USING GEOSPATIAL ANALYSIS TO INFORM DECISION MAKING IN TARGETING HEALTH FACILITY-BASED PROGRAMS

A GUIDANCE DOCUMENT

Josh M. Colston

Clara R. Burgert

MEASURE Evaluation

www.cpc.unc.edu/measure

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement GHA-A-00-08-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with Futures Group, ICF International, John Snow, Inc., Management Sciences for Health, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government.

June 2014 MS-14-88

Table of Contents

List of Tables and Figures

Acronyms

AIS	AIDS Indicator Survey
ANC	antenatal care
DHS	Demographic and Health Surveys
GIS	geographic information system
GPS	global positioning system
GPW	Gridded Population of the World
GRUMP	Global Rural Urban Mapping Project
KDE	kernel density estimation
LSMS	Living Standard Indicator Survey
MICS	Multiple Indicator Cluster Survey
MIS	Malaria Indicator Survey
MoHSS	Namibia Ministry of Health and Social Services
SARA	Service Availability and Readiness Assessment
SPA	Service Provision Assessment
UNDP	United Nations Development Programme
WHO	The World Health Organization

Overview

To date, there has been little guidance available on using geospatial methods for targeting health system resources so as to maximize the impact and meet the needs of growing populations. The data necessary for such an analysis comes from a variety of different sources and is not routinely brought together in a single place. This document addresses this gap by providing guidance for using geographic information systems (GIS) to inform decision making about allocating resources for facility-based health services. It presents a process for deciding the best geospatial method to use, given the types of data available, and outlines the steps to take in order to reach a decision. This document is aimed at program planners and managers who have some basic knowledge of GIS terminology and of geospatial analytic methods.

The decision about where to direct resources should be guided by thorough analysis of available data to ensure that investments in health services are targeted effectively. Resources must be allocated carefully in order to maximize the health *impact* of investments (for definitions of words in italics, see Annex A: Glossary, or click on the term for a link to the glossary). This is particularly true for services that are delivered through health facilities. Managers of programs offering these kinds of services need to identify geographic areas where the population has the greatest need. This involves bringing together two types of information:

1. **Data on the existing health service environment:** It is necessary to know the distribution and capacity of the health facilities already offering the services. These are examples of information about the "supply" of the service. Other supply-side information includes input prices and availability, technology, and staff efficiency (Ensor & Cooper, 2004).
2. **Population data:** It is also necessary to know about the size, distribution, and needs of the population that the health facilities serve. This is one component of "demand-side" information. Others include perception of cost and household, community and cultural preferences (Ensor & Cooper, 2004).

Both types of information are needed for calculating the coverage of health services and the population's *access* to the facilities that provide them. Bringing together these two kinds of data is therefore essential for identifying and quantifying the population of potential clients that are *eligible* to use the services of each facility.

Geospatial analysis and GIS software are useful tools for decision making. This is because, to maximize health service coverage and access, it is important to assess whether the locations of health facilities are appropriately distributed throughout populated areas. "Access" to a service refers to the physical presence of a facility offering that service within a given distance or catchment area of its potential clients. A facility's *accessibility* refers to the "travel impedance" — the distance or travel-time — between the client's location and the facility (Guagliardo, 2004; Tanser, Gijsbertsen & Herbst, 2006). A client's access to a particular health service depends on geographic factors such as distance from their home to the facility and their connection to the transportation network. To make the most of a GIS, the two types of data (health service and population data) must be linked by a third type — geographic data. Geographic data are about the physical location — in this instance, the geographic locations of health facilities and the population they are meant to serve.

In recent years there has been a trend towards making data more open-access and increasingly georeferenced. This means that some kinds of geographic data sets are publicly available for most countries in the world. However, the quality, detail, and accuracy of these data vary considerably, putting limitations on the type of geospatial analysis that can be done. Furthermore, detailed geographic health data on both the demand- and supply-side may not be available for a specific country or area of interest. Regardless of the data situation, it is important to have a standard approach to using the data that is available in order to maximize its potential while understanding its limitations.

