUND

Could you explain the administrative set up of the AMMP project? ---
----

What is your main task/role in AMMP project? --

SECTION B: CURRENT FIELD WORK METHODS

Could you explain step by step how do you carry out your fieldwork? --
----

In areas like Mtoni Relini there is evidence that densification is still taking place very rapidly. How important is it for your fieldwork to have (updated) maps and what steps do you take to ensure that they are available?

Before starting the fieldwork, how do you update your maps? If you don't use maps what other techniques do you use to identify the new buildings? --

In the fieldwork you might be facing some problems, could you explain the main problem(s) do you encounter during the field survey? -------
----

How do you ensure consistency in your fieldwork survey? --

SECTION B: PROPOSED METHOD

The main objective of this research is to develop a methodology for geo-referencing the AMMP health data. What criteria/specific requirements do you think should be considered in order to generate a feasible geo referencing methodology that may suit your needs? ----------------
----

As part of data collection method the researcher has prepared detailed maps of the two branches of Relini and Ilala Ilala, and divide the branches into small working blocks (A, B, C and D). Each building on the map has been assigned an ID number, the task of the enumerator will be to identify the building on the ground as well as on the map and record its information on the form. According to your experience do you think the enumerators can perform such a task? If not what do you think will be the main problem(s)? (The researcher will demonstrate how to use the maps) --
----

SECTION D: ANALYSIS

How do you think examining spatial elements of AMMP health data could be useful in order to achieve the objectives of the AMMP project?
How do you think the use of GIS and remote sensing may improve the data   collection, analysis and visualization capabilities of the AMMP project?
What is the problem to be solved?
What kinds of analysis are to be carried out?
What are the final products expected of the GIS?

SECTION D: GENERAL QUETSIONS

The National Institute of Medical Research (NIMR) is also conducting research of this nature, what is the difference between your work and that performed by the institute?
The AMMP project is under the ministry of health, what role does the Dar es Salaam city council play in this project?

QUESTIONS FOR SUPERVISORS

Name of the interviewee---
Address--
Occupation/position--
Date--
What is the highest level of education attained? (Please mark with X)

Primary school   ⌂
Secondary school   ⌂
College   ⌂
University   ⌂
Any other (specify)--

Please mention any other course you have attended? --

--

How many times have you been a member of fieldwork supervisor's team for the AMMP project?

Please explain briefly, what your main responsibility in the AMMP project is? --
----

In general, what are the main problems you encountered in the fieldwork during the October 2001 census? -----------------------
----

Questions related to the use of maps during fieldwork
During this census (October 2001) Enumerators have been using maps, which were prepared by ITC. Were these maps useful in carrying out the household interviews and, if so, in what ways? --
----

According to your experience do the maps have any deficiencies (e.g. the buildings are too small, the numbering is not clear, some important information is missing, etc)? Please explain. --
----

What improvements should be made to the maps in order to make them easier to use in the field? ---------------------------------
----

On the basis of your experience would you recommend that the AMMP adopt the use of maps as a standard practice? (Please explain your answer) --
Do you have any other additional remarks to make about the use of maps in this year census activities.

QUESTIONS FOR ENUMERATORS

Name of the interviewee---
Address---
Date--
Occupation/position---
Name of interviewer---

SECTION A: Background
What is the highest level of education attained?
Primary school
Secondary school
University
Any other (specify)---

What other courses attended? --

For how long have you been working as enumerator with AMMP project? -------------------------------
----
Do you use maps in your fieldwork survey? If yes? For what purpose? (get a copy of the maps if possible and find out if they are happy with the maps and if not how they think they should be improved)--
--- -
What is your ability to use maps?
Very good
Good
Poor
SECTION B: CURRENT FIELD WORK METHODS

Could you explain step by step how do you carry out your fieldwork? ---
----
In areas like Mtoni Relini there is evidence that densification is still taking place very rapidly, before starting the fieldwork, how do you update your maps? If any? If you don't use maps what other techniques do you use to identify the new buildings? ---------------------------
----

In the fieldwork you might be facing some problems, could you explain the main problem(s) do you encounter in the fieldwork?
How do you ensure that you visit all houses in your fieldwork survey? --
----

How many questionnaires do you complete per day? ---

Questions related to the use of maps during fieldwork

During this census (October 2001) you have been using maps, which were prepared by ITC. Were these maps useful in carrying out the household interviews and, if so, in what ways? --
----

According to your experience do the maps have any deficiencies (e.g. the buildings are too small, the numbering is not clear, some important information is missing, etc)? Please explain. --
----

What improvements should be made to the maps in order to make them easier to use in the field? --
----

On the basis of your experience would you recommend that the AMMP adopt the use of maps as a standard practice? (Please explain your answer) --

Do you have any other additional remarks to make about the use of maps in this year census activities