MEMOIRE DE RECHERCHE

LINKING MULTIPLE GEO-REFERENCED HEALTH DATASETS USING GEOGRAPHIC INFORMATION SYSTEMS (GIS) : METHODS & POTENTIAL APPLICATIONS

A case study for MALAWI

Data sources

National Statistical Office of Malawi
World Health Survey – World Health Organization
Health Facilities Inventory – Japanese International Co-operation Agency

Supervised By :
Dr. Steeve EBENER

Prepared By :
Imed BEN HAMADI

MARS 2006
LIST OF FIGURES

Figure 1: Possible use of GIS as a tool for the integration of data and analytical tools ...................... 6
Figure 2: Main steps of the process ..................................................................................................... 7
Figure 3: Malawi Administrative Districts (2002) ........................................................................... 8
Figure 4: Distribution of the clusters part of the WHO WHS .......................................................... 12
Figure 5: Sections 0100 (sampling information) and 0200 (geocoding information) of the WHS questionnaire ......................................................................................................................... 13
Figure 6: Description of cleaning process used for the geographic component in the context of the WHO WHS .................................................................................................................................................. 14
Figure 7: District level links .............................................................................................................. 18
Figure 8: Facilities/Households levels links ...................................................................................... 19
Figure 9: Distribution of selected (a) and quoted (b) health facilities. ................................................ 20
Figure 10: Distribution of selected facilities by type (a), ownership (b), road access (c) and frequency of use (d). .................................................................................................................................................. 21
Figure 11: Global representation of distances travelled to selected facilities. .................................... 23
Figure 12: Distribution of selected health facilities / Average distance travelled ......................... 25
Figure 13: Lilongwe Central Hospital catchment’s area. .................................................................. 26
Figure 14: St Gabriel catchment’s area. ........................................................................................... 26
Figure 15: Matiki Health Centre catchment’s area ......................................................................... 27
Figure 16: Distances travelled to Ndamera health centre per household expenditures ...................... 28
Figure 17: Catchment’s area for St. Johns Hospital a) and Mzuru Central hospital b) .................... 29
Figure 18: Salima District Hospital catchment’s area ....................................................................... 30
Figure 20: Salima District Hospital catchment’s area ..................................................................... 31
Figure 21: Lifuwu Health Centre catchment’s area ......................................................................... 31
Figure 22: Distribution of patients per type of health care ............................................................... 31
Figure 23: Patient global assessment of Ndamera health centre service. .......................................... 32
Figure 24: % surveyed households that have used inpatient health care services for Adults ........... 33
Figure 25: % surveyed households that have used inpatient health care services for Childs ........... 33
Figure 26: % surveyed households that have used ambulatory health care services for Adults ....... 33
Figure 27: % surveyed households that have used ambulatory health care services for Childs ...... 33

LIST OF TABLES

Table 1: Country and district level information/indicators .................................................................. 10
Table 2: Household level information/indicators ............................................................................. 11
Table 3: Health facility level information/indicators ....................................................................... 15
Table 4: Average household’s size / sub-groups ............................................................................. 34
Table 5: Household’s age classes / sub-groups .............................................................................. 34
Table 6: Household’s header education / sub-groups .................................................................... 35
Table 7: Household’s years of formal education / sub-groups ......................................................... 35
Table 8: Household’s expenditures / sub-groups .......................................................................... 35
Table 9: Household’s header main occupation / sub-groups .......................................................... 35
I am greatly indebted to:

- Dr. Steeve Ebener who supervised this research and whose assistance, suggestions and encouragement helped me to accomplish this work.
- Dr. Somnath Chatterji for having facilitated and encouraged this research by his interest and valuable hints as well as for having provided access to the WHO WHS data.

I would like to express my gratitude to the Department of Geography of the University of Geneva for efforts in providing a comprehensive and applicable programme in “Geographic Information Systems”.

My special thanks are due to Mr. Antonio Martin, who accepted to examine and discuss this report. I want to thank him for his help and interest.

Finally, I want to dedicate this research to my fiancée Hayat who has been a great source of motivation and whose patient love enabled me to complete this work.
I. CONTEXT

This research has been initiated in the continuity of an internship which took place at the World Health Organisation (WHO) between November 2004 and January 2005 within the Knowledge Mapping and GIS group of the Knowledge Management & Sharing department (KMS) in close collaboration with the multi-country studies group in the Measurement and Health Information Department (MHI).

The Knowledge Mapping and GIS group is working in the area of Knowledge Mapping, acquisition, management, analysis and modeling of the spatial and temporal dimensions of Health in the area of Health Systems performances and Health Inequalities.

The main projects focusing on the translation of data to geographic information and/or the translation of geographic information to knowledge are:

- The development of Knowledge Mapping methods and guidelines
- The Second Administrative Level Boundaries dataset project (SALB¹),
- Measuring physical accessibility to health care (AccessMod²),
- Health inequalities mapping (poverty mapping),
- The technical support to the WHO World Health Survey (WHS³).

The objectives of the internship were to work in the context of the support given to the WHO WHS by performing the following tasks:

- Generation of geographic variables:
  - Generation of geographic variables for the countries that have used GPS devices in the context of the WHO WHS,
  - Testing the protocol related to geographic variables generation method and providing appropriate improvements,
  - Integration of significant statistics describing the spatial distribution of surveyed households.
- Improvement of methods:
  - Bibliographic search regarding the integration of geography and Geographic Information Systems (GIS) in the context of health surveys;
  - Investigation of the potential uses of GIS for analyzing spatially referenced data and its applications.
  - Development of a tool to support survey sampling frame design

¹ http://www3.who.int/whosis/gis/salb/salb_home.htm
² http://www.who.int/kms/initiatives/accessmod/en/index.html
³ http://www.who.int/healthinfo/survey/en/
II. INTRODUCTION

In many developing countries, the main weaknesses of health systems remains:

- The lack of reliable data
- The inadequate appreciation and use of available information in planning and management of health systems.

In this regards, GIS might offer a working platform for generating, health planning and management information systems that could integrate both data and analytical tools (Figure 1).

The continuous expansion of the capacities of GIS and the increasing awareness of these tools are offering new opportunities for public health decision makers willing to enhance their planning, analysis and monitoring capabilities.

Among these capacities we can for example mention:

- Integration of socio-economic and spatially distributed information in a unique health information system;
- Setting priorities for the allocation of health care resources;
- Monitoring and surveillance of health status;
- Evaluation of health programmes & health care outcomes;
- Identification of environmental, socio-economic and other factors, which influence health, under serviced, poor, inaccessible areas and other geographic and demographic factors.

Among possible sources of data, we can mention households and health facilities surveys.
The role of surveys is expressed by the following major points:

- Providing updated field data;
- Structured data depending on sampling stratification variables;
- Peoples health state description and valuations;
- Health systems responsiveness;
- Results can be related to many significant factors (income, education, location, ..etc).

Also, the growing use of GPS devices in the context of households and health facility surveys are offering new source of information that can be analysed from a spatial perspective using GIS tools.

This research aims at exploring ways of using GIS for strengthening public health services planning, issue of concerns especially for countries facing crucial health challenges.

This study therefore focused on:

- The integration of the geographic dimension in the context of households and health facilities surveys,
- The process for linking several geographic data sets,
- The potential uses and application of GIS tools for analysing merged spatially referenced facility and health data.

The whole process of this study could be summarized as shown in Figure 2.
III. THE CASE STUDY AREA

In the context of this study, Malawi has been chosen for applying and developing the new methods.

Malawi is bordered with Mozambique to the South, south East and South West, Zambia to the West and Tanzania to the North and North East. Administratively, Malawi is divided into 3 regions: North, Central and South, with a total of 27 districts. It has a population of about 12,000,000 people unequally distributed between regions and 45% of these are children below 15 years. Almost 90% of the population lives in the rural areas with agriculture as the main occupation. Literacy in the country is estimated at 52%.

Landlocked Malawi ranks among the world’s least developed countries. Population growth, increasing pressure on agricultural lands, and HIV/AIDS pose major problems for the country.