How to Use this Guidance Document

This document starts by describing the questions that need to be answered and the data required to make a decision about resource allocation. The flowcharts in figures 1 and 2 represent the sequence of decisions that need to be made as part of the geospatial analysis. They are intended to help guide the reader through the process of identifying potential analytical approaches. The process starts with answering the first two questions listed in Sequence of Questions (at right), which pertain to the available data on the health service environment and the population of potential clients.

Sequence of Questions

The sequence of questions to be answered in a geospatial analysis of health facility access are:

1. Where are existing services currently located?
2. Where is there eligible population that does not have access to the service?
3. How should future investments in facility-based services be allocated?

This document then presents an illustrative scenario, in which geospatial techniques are used to target investments in improving antenatal care (ANC) coverage in Namibia. While the scenario is fictitious, the data used are from a real household survey and health facility survey. The annexes give more information about some of the particular topics covered:

- **Annex A** is a glossary giving definitions of specialized terms (which are italicized on first use within this document for ease of reference).
- **Annex B** discusses the structure and sources of health service environment data.
- **Annex C** discusses population data and the advantages and limitations of the various formats in which this data comes.
- **Annex D** details the geospatial analysis methods that are presented in the document.
- **Annex E** presents more general geospatial methods that can be used to convert between different data formats.

Geospatial Analysis Summary

The decision of where to allocate program resources can be guided by the answers to a series of questions that begin with understanding where services are located; then where a target population lives; and finally where there is a need for expanded or new services. The key questions for this analysis are described in more detail in this section.

Question 1: Where Are Existing Services Currently Located?

The first step in the process is to find information about the existing health-service environment. This means knowing the current geographic distribution of health facilities across the area of interest. This information may come in the form of a database or table which lists the health facilities (e.g. a master facility list) and includes a *geographic identifier*.

There are two types of geographic identifier, which are explained further in Annex B: *Administrative unit* identifiers and *location identifier*. Administrative unit identifiers give the name or identification code of the state, region, district or county in which a facility is located. Location identifiers can be either addresses or coordinates. Coordinates typically give the latitude and longitude values for the precise location of the facility.

The type of geographic identifier the data has will determine the kind of analyses that are possible (figure 1, in the next section of this guide). In general, more rigorous and detailed analyses are possible using coordinates. However, data on the latitude and longitude of health facilities' locations tend to be less widely available. In the absence of coordinate data, cruder geospatial analyses can still be conducted based on administrative unit data.

Question 2: Where Is the Eligible Population that Does Not Have Access to the Service?

After establishing the locations of the health-facilities, the next step is to compile data on the eligible population of potential service users for each facility. The eligible population can be defined as those individuals who have the potential to benefit from an intervention. If appropriate, this population data should be disaggregated by age and sex, since the population that is eligible for a particular service is often a subset of the total population. Eligibility for a service can either be defined according to age and/or sex (according to a norm e.g. women of childbearing age are eligible for contraceptive counseling) or by having a certain health condition (e.g. children with diarrhea are eligible for oral rehydration therapy). Examples of facility-based interventions to which these methods may be applied are listed at right.

Examples of facility-based interventions to which methods covered in this document can be applied:

- antenatal care
- family planning
- emergency/basic obstetric care
- HIV voluntary counseling and testing
- malaria diagnosis
- integrated management of childhood illnesses; treatment for diarrhea, fever or cough
- directly observed therapy for treatment of tuberculosis

Georeferenced population data can come in three forms: *aggregated statistics*; *raster* (gridded surface) layers; or *gazetteer* data sets. These are discussed further in annex C. Each of these can be used in different ways to determine where the population with low access to the health-service of interest is located.

Question 3: How Should Future Investments in Facility-Based Services Be Allocated?

Once questions 1 and 2 have been answered, information needs to be brought together to assess how future investment in facility-based services should be allocated. Ideally, resources for health facility-based services should be targeted to where they are likely to have the largest *impact*. This means the largest net improvement in population health status that can be attributed to that intervention. This improvement can be difficult to predict or measure; so often, researchers rely on indicators of *health system performance* instead, such as rates of coverage or *utilization* of a service. A service's *coverage gap* is the proportion of the eligible population that did not utilize the service, or the inverse of the rate of utilization.