Figure 3 : Malawi Administrative Districts (2002)
This country has firstly been selected because of the availability of several geo-referenced datasets:

- The WHO World Health survey (WHO WHS) performed in Malawi in 2002,
- The Demographic Health Survey (DHS) performed in Malawi in 2004,
- The health facilities inventory funded by the Japanese International Co-operation Agency (JICA) in 2002.

More information regarding these datasets are provided in the next section of this document.
IV. DATA

As we have seen earlier, good decisions and good planning require good information. Then, gathering essential information and indicators from available data sources at different levels is the key step of this process. For this project 3 types of indicators have been identified according to their geographical level:

- The country and district level (Table 1),
- The household level (Table 2),
- The health facility level (Table 3).

The source and contents for these information and indicators used in the context of this study are presented in the coming sections.

IV.1 – The 2005 Statistic Yearbook of Malawi (National Statistical Office of Malawi, 2005)

The "National Statistical Office of Malawi" is the main government department for the collection and dissemination of official statistics.

The Statistical Yearbook is an annual publication, intended to provide a summary of the most recent statistical data available from official sources. It includes key social and economic indicators, together with statistics on population, education, economic activities, utilities,...etc.

General information, structured at district level, is available in Excel format.

A summary table, with key indicators useful for this study was prepared. This table includes:

- Population per district,
- Population density,
- Pupils, teachers, classrooms, schools per district,
- Pupils per teacher, per classroom, per school,
- Orphans per district,

<table>
<thead>
<tr>
<th>Table 1 : Country and district level information/indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicators</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Spatial distribution of the population</td>
</tr>
<tr>
<td>Spatial distribution of educational infrastructures</td>
</tr>
<tr>
<td>Spatial distribution of health infrastructure</td>
</tr>
<tr>
<td>Health status</td>
</tr>
</tbody>
</table>

Page 10 / 54
Also, recent data on health facilities is provided by JICA inventory, however, the Ministry of Health and Population of Malawi, published statistics for HIV/AIDS, Malaria, TB and under five diarrhoea, number of cases, admissions and deaths by district (2004). Also, we selected the percentage of child under 1 year who are fully immunized as an indicator of child health care.

The 2005 statistical yearbook also provides relevant data and indicators on poverty at national level. These indicators are very useful for checking coherence of WHS results.


**IV.2.1 – Description of the WHO WHS**

The WHO has developed and implemented a Survey Programme and a “World Health Survey” to compile comprehensive baseline information on the health of populations, the outcomes associated with the investment in health systems and the baseline evidence on the way health systems are currently functioning.

One of the main objectives of the WHS is to provide policy-makers with the evidence they need to adjust their policies, strategies and programmes as necessary.

The module covers the following:

- Socio-demographic and economic characteristics of surveyed households: Age classes, education, expenditures, etc;
- Health states of population;
- Responsiveness of health systems;
- Coverage, access and utilization of health services;

<table>
<thead>
<tr>
<th>Table 2: Household level information/indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicators</strong></td>
</tr>
<tr>
<td>Household’s size</td>
</tr>
<tr>
<td>Age groups</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Health training</td>
</tr>
<tr>
<td>Health insurance</td>
</tr>
<tr>
<td>Household’s expenditures</td>
</tr>
<tr>
<td>Health state descriptions</td>
</tr>
<tr>
<td>Economic activities</td>
</tr>
</tbody>
</table>

A detailed list of the variables selected for this study is presented in Appendix B. However, the sampling guidelines, samples size and distribution are presented in appendix D.

In the case of Malawi, the WHO WHS visited 5306 households located in 194 clusters (Figure 4).
IV.2.2 – Collection of the geographic component of the survey

By integrating Geography, the 2002 WHS becomes the second biggest effort ever, after the DHS survey that collects Geographic Information at the cluster level.

In this context the objectives were to:

- Obtain the location of each surveyed cluster with the highest precision possible without breaking the confidentiality of the respondents;
- Collect maps that would be used as background for the creation of the thematic map, to perform spatial analysis or apply specific models.

For the localisation of the surveyed cluster two approaches have been used:

- Global Positioning System (GPS) devices were used in 26 countries in order to collect the location of the surveyed household directly in the field;
- Digital maps delimiting the surveyed cluster were collected for the remaining 44 countries.

The collection of geographic variables is based on the sections 0100 (sampling information) and 0200 (geocoding information) of the WHS questionnaire (Figure 5).
In order to ensure the best use of GPS devices in the field, the steps to follow in order to fill in these sections are described in several document and guides which can be downloaded from the WHO WHS web site at [http://www.who.int/healthinfo/survey/en/](http://www.who.int/healthinfo/survey/en/). This includes:

- The GPS field guide.
- The GPS test and use document
- The description of the GPS training materials
- The GPS data collection protocol.

**IV.2.3 – Cleaning of the geographic component of the survey**

The cleaning of the geographic component of the WHS has been done in several steps as described in Figure 6.
a) emergency check:
The aim of this check is to detect early any acquisition problem like GPS settings error (coordinate system or units), inadequate use of the GPS or systematically missing data.

b) data cleaning:
This step is done once the full data set is completed, correction are proposed and submitted to the field validation before to go further.

Figure 6: description of cleaning process used for the geographic component in the context of the WHO WHS

IV.2.4 – Generation of the Geo subset

After having cleaned the geographic component of the survey additional geovariables have been calculated including:
- The aggregation of data to cluster level, taking into account country respective data confidentiality policies;
- The calculation of the weighted centre of gravity for each surveyed cluster;
- The generation of different parameters and indexes giving an indication of the dispersion of the interviewed households in the cluster around its weighted centre of gravity.

Finally the names of the administrative units have been integrated in the data set using the information collected through the Second Administrative Level Boundaries data set project (SALB).

IV.3 – The Malawi health facility census (JICA, 2002)

JICA conducted a full health facility census based on a set of questionnaires. The survey covered all the 670 public and NGO/Mission operated facilities above health centre level, voluntary counselling and testing centres, some private facilities, MOHP offices, and rehabilitation centres.

The Malawi Health facility inventory collected data on available types of services, human resources and condition of equipment and physical infrastructure.

The data items collected include the following data:
- Health facility physical infrastructure,
- Type of service provided,
- Equipment,
- Ownership of the facility,
- Road access. ..etc.
<table>
<thead>
<tr>
<th>Indicators</th>
<th>Variables</th>
<th>Pertinence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health infrastructure</td>
<td>Total health facilities per district,</td>
<td>It is evident that many questions concerning the provision of health care are related to space. Indeed, health problems vary in space and so do the needs of the people. Therefore, the spatial distribution of health facilities and services they offer should answer the needs of populations varying in numbers, densities, and health problems.</td>
</tr>
<tr>
<td></td>
<td>Health facilities per types,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health facilities per ownership,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health facilities per road access,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Services provided.</td>
<td></td>
</tr>
</tbody>
</table>

A detailed list of the variables selected for this study is presented in Appendix C.

**IV.4 – The Second Administrative Level Boundaries data set project (SALB)**

SALB is a UN project which has been launched in the context of the activities of United Nations Geographic Information Working Group (UNGIWG) and in the continuity of different efforts that took place in the middle of the 90’s where the delimitation of the administrative boundaries was needed for the creation of population distribution grids (Tobler et al., 1995).

From its original objective, which was to provide the international community with a global standardized GIS layer containing the delimitation of the administrative boundaries down to the second sub national level representative of January 2000, the project has moved into providing a working platform for the collection, management, analysis, visualization and sharing of sub national data for a more larger period as this data set contains information for the first sub national level since 1990 and for the second sub national level since 2000.

For the context of the present study, the map currently under validation for 2000 has been used in order to be able to support the management and analysis of data at the district level.

**IV.5 – GIS, sampling and data quality assurance**

In every data collection initiative results depend on the input; as the principle goes *garbage in - garbage out*. Whatever are the quality of the instrumentation and analytical techniques, basic quality of the results depends heavily on the implementation of the survey, sampling and proper conduct of the questionnaire.

**IV.5.1 – Sampling design**

A better integration of GIS capacities in the sampling framework seems to be important according to the following observations:

- A proper sampling is essential for the generation of the survey results. To ensure sampling quality, accurate sampling strategies have to be set up;

- The survey sampling aims to:
  - Adhere to the strictest scientific standards possible;
  - Provide a Probabilistic Sampling Design;

---

4 More information regarding this project can be found at: http://www3.who.int/whosis/gis/salb/salb_home.htm.

7 Malison et al. 1987, in :
o Be a Nationally Representative Sample;

o The health survey sampling design is based on multi-stage stratified random cluster sample;

o The stratification aims to reduce sampling variance and to ensure accurate generalization of samples results to the entire population. Stratification factors are related to outcome:
  o Demographic, geographic or administrative variables;
  o Socio-economic;
  o Epidemiological;
  o Etc...

o Capacity building in countries is crucial.

The importance of this specific process to be conducted at the beginning of a survey as well as the lack of tools for supporting it have been at the origin of the generation of the Survey Sampling Designer (SSD). This application, which has been developed in the context of the internship performed at WHO, is presented in Appendix A of this report.

**IV.5.2 – Sample deviation index (SDI)**

The SDI is mainly used to see whether the survey data represent the census data in terms of the age-gender structure. It indicates the representativity of the survey sample and is calculated for each category as:

$$SDI(n) = \frac{\text{Sampled value proportion}}{\text{All population proportion}}.$$  

A value of 1 is expected, a value close to zero indicates that the sample is under represented in that category, a value in excess of 1 indicates the category is over represented. This method gives an idea of the representativity of the sample but does not indicate the "distance" between data sets.

It is therefore necessary, before starting any data analysis, to make sure that the distribution of the surveyed households according to the generated geographic variables is representative of the distribution of the total households in the country so that the results of the analysis will be spatially representative of the country.
V. METHODS

This section describes the methods which have been used in the context of this study.

V.1 – Standardisation of the different data sets

The geo-referenced data used in the context of this study being all un-projected (Geographic projection) in the same reference system (WGS 84) it has not been necessary to perform any modification of this type prior to the analysis.

Nevertheless, as the administrative unit names collected through the SALB project are representing a validated source of information, it has been necessary to integrate these names and codes in all the other databases in order to facilitate the merge of data tables as well as the analysis at the country level (Figure 7).

Also, the use of different coding scheme as well as the necessity to be in a position to link each household with the corresponding health facility asked for some standardization work.

The establishment of the link between the household ID from the WHO WHS and the health facility ID coming from the Health facility census (JICA) was based on section 7 of the WHS questionnaire (seeing health care providers). Questions q7300 and q7400 was containing the name of the last visited health facility.

The challenge was then to find the same facility among the JICA database in order to attribute a common code in both databases and therefore being able to draw catchment’s areas like reported in Figure 8.

V.1.1 – Integration of the SALB codes in the different data sets

As the administrative unit names collected through the SALB project are representing a validated source of information it has been decided to integrate these names and codes in all the other database in order to facilitate the analysis at the country level (Figure 7).
SECOND ADMINISTRATIVE LEVEL BOUNDARIES

<table>
<thead>
<tr>
<th>ADM1_NAME</th>
<th>ADM2_NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Region</td>
<td>Balaka</td>
</tr>
<tr>
<td>Southern Region</td>
<td>Blantyre</td>
</tr>
<tr>
<td>Southern Region</td>
<td>Chikawa</td>
</tr>
<tr>
<td>Southern Region</td>
<td>Chiradzulu</td>
</tr>
<tr>
<td>Northern Region</td>
<td>Chitipa</td>
</tr>
<tr>
<td>Central Region</td>
<td>Dedza</td>
</tr>
<tr>
<td>Central Region</td>
<td>Dowa</td>
</tr>
<tr>
<td>Northern Region</td>
<td>Karonga</td>
</tr>
<tr>
<td>Central Region</td>
<td>Kasugu</td>
</tr>
</tbody>
</table>

NATIONAL STATISTICAL OFFICE

<table>
<thead>
<tr>
<th>Pop 2004</th>
<th>Area</th>
<th>Density</th>
<th>Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>295'623</td>
<td>2'193</td>
<td>135</td>
<td>154</td>
</tr>
<tr>
<td>1'027'808</td>
<td>2'012</td>
<td>511</td>
<td>224</td>
</tr>
<tr>
<td>425'080</td>
<td>4'755</td>
<td>89</td>
<td>116</td>
</tr>
<tr>
<td>273'993</td>
<td>767</td>
<td>357</td>
<td>81</td>
</tr>
<tr>
<td>152'691</td>
<td>4'288</td>
<td>36</td>
<td>171</td>
</tr>
<tr>
<td>582'289</td>
<td>3'624</td>
<td>161</td>
<td>206</td>
</tr>
<tr>
<td>469'924</td>
<td>3'041</td>
<td>155</td>
<td>233</td>
</tr>
<tr>
<td>230'026</td>
<td>3'355</td>
<td>69</td>
<td>154</td>
</tr>
<tr>
<td>589'019</td>
<td>7'878</td>
<td>75</td>
<td>323</td>
</tr>
<tr>
<td>9856</td>
<td>18</td>
<td>548</td>
<td>101</td>
</tr>
</tbody>
</table>

WORLD HEALTH SURVEY

<table>
<thead>
<tr>
<th>&lt; 5</th>
<th>HH_Mmbrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.23</td>
<td>3.9</td>
</tr>
<tr>
<td>13.33</td>
<td>3.9</td>
</tr>
<tr>
<td>13.62</td>
<td>3.8</td>
</tr>
<tr>
<td>12.50</td>
<td>3.3</td>
</tr>
<tr>
<td>13.80</td>
<td>4.9</td>
</tr>
<tr>
<td>16.58</td>
<td>4.2</td>
</tr>
<tr>
<td>16.42</td>
<td>4.9</td>
</tr>
<tr>
<td>16.09</td>
<td>4.8</td>
</tr>
<tr>
<td>13.51</td>
<td>4.0</td>
</tr>
</tbody>
</table>

HEALTH FACILITIES INVENTORY

<table>
<thead>
<tr>
<th>Hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>33</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

Figure 7: District level links
V.1.2 – Establishment of the link between the household ID from the WHO WHS and the health facility ID coming from the Health facility census (JICA)

In order to be able identifying to which health facility the persons leaving in a particular household surveyed during the WHO WHS went it has been necessary to identify these links using the information reported in the WHS questionnaire regarding the name of the last visited health facility.

The challenge was then to find the same facility among the JICA database in order to attribute a common code to both objects and therefore being able to draw catchments areas like reported in Figure 8.

Figure 8: Facilities/Households levels links
V.2 – Selection of health facilities for the case study

In order to explore the potential uses of linking GIS tools with households and facilities surveys datasets, 29 health facilities were selected in order to perform a case study on the method capacities.

This selection has been based on the following criteria and process:

- JICA database contains 670 health facilities with a valid geographic coordinate
- 243 of these facilities have been quoted by 2830 of the 5727 households surveyed in the context of the WHO WHS.
- Unique codes were assigned to heath facilities in both JICA and WHS databases.
- Facilities which are quoted in less than 3 WHS questionnaires (102 facilities) are not considered for this case study: 141 facilities remains.
- 29 Health facilities (20 % of remaining facilities) selected for this case study.
- The selection criteria were:
  - The geographic distribution (Regions / Districts) of facilities,
  - Type of health facility (District hospital, Central hospital, Clinic, Rural dispensary,..),
  - Ownership,
  - Road access,
  - Most / Least frequently visited (WHS questionnaires).

V.2.1 – Distribution of selected facilities

The distribution of all the facilities quoted in the WHS as well as those selected for this case study are reported in Figure 9

![Selected health facilities (29)](image1)

![Quoted Health facilities – WHS (243)](image2)

Figure 9: Distribution of selected (a) and quoted (b) health facilities.
V.2.2 – Characteristics of selected facilities

In order to ensure the relevance and accuracy of results of the case study, the selected health facilities should be representative of the country’s facilities key characteristics. Main types of facilities and ownerships were represented in the sample. Also, the quality of road access to the facility as well as the frequency of its use (as reported in WHS questionnaires) were considered. Figure 10 display the distribution of these indicators.