Targeting for investment in health-facility based services involves identifying locations where there is eligible population with no access to the service, or with services that are not of an adequate scale to meet their needs. However, access is just one of the many barriers that may prevent the utilization of a service by an eligible individual. Other barriers may include cost, quality, and acceptability of care (Burgert et al. 2011, McLafferty and Grady 2004). Information on utilization and coverage gaps can also inform decisions.

Nevertheless, geospatial analysis lends itself most readily to the assessment of physical *accessibility* component of coverage. This is because accessibility is defined in this document in terms of relative physical location and proximity alone. Other barriers that determine rates of actual service utilization, such as cost, quality, and acceptability, may have spatial components. However, they require very rich data and sophisticated methods to analyze. In addition, the main sources of data on service utilization are population- or facility-based surveys, which only collect information from a sub-set or sample of the total population or health facilities. On the other hand, with the right data, access and accessibility can be analyzed in a GIS for the whole population of an area of interest.

This document will mainly focus on analyzing access to health services in order to show how to estimate the areas of greatest potential impact. The document will also mention how to bring data on utilization and coverage gaps into the analysis.

Sequence of Decisions that Are Part of a Geospatial Analysis

Figures 1 and 2 are flowcharts representing the sequence of decisions that need to be made as part of the geospatial analysis. Figure 1 focuses on how to decide whether to make a *choropleth map* or a catchment area analysis based on the nature of the geographic identifier in your data. Figure 2 gives guidance on deciding what kind of catchment area analysis you can carry out with the data that you have. To demonstrate the process of identifying gaps in access or coverage by following the flow charts, the next section in this guide provides an illustrative example using data from Namibia. To use the flow charts, start at the top of figure 1 and answer the questions about the nature of the available data in order to follow the best path for analyzing it in a GIS.

The blue boxes represent questions that analysts need to answer about the available data and the kinds of outputs they wish to produce. The orange boxes represent methods that can be applied to the data. Where relevant, these boxes include references in parentheses that refer the reader to corresponding sections in the annexes (for example, D.c.i refers to annex D, section c, subsection i). The green boxes represent the outputs of the analysis.

Examples of what to look for when assessing service accessibility based on the outputs of a geospatial analysis include the following:

- Areas with low *accessibility ratios* could indicate a need for increasing the number of facilities offering the service of interest. Where high accessibility ratios overlap with low levels of utilization (high coverage gaps), it can suggest that physical accessibility may not be the main determinant of utilization and that other barriers might exist.
- The population living outside of a health facility's catchment area can be thought of as lacking access to that facility's services. Therefore, *settlements* that do not fall within the service catchment area polygon or buffer of any facility might be potential sites for establishing new facilities.
- It is worth looking at the distribution of health facilities that are not currently offering the service of interest. If any of these are located in areas of low access to that service (i.e., outside any catchment areas), then resources may be better invested in those existing facilities so that they do offer that services, rather than investing in entirely new facilities.

It is important to bear in mind that the process of developing maps or other GIS outputs to support decisions, requires the active engagement of diverse stakeholders. There is rarely a case in which a single map yields an obvious, unambiguous answer to the analysis question. It is also an iterative process since often the interpretation of a first map reveals the need for a second map or the inclusion of more data elements on the first map (Moise, Cunningham & Inglis, ND). This is apparent in the Namibia example in the next section, in which an interpretation of the first map in figure 4 prompts a second question about a specific region, which leads to a subsequent map (figure 5). While the resulting maps may inform a decision-making process, they should be seen as just one data product among many. The production of multiple maps is typically just one part of a larger data use process that involves making multiple smaller decisions to get to the final programmatic decision. Geospatial analysis can be useful for bringing together different types of data from various sources, and presenting them in a more easily interpretable form. However, it is still necessary for the decision maker to understand the limitations of the map and assess it in conjunction with other data products.

Figure 1: **Determining the appropriate geospatial method based on the available geographic identifiers.**
(Letters in parentheses refer to steps outlined in annexes, e.g. “D.a” refers to item a in annex D.)