Regarding figure 10 (d), the frequencies of use of the selected health facilities were classified using the method of “Natural Breaks” in ArcView. The class’s significance is as follows:

- Class 1: More than 80 respondents had used the facility,
- Class 2: 65 – 80 respondents,
Class 3: 50 – 64 respondents,
Class 4: 35 – 49 respondents,
Class 5: 10 – 34 respondents,
Class 6: Less than 10 respondents.

V.2.3 – Distances travelled to the selected facilities

The steps that have been followed in order to calculate the distance travelled to the selected facilities are the following:

- For each selected Health facility, two tables have been created and imported in the GIS Software (ArcView):
  - the first one containing the geographic coordinates of each health facility and corresponding characteristics (JICA database),
  - the second one containing the geographic coordinates of each respondents in the WHS who had visited one of JICA’s facility
- In ArcView, shape files has been created based on each table and their projection changed to metric,
- A script (Spider diagram) has been used to calculate the linear distances between each household and the visited facility,
- A new theme with corresponding distances has been created.

This process resulted in the creation of 87 shape files, 3 for each health facility:

- the first one containing the facility characteristics (JICA database),
- the second one containing the information for the respondents (users) characteristics (WHS database),
- the last one containing the linear distances between each household and the visited facility.
VI. RESULTS

This section presents the results obtained by applying the different methods which have been identified for analysing the health system in place on the basis of the available data.

On the different figures the facilities are represented by red crosses while the selected households are represented by a point.

The link between each household and the visited facility is represented by a blue line.

VI.1 Health facilities catchments area analysis

VI.1.1 —Country level analysis

VI.1.1.1 — General overview

Figure 11 contains all the links that it has been possible to identify between the households surveyed in the context of the WHO WHS and the health facility census founded by JICA.

Figure 11 : Global representation of distances travelled to selected facilities.
The first thing we can observe on this figure is the fact that some patients might travel a long distance to obtain care. This first observation does already raise some questions:

- did these patients really travelled that far?
- is there not a facility closer to their place where they could have obtained the same treatment?
- was the question in the WHS correctly spelt and/or were some information missing?

These are some of the questions that this section will try to answer by focusing on some particular examples.

For this, different type of information will be extracted from:

- the WHS data:
  - Separate Outpatient / Inpatient,
  - Satisfied / Dissatisfied,
  - Transportation media, ...
  - Respondent main characteristics (Education, income,...)
  - ......

- And from JICA's data set:
  - Services provided,
  - Type of facility,
  - Ownership,
  - Road access,
  - ......

### VI.1.1.2 – Average distance travelled to obtain care

Distances travelled to receive treatment is a pertinent indicator of health services accessibility and could be linked to several variables related to:

- Households characteristics : Income groups, education level,
- Health facilities characteristics : Services provided, cost of care, road access,
- Patient conditions : More / less severely ill.

On the other hand, it’s now assumed that proximity to health services does not necessarily guarantee use’. Situations of bypassing, patient preferences, social barriers (minorities, ethnic groups, ...) should therefore be identified and analysed.

Figure 12 class the 29 selected health facilities according to the mean linear distance calculated on the basis of all the households who reported having visited the respective facility:

- Class 1 : More than 39 Km,
- Class 2 : 28 – 39 Km,
- Class 3 : 15 – 28 Km,
- Class 4 : 9 – 15 Km,
- Class 5 : 3 – 9 Km,
- Class 6 : Less than 3 Km.
VI.1.2 – Health facility level analysis: Calculation of statistical indicators

Data on distances travelled from each household to attend each selected facility could be exploited in calculating several statistical indicators as present for some of the selected hospitals in Figure 13 to 15.
Lilongwe Central Hospital:
- Central Hospital – MOHP,
- Road Access : Tarmac,
- Services provided : 18 / 21,
- Respondents (WHS) who had visited the facility : 63,
- Distances travelled:
  - Mean : 55.3 Km
  - Min : 1.7 Km
  - Max : 393.1 Km
  - Standard Deviation : 92.2
  - Skweness : 2.07
  - Kurtosis : 3.26

Figure 13: Lilongwe Central Hospital catchment’s area.

St Gabriel Hospital:
- Hospital – CHAM,
- Road Access : Earth Fair,
- Services provided : 18 / 21,
- Respondents (WHS) who had visited the facility : 53,
- Distances travelled:
  - Mean : 9.1 Km
  - Min : 3.1 Km
  - Max : 20.6 Km
  - Standard Deviation : 4.3
  - Skewness : 0.88
  - Kurtosis : -0.41

Figure 14: St Gabriel catchment’s area.
Matiki Health Centre:
- Health Centre – Private,
- Road Access: Gravel,
- Services provided: 19/21,
- Respondents (WHS) who had visited the facility: 7,
- Distances travelled:
  - Mean: 2.3 Km
  - Min: 2.1 Km / Max: 2.4 Km
  - Standard Deviation: 0.09
  - Skewness: -0.13 / Kurtosis: -0.77

These statistical indicators are very useful for classifying health facilities and establishing homogenous groups regarding to selected criteria.

**VI.2 Prospective applications of the method**

**VI.2.1 – Analysing health services accessibility**

Accurate data and adequate analytical tools are especially critical in tackling health inequalities, which covers economic, geographical and cultural facets.

In fact, many factors are affecting the access to public health, such:

- The availability of health facilities;
- The ability and willingness to pay;
- The quality of care;
- The degree of education and awareness;
- Etc...

Data on households income, health expenditures, health status, education and health facilities characteristics,...etc are very useful to analyse this question. Furthermore, linking these data with geographical information using GIS can represent a powerful way to develop practical strategies regarding to health system responsiveness.
For example, Figure 16 shows the distribution of distances travelled to a selected health facility linked to respective households’ expenditures.

This spatial analysis aims to determine whether and how households’ income affects access to health services.

In fact, similar presentations could be produced for the main country’s health centres and then analysed taking into account the types of facilities (government, NGO, private,..etc) or costs of health care.

Then, the other factors that could affect access to public health, for example education or social groups could be analysed and appropriate maps created.

**VI.2.2– Analysing proximity: Close-set facilities**

The case of availability of government, NGO and private sources of health care in relatively close proximity is valuable for analyzing patient preferences and for looking deeply in factors explaining their choices.

Figures 17 a) and b) above shows the case of two different types of health facilities (Government / NGO), providing equivalent services and located in a close area.
VI.2.3 – Analysing preferences (bypassing)

“That a person bypasses a facility is almost certainly indicative either of significant problems with the quality of care at the bypassed facility or of significantly better care at the alternative source of care chosen. When it is a poor person choosing to bypass a free public facility and pay for care further away, such action is especially bothersome to public policy-makers”.

---

VI.2.4 – Analysing road access

Several questions could be considered and analyzed:

- Does distances travelled to health facilities depend on quality of road access?
- Does accessibility to health services results in improved quality of road access?

Data should be carefully analyzed taking account of transport modes, urban / rural areas, income and social groups and then situations of scattered communities where people walk for hours to get to health facilities could be revealed.
VI.2.5 – Outpatient / Inpatient health care

Effective use of health care services and type of services needed are essential for community health needs assessment.

As an example, correlations between age groups structure in surveyed households and type and use of health services are very valuable for planning health services delivery.
VI.2.6 – Patients assessment of the quality of health services

"Studies are needed to examine the management of health facilities and the clients' perception of health-care providers".\textsuperscript{10}

Types of facilities (government / private / NGO, etc) and patients conditions (income, education) should be analyzed. Also, regional disparities (urban/rural, richest/poorer) and ethnic groups distribution (minorities) could be largely linked to patients assessment of health services.

![Figure 23: Patient global assessment of Ndamera health centre service.](image)

VI.3 – Health services use analysis

Effective use of health care services is necessary for:

\begin{itemize}
  \item Health needs evaluation,
  \item Spatial distribution of health services analysis,
\end{itemize}

In another hand, the relationship between household’s socioeconomic characteristics, health status and health services use is valuable for health services delivery planning.

For this section, thematic maps were systematically produced in order to overlay the spatial distribution of health care services use and availability.

VI.3.1 – Global use of health care services

Global use of health care services were assessed using WHS-Malawi individual questionnaires (2002), section 7000 (Health system responsiveness), sub-sections: “Needing health care and General Evaluation of Health System” & “Seeing Health Care Providers”.