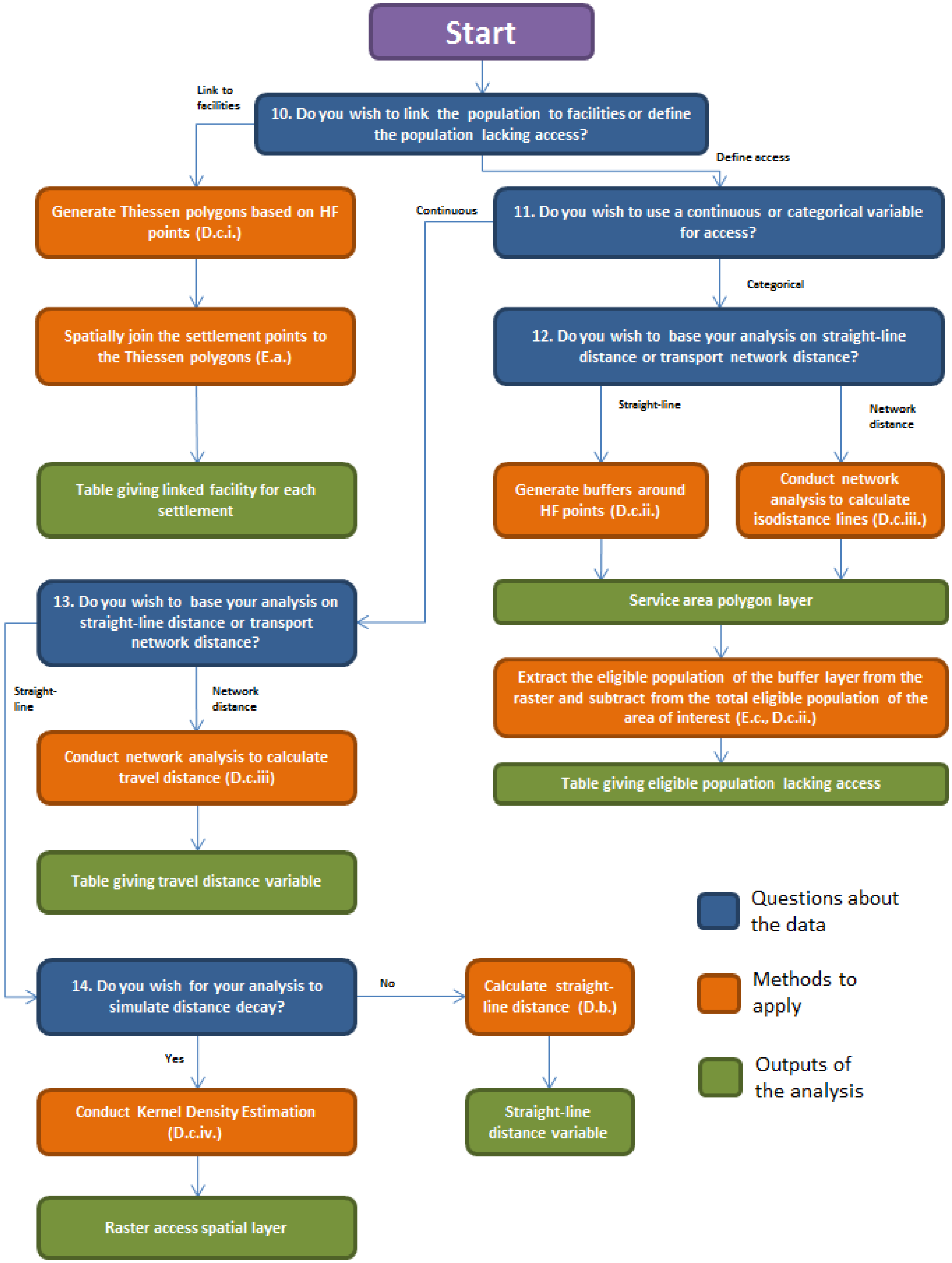


Figure 2: **Determining the appropriate catchment area analysis based on the available data.**
(Letters in parentheses refer to steps outlined in annexes, e.g. “D.a” refers to item a in annex D.)

Example: Access to Antenatal Care in Namibia

This section presents a fictitious scenario to illustrate the data-informed decision-making process and demonstrate the use of geospatial analysis techniques. In this example, these techniques are used to identify geographic areas that might benefit from the investment of resources to improve coverage of antenatal care. The example does not include every method discussed in this document's annexes or in figures 1 and 2, nor does it address every possible data source that can be used. However, it is intended to be illustrative of the type of approach that can be taken to address a certain type of scenario.

Scenario

According to the 2006-07 Namibia Demographic and Health Survey (DHS), 94.6% of women of childbearing age in Namibia reported having received ANC from a skilled provider during their last pregnancy (Namibia Ministry of Health and Social Services & Macro International Inc, 2008). A donor has come forward wishing to invest funds to address gaps in ANC coverage at the sub-national level in order to bring underserved populations closer to the national average. The government of Namibia wants to know whether this high coverage percentage for ANC is masking any sub-national differences in utilization and, if so, where the government should target these new resources in order to get the largest increase in ANC coverage.

The analysis starts by addressing the following questions:

1. Where are the existing ANC services located?
2. Where are the women of childbearing age (15-49) who lack access to ANC services located?
3. How should future investments in ANC services in Namibia be allocated?

The first step in the analysis is to gather together all the available health facility and population data. Table 1 presents some of the regional level indicators that can be used in a geospatial analysis of service access. The methods for compiling and calculating these statistics are explained in the following sections.

Question 1: Where Are Existing Services Currently Located?

The Namibia 2009 Health Facility Census collected information about every facility in the country including the coordinates of their locations and whether or not they offered ANC services (Ministry of Health and Social Services [Namibia] & ICF Macro, 2010). Therefore, the geographic identifiers available for the health facilities are geographic coordinates (the answer to question 1 in figure 1). These were imported into a GIS and mapped over a polygon *shapefile* of the boundaries of Namibia's 14 regions to make the map shown in figure 3. It is important to be able to distinguish between the health facilities that do and those that do not offer ANC, so these are represented by different colored dots in figure 3.

To calculate the aggregated statistics on health facilities, a *spatial join* was carried out to assign region names from the polygon layer to each health facility offering ANC (see annex E, section a). The output of the spatial join was a new point layer. In the *attribute table* of this layer was a column giving the name of the region in which each facility is located. This table was exported to a spreadsheet to calculate the number of health facilities per region (column 2 of table 1).

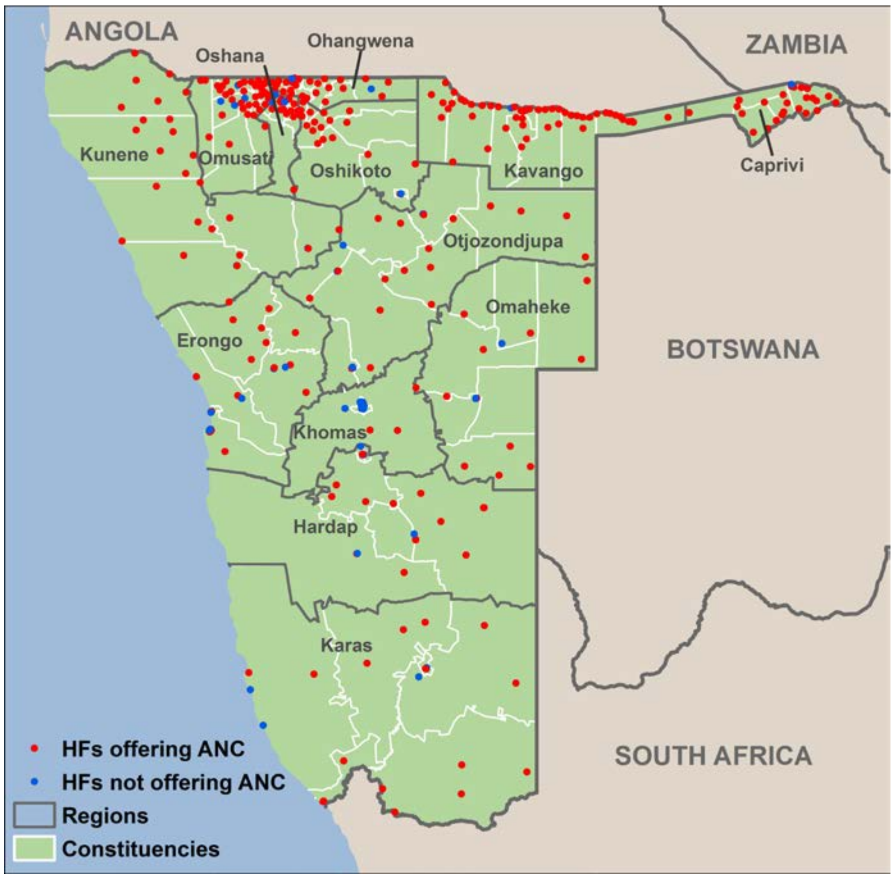


Figure 3: Health facility locations and region boundaries in Namibia.