Respondents were classified in the following sub-groups regarding to their use of health care services:

- Category 1: Respondents who haven’t needed health care,
- Category 2: Respondents who have needed Outpatient health care (Adult / Child),
- Category 3: Respondents who have needed inpatient health care (Adult / Child).

**VI.3.2 – Inpatient care services use**

![Figure 24: % surveyed households that have used inpatient health care services for Adults.](image)

![Figure 25: % surveyed households that have used inpatient health care services for Childs.](image)

**VI.3.3 – Outpatient (Ambulatory) care services use**

![Figure 26: % surveyed households that have used ambulatory health care services for Adults.](image)

![Figure 27: % surveyed households that have used ambulatory health care services for Childs.](image)

---

11 WHS – Malawi (2002), Individual questionnaire, section 7000 (Health system responsiveness), sub section, "Inpatient Hospital"
12 Idem 11
**VI.3.4 – Factors correlated with health care services use**

In order to increase the use of health services, policymakers should assess the factors determining the use of existing facilities, before planning more health facilities.

**VI.3.4.1 – Method**

The surveyed households were divided into 2 sub-groups regarding to health care services utilization:

- Respondents that haven’t needed health care,
- Respondents that have needed health care,

Then, the main characteristics of sub-groups are compared.

**VI.3.4.2 – Results**

Tables below show results of comparisons between sub-groups for selected variables.

**Table 4 : Average household’s size / sub-groups**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Males %</th>
<th>Females %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haven’t needed health care</td>
<td>3.6</td>
<td>48.2</td>
<td>51.8</td>
</tr>
<tr>
<td>Needed health care</td>
<td>4.5</td>
<td>47.8</td>
<td>52.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4.3</strong></td>
<td><strong>47.9</strong></td>
<td><strong>52.1</strong></td>
</tr>
</tbody>
</table>

**Hypothesis for means comparisons test**: Health services use is significantly correlated to household’s size.

**Table 5 : Household’s age classes / sub-groups**

<table>
<thead>
<tr>
<th></th>
<th>Haven’t needed health care</th>
<th>Needed health care</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 5</td>
<td>9.2</td>
<td>17.1</td>
<td>15.3</td>
</tr>
<tr>
<td>5 To 11</td>
<td>20.0</td>
<td>22.7</td>
<td>22.1</td>
</tr>
<tr>
<td>12 To 18</td>
<td>19.9</td>
<td>16.5</td>
<td>17.3</td>
</tr>
<tr>
<td>19 To 40</td>
<td>30.7</td>
<td>31.9</td>
<td>31.6</td>
</tr>
<tr>
<td>41 To 60</td>
<td>11.7</td>
<td>8.4</td>
<td>9.1</td>
</tr>
<tr>
<td>61 To 70</td>
<td>4.9</td>
<td>2.2</td>
<td>2.8</td>
</tr>
<tr>
<td>&gt; 70</td>
<td>3.6</td>
<td>1.3</td>
<td>1.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

**Hypothesis for means comparisons test**: Significant higher percentage of under 5 years old children in households that have needed health care.

---

13 WHS – Malawi (2002), Individual questionnaire, section 7000 (Health system responsiveness), sub section, "Outpatient and care at home"
14 Idem 13
15 WHS – Malawi (2002), Crosstable:
  - Individual questionnaire, section 7000 (Health system responsiveness), sub section, "Needing health care and General Evaluation of Health System",
  - Household questionnaire, section 0400 (Household roster).
16 Idem 15
Table 6: Household’s header education / sub-groups

<table>
<thead>
<tr>
<th></th>
<th>Haven’t needed health care</th>
<th>Needed health care</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No formal schooling</td>
<td>30.9</td>
<td>29.4</td>
<td>29.7</td>
</tr>
<tr>
<td>Less than primary school</td>
<td>53.9</td>
<td>54.3</td>
<td>54.2</td>
</tr>
<tr>
<td>Primary school completed</td>
<td>11.5</td>
<td>12.3</td>
<td>12.1</td>
</tr>
<tr>
<td>Secondary school completed</td>
<td>3.1</td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td>High school completed</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>College completed</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Post graduate degree</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

Table 7: Household’s years of formal education / sub-groups

<table>
<thead>
<tr>
<th>Years of formal education completed</th>
<th>Haven’t needed health care</th>
<th>Needed health care</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.3</td>
<td>6.4</td>
<td>6.1</td>
</tr>
</tbody>
</table>

Hypothesis for means comparisons test: No significant differences between groups regarding to household’s header education level. However, in terms of years of education completed, household’s that have needed health care seems to be slightly more educated than the others.

Table 8: Household’s expenditures / sub-groups

<table>
<thead>
<tr>
<th>Household’s Expenditures (Kwacha)</th>
<th>Total</th>
<th>Per household member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haven’t needed health care</td>
<td>2007.4</td>
<td>720.6</td>
</tr>
<tr>
<td>Needed health care</td>
<td>2649.7</td>
<td>669.8</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>2476.9</strong></td>
<td><strong>683.4</strong></td>
</tr>
</tbody>
</table>

Hypothesis for means comparisons test: Highest expenditures in households that have needed health care are only related to household’s size.

Table 9: Household’s header main occupation / sub-groups

<table>
<thead>
<tr>
<th>Job</th>
<th>Haven’t needed health care</th>
<th>Needed health care</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government employee</td>
<td>2.4</td>
<td>3.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Non-government employee</td>
<td>10.2</td>
<td>7.8</td>
<td>8.4</td>
</tr>
<tr>
<td>Self-employed</td>
<td>33.1</td>
<td>37.9</td>
<td>36.6</td>
</tr>
<tr>
<td>Employer</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Not working for pay</td>
<td>53.8</td>
<td>50.1</td>
<td>51.1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

---

17 WHS – Malawi (2002), Crosstable:
  - Individual questionnaire, section 7000 (Health system responsiveness), sub section, “Needing health care and General Evaluation of Health System”
  - Individual questionnaire, section 1000 (Respondent’s Socio Demographic characteristics).
18 Idem 17.
19 WHS – Malawi (2002), Crosstable:
  - Individual questionnaire, section 7000 (Health system responsiveness), sub section, “Needing health care and General Evaluation of Health System”
  - Household questionnaire, section 0800 (Household Expenditure).
20 WHS – Malawi (2002), Crosstable:
  - Individual questionnaire, section 7000 (Health system responsiveness), sub section, “Needing health care and General Evaluation of Health System”
  - Individual questionnaire, section 1000 (Respondent’s Socio Demographic characteristics).
Hypothesis for means comparisons test: Highest proportion of respondents without economic activity or regular income in the category of households that haven't needed health care.

This could expose the cases of social and/or financial barriers to get health care.
VII. CONCLUSIONS

This study demonstrates the significant contribution of GIS in this process of:

- Linking several spatially distributed datasets,
- Integrating spatial analysis for a better understanding of health issues,
- Developing appropriate tools and applications for analysing health information.

This necessitates:

- The integration of GIS in the context of health surveys (at households and facilities levels) by the use of GPS for data collection,
- The generation of the geographic component of the surveys.

The role of surveys is crucial for health information systems by providing updated and reliable field data and the integration of geography is valuable for analysing these data from a spatial perspective.

In this context, this study developed a method for linking multiple geo-referenced datasets using GIS.

Various potential applications of the method for analysing health system responsiveness and accessibility were explored. Among these capacities we can mention:

- The bypassing phenomenon study,
- The barriers to health services access,
- The effect of distances travelled, cost and quality of services related to households income and education level.

The implications of these applications on the planning and management of health services are multiple:

- The establishment of decision tables based on the classification of health facilities in regard to selected factors related to health services accessibility,
- The consideration of population socio-economic characteristics in the assessment of health care needs and the planning of health services delivery,

Also, the study analysed the effective use of health services reported in WHS survey which is valuable for:

- The process of assessment of health services needs,
- The analysis of the spatial distribution of health services.
- The understanding of the relationship between health services availability and use.

Finally, it is worth noting that most of the middle and least developed countries face the problem of organizing health service delivery in a manner that provides adequate quality and coverage of health care to their populations against a background of economic recession and limited resources.

In this regard, assisting efforts in collecting reliable data for these countries and research projects demonstrating ways to analyse data collected using GIS and aiming to improve health services seems to be highly promising.
VIII. BIBLIOGRAPHY


WHO (2002): The WHS sampling guidelines for participating countries.


Appendix A : The Survey Sampling Designer (SSD)

The survey sampling designer (SSD) is an application developed in "MS Access" which generates a stratified sample based on 4 stratification variables. Using this application, the goal is to obtain a sample from a population that effectively captures its main characteristics. Thus, the precision of sample estimates will increases and also the size of sample need will decreases. The stratification variables selection is a determining factor depending on the survey objectives.

As shown in the application main menu (Figure 1), the steps required in regards to achieve this objective are to:

- Import the sampling data and declaring units and variables labels,
- Calculate the basic statistics for each stratification variable used,
- Select the appropriate classification method and define homogenous classes for each stratification variable used,
- Define the strata which results from the combination of homogenous classes of all stratification variables used,
- Provide the total sample size, sample by stratum, surveyed clusters, sampling standard errors and final table export.

The user is then going through each window following the order of the section mentioned in the main menu.

The description of the content of each of these windows is reported in the following sections.

To test the application, a virtual example with 4 stratification variables representing the extreme possible cases has been built. These extreme cases were:
First variable: Standard deviation greater than the mean,
Second variable: Standard deviation equal to the mean,
Third variable: Standard deviation lesser than half the mean,
Fourth variable: For this case, data was generated randomly.

In the total, 964 clusters were used to run and test the application.

**I - DATA, LABELS AND UNITS**

The input sampling data can be structured by administrative units, health sectors or any other territorial subdivisions, depending on the survey objectives. These elementary subdivisions represent the initial clusters. In addition to that, the application also runs with data arranged by equal area grid (pixels). The last point concerning the sampling input data, is that administrative or technical boundaries levels should be as lowest as possible in order to have the clusters as smaller as possible improving therefore the sampling accuracy.

Thus, the first step consists in importing the sampling data, declaring the stratification variables labels and the area units to be used (Section A in the Main menu window, Figure 1 above). The related message box informs the user about how to import data, file format and structure. Then area units and variables labels are declared.

**II - BASIC STATISTICS**

The basic statistics are automatically calculated for each stratification variable being used. The form related to basic statistics results is shown in the following figure:

![Basic statistics form](image)

For each stratification variable used, the application calculates the mean, the standard deviation and the minimal and maximal values.

The standard deviation of the total population gives us information about the data distribution around the mean.
III- CLASSIFICATION

The next step consists in choosing the stratification variables classification method. For each stratification variable, the user can choose between the:

- Manual classification: The user has to type the classes break. This will be for example used for variables for which well defined homogenous classes are already known based on regional or local norms (e.g.: Urban / Rural, population density).
- Standard deviations classification method.

The menu related to this step of the process (figure 3), also allows the user to don't classify one or more variables. Consequently, it will be possible to test the impact of each variable in the stratification results and then to decide about the final significant variables to keep for the sampling.

Finally, manual and standard deviations classification method can also be combined for any of the used variables.

A - The standard deviations classification method

The classification method used by SSD is the standard deviations method. In standard deviations classifications, data are assigned to classes based on where they fall relative to the mean and standard deviations of the data distribution. The major advantage of using the mean as a dividing point is that a contrast of values above and below the mean is readily visualized.

This method works well for a dataset that is normally distributed. An even number of classes should be used, such that the mean of the data serves as the dividing point between the number of classes above and below the mean. The major disadvantage of the standard deviations classification is that it requires a basic understanding of statistical concepts, and hence may be difficult to interpret.

In our case, the application first calculates standardized amplitudes above (1) and below (2) the mean:

- \( \text{amplitude}_{above} = \frac{\text{Maximal value} - \text{Mean value}}{\text{Standard deviation}} \)
- \( \text{amplitude}_{below} = \frac{\text{Mean value} - \text{Minimal value}}{\text{Standard deviation}} \)
Then places class breaks at intervals of ½ minimal standardized amplitude above and below the mean. Finally, extreme values exceeding 3*Std Dev from the mean are aggregated into two classes (> 3*Std Dev above the mean & < - 3*Std Dev below the mean).

**B - The classification results**

In SSD, the classification results, shows (figure 4), for each variable used and for each class:

- The number and the % of clusters in the class,
- The area of the class and the % of total area,
- The population of the class and the % of total population,
- The mean, standard deviation, minimal and maximal values of the considered variable.

<table>
<thead>
<tr>
<th>Variable Label</th>
<th>Variable 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td></td>
</tr>
<tr>
<td>Number of clusters</td>
<td>324</td>
</tr>
<tr>
<td>% of clusters</td>
<td>23.65%</td>
</tr>
<tr>
<td>Sum of area</td>
<td>36449.22</td>
</tr>
<tr>
<td>% of area</td>
<td>14.49%</td>
</tr>
<tr>
<td>Sum of sample units</td>
<td>164089</td>
</tr>
<tr>
<td>% of sample units</td>
<td>33.49%</td>
</tr>
<tr>
<td>Average</td>
<td>571.09</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>240.24</td>
</tr>
<tr>
<td>Min</td>
<td>42.52</td>
</tr>
<tr>
<td>Max</td>
<td>907.39</td>
</tr>
</tbody>
</table>

![Figure 4: Classification results](image)

For each class, obtaining low standard deviations in relation to the mean inform the user about the classification quality and classes homogeneity. Generally, a standard deviation equal or smaller than the half of the mean indicates homogenous classes.

When homogenous classes are defined, the next step to be performed concerns the stratification.

**IV - STRATIFICATION**

**A – The stratified sampling**

The most commonly used sampling strategy is based on the stratified sampling design. Stratified sampling designs are frequently used in demographic, socio-economic and health surveys. In this sampling technique, the whole population is divided into mutually exclusive subgroups or strata. The segments are based on some predetermined criteria such as geographic location, size or demographic characteristic. It is important that the segments be as heterogeneous as possible.
Geographical regions corresponding to the lowest administrative level are generally used as elementary clusters in most of the stratified sampling surveys. Smaller are the elementary clusters, better is the stratification results. Clusters can also correspond to pixels in the case of grid data.

If each stratum is homogenous, a precise estimate of any stratum parameter can be obtained from a small sample in that stratum. These estimates can then be combined into a precise estimate for the whole population. In this context, the value of characteristic being measured is the most suitable variable to be used for the stratification of elements in a population. However, in general, this information is not available. Therefore, other variables that are highly correlated to the characteristics of interest are used.

A.1 - Stratified sampling estimators

The following stratified sampling estimators are generally used:

a. Estimate of the mean per stratum :

\[ \bar{x}_j = \sum_{j=1}^{M} P_j \bar{x}_j \]

b. The Estimate of the mean for the population:

\[ \bar{x} = \frac{\sum_{j=1}^{M} N_j \bar{x}_j}{N} \]

c. The Estimate of the total for X for the entire population:

\[ \hat{X} = \sum_{j=1}^{M} N_j \bar{x}_j \]

d. Variance of the mean for the population:

\[ s_{\bar{x}}^2 = \frac{1}{N} \sum_{j=1}^{M} \left[ N_j^2 s_j^2 \left( \frac{N_j - n_j}{N_j} \right) \right] \]

e. The standard error of the total estimate of x:

\[ s_X^2 = N^2 s_{\bar{x}}^2 \]

With:

<table>
<thead>
<tr>
<th>M</th>
<th>Number of strata in the population.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Total number of sampling units measured for all strata (sample)</td>
</tr>
<tr>
<td>nj</td>
<td>Total number of sampling units measured in the jth stratum.</td>
</tr>
</tbody>
</table>
Appendix

30.03.2006

| N     | Total number of sampling units in the population. |
| N_j   | Total number of sampling units in the jth stratum. |
| X_ij  | Quantity X measured on the ith sampling unit of the jth stratum. |
| \bar{X}_j | Mean of X for the jth stratum. |
| \bar{X} | Estimated mean of X for the population. |
| P_j   | Proportion of the total population in the jth stratum. |
| \hat{X} | Estimated total of X for the population. |
| \hat{S}^2_j | Variance of X for the jth stratum. |
| \hat{S}^2_{\bar{X}} | Estimated variance for the mean for the population |
| \hat{S}^2_X | Estimated variance of \bar{X} |

A.2 – Advantages

- Stratification will always achieve greater precision provided that the strata have been chosen so that members of a same stratum are as similar as possible in respect of the characteristic of interest. The bigger the differences between the strata, the greater the gain in precision.
- It is often administratively convenient to stratify a sample. Interviewers can be specifically trained to deal with a particular age-group or ethnic group, or employees in a particular industry. The results from each stratum may be of intrinsic interest and can be analysed separately.
- It ensures better coverage of the population than simple random sampling.

A.3 – Disadvantages

- Difficulty in identifying appropriate strata.
- More complex to organise and analyse results.

B - Stratification

Stratification consists in the combination of homogenous classes obtained for all considered variables. Thus, the first class of the first stratification variable, will be divided on sub-classes based on the second variable classes and so on until all variables classes are combined. Then, each unique classes combination will constitute a homogenous stratum, as shown in Figure 5:

<table>
<thead>
<tr>
<th>CLASSIFICATION</th>
<th>STRATIFICATION</th>
<th>Variable_2</th>
<th>Variable_2</th>
<th>Strata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var1_1st class</td>
<td>Var1_1st class / Var2_1st class</td>
<td>Strata_1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_1st class / Var2_2nd class</td>
<td>Strata_2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_1st class / Var2_3rd class</td>
<td>Strata_3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var1_2nd class</td>
<td>Var1_2nd class / Var2_1st class</td>
<td>Strata_4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_2nd class / Var2_2nd class</td>
<td>Strata_5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_2nd class / Var2_3rd class</td>
<td>Strata_6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var1_3rd class</td>
<td>Var1_3rd class / Var2_1st class</td>
<td>Strata_7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_3rd class / Var2_2nd class</td>
<td>Strata_8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var1_3rd class / Var2_3rd class</td>
<td>Strata_9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: Example of stratification generating homogeneous stratum
The strata represent groups within a population that should be different one from another, but the members within a same stratum should be similar to each other as much as possible. Then stratification:

- Increases the precision of the sample estimates.
- Decreases the number of sample need.

C – Stratification results

In SSD, the stratification results (figure 6), shows for each stratum:

- The number and the % of clusters in the stratum,
- The area of the stratum and the % of total area,
- The population of the stratum and the % of total population,
- The mean, standard deviation, minimal and maximal values for each considered variable.

<table>
<thead>
<tr>
<th>STRATA SUMMARIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strata</td>
</tr>
<tr>
<td>Number of clusters</td>
</tr>
<tr>
<td>Sum Of Area</td>
</tr>
<tr>
<td>Sum Of Sample Units</td>
</tr>
<tr>
<td>Variable Label</td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Standard dev.</td>
</tr>
<tr>
<td>Variable Label</td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Standard dev.</td>
</tr>
</tbody>
</table>

Figure 6: Stratification results

Like for the classes, the standard deviations must be as low as possible in relation to the mean, which will indicate homogenous strata.

V - SAMPLING

In stratified sampling procedure, the goal is to obtain a sample from a population that effectively captures its main characteristics, so that conclusions made from this sample can be extrapolated to the general population. The goal is to obtain a sample that includes all of the different strata within the population. Two steps are therefore required:

- Defining the total sample size;
- Dividing the sample between strata.

A - Defining the Sample Size
Determining the sample size is a very important issue because too large samples may waste time, resources and money, while too small ones may lead to inaccurate results. In many cases, we can easily determine the minimum sample size needed to estimate a process parameter, such as the population mean $\mu$.

When sample data is collected and the sample mean $\bar{x}$ is calculated, that sample mean is typically different from the population mean $\mu$. This difference between the sample and population means can be thought of as an error. The margin of error $E$ is the maximum difference between the observed sample mean $\bar{x}$ and the true value of the population mean $\mu$:

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

where:
- $z_{\frac{\alpha}{2}}$ is known as the critical value. At 95% degree of confidence, the critical value is therefore $z_{\frac{\alpha}{2}} = 1.96$.
- $\sigma$ is the population standard deviation.
- $n$ is the sample size.

Rearranging this formula, we can find the sample size necessary to produce results accurate to a specified confidence and the margin of error by using the following formula:

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{E}\right)^2$$

This formula can be used when $\sigma$ is known and when we need to determine the sample size necessary to be established, with a confidence of $1 - \alpha$, the mean value $\mu$ to within the margin of error $(\pm E)$.

But in several cases:
- The population standard deviation is rarely known.

AND
- In a stratified sampling design, many factors are considered to define strata.

We can therefore use Bernoulli’s formula which is more adapted for such situations:

$$n = \frac{\left(\frac{z_{\frac{\alpha}{2}} \sigma}{E}\right)^2}{\left(\frac{\sigma^2}{N} + \left(\frac{I^2}{N-1}\right)\right)}$$

Where:
- $n =$ Sample size.
- $N =$ Total population (sample units).
- $Z$ factor $= 1.96$ for confidence level of 95%.
- $I =$ Confidence interval.

The SSD application calculates the optimal sample size using this formula as an indication to the user who is manually entering the final sample size considered by the application during the calculation process.
B – Dividing the sample between strata

Density of samples in each stratum should be proportional to the population counts in the stratum. In this case the maximum precision is reached, and thus, the stratum with higher population density should be sampled more intensively.

After calculating sample size for each stratum, the number of surveyed clusters is defined according to the minimal clusters sample size. Also, strata with sample size lower than this limit will not be surveyed.

It is important to note that several other sample allocation methods exists. It would therefore be possible to integrate them in the application if needed in order to test the optimal allocation to use.

Finally, the sample summarizes form (figure 7) shows general information about total and surveyed strata and clusters.

![Sample Summarize](image)

**Figure 7: The sample summarize**

Concerning sampling quality, the standard deviations of the estimated mean for the population is firstly calculated.

SSD then compares:

- The true population mean, WITH
- The estimated mean for the overall population, assuming only sample data.

VI – FINAL SAMPLING DATA TABLE

The final step of the process consists in exporting the final data table in Excel with all input data and calculated fields, such as:

- Classes for each variable,
- Strata,
- Sample by strata,
- Surveyed clusters by strata.

Once imported in Arc View, it would be possible to visualize the strata distribution as well as the distribution of the clusters to be surveyed and the sample size in each of them.
# Appendix B : Selected variables - WHS

## Sampling information
- Districts & Clusters
- Urban / Rural

## Geocoding informations
- Geographic coordinates

## Household roster

### Age

### Education
- No formal schooling
- Less than primary school
- Primary school completed
- Secondary school completed
- High school (or equivalent) completed
- College / pre-university / university completed
- Post graduate degree completed

## Household members ever worked or trained in a health-related field

## Health Insurance
- Household members covered by any kind of health insurance plan

## Household Expenditures
- Total household Expenditure (Last 4 weeks)
  - Food
  - Housing, gas, electricity, water, telephone, and heating fuel
  - Education fees and supplies
  - Health care costs
  - Health insurance
  - Others

## Respondent’s Socio Demographic Characteristics

### Mother tongue

### Sex
- Female
- Male

### Age

### Education
- Idem

### Years of school completed

### Ethnic group

### Current job
- Government employee
- Non government employee
- Self-employed
- Employer
- Not working for pay
- If not working for pay

### Main occupation
- Legislator, Senior Official, or Manager
- Professional (engineer, doctor, teacher, clergy, etc.)
- Technician or Associate professional (inspector, finance dealer, etc.)
- Clerk (secretary, cashier, etc.)
- Agricultural or fishery worker (vegetable grower, livestock producer, etc.)
- Craft or trades worker (carpenter, painter, jeweler worker, butcher, etc.)
- Plant/machine operator or assembler (equipment assembler, swing-machine operator, driver, etc.)
- Elementary worker (street food vendor, shoe cleaner, etc.)
- Armed forces (government military)

### Reasons not working for pay
- Homemaker / caring for family
- Looked but can't find a job
- Doing unpaid work / voluntary activities
- Studies / training
<table>
<thead>
<tr>
<th>Health State Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Retired / too old to work</strong></td>
</tr>
<tr>
<td><strong>Ill health</strong></td>
</tr>
<tr>
<td><strong>Other</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health State Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health rate</strong></td>
</tr>
<tr>
<td>Very good</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Bad</td>
</tr>
<tr>
<td>Very bad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health State Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Difficulty with work or household activities</strong></td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>Mild</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Severe</td>
</tr>
<tr>
<td>Extreme / cannot do</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health System Responsiveness: Needing Health Care And General Evaluation Of Health Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household members that haven’t needed health care</strong></td>
</tr>
<tr>
<td><strong>Inpatient health care</strong></td>
</tr>
<tr>
<td>Adult stayed overnight at hospital</td>
</tr>
<tr>
<td>Child stayed overnight at hospital</td>
</tr>
<tr>
<td><strong>Outpatient health care</strong></td>
</tr>
<tr>
<td>Adult received outpatient health care</td>
</tr>
<tr>
<td>Child received outpatient health care</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health System Responsiveness: Needing Health Care And General Evaluation Of Health Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household members that didn’t get health care they had needed</strong></td>
</tr>
<tr>
<td><strong>Reasons</strong></td>
</tr>
<tr>
<td>Could not afford the cost of the visit</td>
</tr>
<tr>
<td>No transport</td>
</tr>
<tr>
<td>The cost of transport</td>
</tr>
<tr>
<td>The health care provider's drugs or equipment are inadequate</td>
</tr>
<tr>
<td>The health care provider's skills are inadequate</td>
</tr>
<tr>
<td>Previously badly treated</td>
</tr>
<tr>
<td>Take time off work or had other commitments</td>
</tr>
<tr>
<td>Did not know where to go</td>
</tr>
<tr>
<td>Were not sick enough</td>
</tr>
<tr>
<td>Were denied health care</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health System Responsiveness: Needing Health Care And General Evaluation Of Health Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global rate of health care in the country</strong></td>
</tr>
<tr>
<td>Very satisfied</td>
</tr>
<tr>
<td>Fairly satisfied</td>
</tr>
<tr>
<td>Neither satisfied or dissatisfied</td>
</tr>
<tr>
<td>Fairly dissatisfied</td>
</tr>
<tr>
<td>Very dissatisfied</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health System Responsiveness: Needing Health Care And General Evaluation Of Health Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Seeing health care provider’s</strong></td>
</tr>
<tr>
<td><strong>Care needed</strong></td>
</tr>
<tr>
<td>Outpatient / Inpatient</td>
</tr>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td>Operated by the government</td>
</tr>
<tr>
<td>Privately operated</td>
</tr>
<tr>
<td>NGO</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health System Responsiveness: Needing Health Care And General Evaluation Of Health Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skills of health care providers</strong></td>
</tr>
<tr>
<td>Adequate / Inadequate</td>
</tr>
<tr>
<td><strong>Equipment</strong></td>
</tr>
<tr>
<td>Adequate / Inadequate</td>
</tr>
<tr>
<td><strong>Drug supplies</strong></td>
</tr>
<tr>
<td>Adequate / Inadequate</td>
</tr>
<tr>
<td><strong>Travelling time to health care provider’s</strong></td>
</tr>
<tr>
<td>Private car or motorcycle</td>
</tr>
<tr>
<td>Public transport</td>
</tr>
<tr>
<td>Ambulance</td>
</tr>
</tbody>
</table>
Health care provider evaluation criteria

- The travelling time to the hospital
- The amount of time for waiting before being attended to
- The experience of being greeted and talked to respectfully
- The way privacy was respected during physical examinations and treatments
- The experience of how clearly health care providers explained things
- The experience of getting enough time to ask questions
- The experience of getting information about other types of treatments or tests
- The experience of being involved in making decisions
- The way the health services ensured patients could talk privately
- The way personal information is kept confidential
- The freedom to choose the health care providers
- The cleanliness of the rooms inside the facility
- The amount of space

Appendix C : Selected variables – JICA Health Facility inventory

<table>
<thead>
<tr>
<th>Name</th>
<th>Geographic location</th>
<th>Type of health facility</th>
<th>Physical access</th>
<th>Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>District</td>
<td>Central Hospital</td>
<td>Earth Bad</td>
<td>MOHP</td>
</tr>
<tr>
<td></td>
<td>Address</td>
<td>District Hospital</td>
<td>Gravel Fair</td>
<td>Cham</td>
</tr>
<tr>
<td></td>
<td>Latitude</td>
<td>Clinic</td>
<td>Tarmac Good</td>
<td>BLM</td>
</tr>
<tr>
<td></td>
<td>Longitude</td>
<td>Dispensary</td>
<td>Other: Water / Path..</td>
<td>ADMARC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Health Centre</td>
<td></td>
<td>Forestry / Military / Police / Prison</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hospital / Mental Hospital</td>
<td></td>
<td>NGO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maternity</td>
<td></td>
<td>Private</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rehabilitation Centre</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rural Hospital</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Voluntary Counselling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
</tr>
<tr>
<td>Fair</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOHP</td>
</tr>
<tr>
<td>Cham</td>
</tr>
<tr>
<td>BLM</td>
</tr>
<tr>
<td>ADMARC</td>
</tr>
<tr>
<td>Forestry / Military / Police / Prison</td>
</tr>
<tr>
<td>NGO</td>
</tr>
<tr>
<td>Private</td>
</tr>
</tbody>
</table>

| Ministry of Health & Population |
| Christian Health Association of Malawi |
| Banja La Mtsogolo |
| Non Governmental Organization |
UNIMA

SERVICES PROVISION

Maternity Service
Orthopaedic Service
HIV/AIDS Testing Service
TB Testing Service
Mental Health Service
Ante Natal Service
Malnutrition Service
Dental Service
HIV/AIDS Counselling Service
Malaria Testing Service
Youth Friendly Service
Dispensary Service
Family Planning Service
Blood Bank Service
Ambulance Service
STDS Service
Water Sanitation Service
Underfive Service
Laboratory Service
Mortuary Service
Immunisation Service

Appendix D : Malawi Sampling Plan

Administrative Structure :

- 3 Regions (Northern, Central, Southern),
- 27 Administrative districts (26 Mainland : 5 North, 9 Central, 12 South – 1 on island on lake),
- Each district divided into Traditional Authorities (TAs),
- Rural : Each TA divided into Enumeration Areas (EAs),
- Urban : Each City divided into Wards, each Ward divided into EAs,
- EAs comprise, on average, 250 households,
- The average household size is 4.5 person,
- Urban Rural split : Urban 14 % - Rural 86 %,

Definition of Survey population :

- Sample size 5880,
- Target / Survey population : Adults (18 years of age and over) currently residing in Malawi excluding the island on lake Malawi,
- Sampling units : Households,
- Observational units : Individuals,

Design specifications :

- Number of stages : 3 stages. PSU : Rural (Traditional Authorities) – Urban (Wards), SSU : Enumeration Areas, TSU : Households,
- Specifications : Mainland Districts (26 strata) + 4 Urban Centers (4 strata – self reporting PSU) – TOTAL : 30 strata,

PSU Selection Summary :

- Rural Districts :
PSUs (TAs) selected with probability proportional to population size (PPS) in each District,
3 TAs randomly selected from each Northern Districts and 2 TAs from each Central and Southern Districts,
Number of rural PSUs in survey sample = 58,

Urban Districts:
PSUs (Wards) selected with probability proportional to population size (PPS) in each District,
4 Wards randomly selected in each City and 3 Wards randomly selected in Zomba Municipality,
Number of urban PSUs in survey sample = 15,

SSU Selection Summary:

Rural TAs:
EAs (SSU) selected with probability proportional to population size (PPS) in each,

Urban Wards:
2 EAs selected in each randomized ward,

Household selection:

Rural EAs:
30 Households systematically selected in each rural EA,

Urban EAs:
32 Households systematically selected in each urban EA,

Frame used: 1998 census dwelling unit lists