Table 1: Regional Level Antenatal Care Coverage in Namibia

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Region	Health facilities offering ANC	Total population	Population eligible for ANC (women age 15-49)	Percentage of total population that are eligible for ANC	ANC Accessibility ratio	Percentage of Women age 15-49 who, during their last pregnancy*			Total ANC coverage gap†
						Received ANC from a skilled provider	Received ANC but not from a skilled provider	Did not receive any ANC	
Caprivi	26	89,125	24,275	27.1%	1.07	93.6%	1.4%	4.1%	5.5%
Erongo	19	114,343	34,650	29.8%	0.55	93.0%	0.2%	6.8%	7.0%
Hardap	17	72,484	19,140	25.6%	0.89	95.5%	0.9%	3.1%	4.0%
Karas	19	73,629	22,012	28.9%	0.86	98.8%	0.0%	0.9%	0.9%
Kavango	53	273,660	69,615	25.1%	0.76	91.6%	1.5%	5.9%	7.4%
Khomas	9	348,171	113,350	32.6%	0.08	96.8%	0.4%	2.4%	2.8%
Kunene	25	77,582	19,206	23.9%	1.30	81.4%	10.8%	7.4%	18.2%
Ohangwena	29	270,757	70,271	24.4%	0.41	95.7%	2.0%	2.3%	4.3%
Omaheke	13	81,474	20,236	23.9%	0.64	91.5%	1.3%	6.5%	7.8%
Omusati	41	247,948	68,747	25.9%	0.60	96.9%	1.3%	1.5%	2.8%
Oshana	14	182,030	55,928	30.5%	0.25	98.5%	0.0%	1.2%	1.2%
Oshikoto	20	187,099	51,711	26.3%	0.39	95.3%	0.0%	4.5%	4.5%
Otjozondjupa	18	167,050	45,787	26.5%	0.39	93.4%	0.2%	6.4%	6.6%
Total	**303**	**2,185,352**	**614,928**	**28.1%**	**0.49**	**94.6%**	**1.2%**	**3.8%**	**5.0%**

Notes:
Source of table MoHSS & Macro International Inc 2008; National Planning Commission & UNFPA, 2012.
* Skilled provider" includes doctor and nurse/midwife. Percentages do not add up to 100% because the percentages of women for whom information on ANC was missing were not included in this table
† The percentage of women age 15-49 who, during their last pregnancy, did not receive ANC from a skilled provider (though they may have received it from another provider.

Question 2: Where Is the Eligible Population that Lack Access to ANC?

The population group that is eligible for ANC is all women of childbearing age. Therefore, it is important to gather data on women who are in the 15-49 year age group. For this example, three different kinds of population data were available for Namibia. These data types are explained further in annex C.

Data on Population Distribution and Structure

Aggregated data: Namibia is divided at the first administrative level into 14 regions and at the second into 121 constituencies (see figure 3). There were two types of aggregated data used in this example:

- **Data on population structure:** Population statistics were available from the report of the 2011 Namibia census and are presented in column 3 of Table 1 (National Planning Commission and UNFPA 2012). The data was disaggregated by sex down to constituency level, but data broken down by age-group was not included in the document. A report from the Namibian Central Bureau of Statistics gave region-level population projections for 2011 disaggregated by sex and five-year age group (Central Bureau of Statistics, 2006). From this it was possible to sum the eligible population (women aged 15-49 years) by region for 2011 (column 4 in Table 1). However, since these projections were not available for the constituencies, the lowest level at which aggregated statistics were available for the eligible population was the region (the first administrative level). Note that the population projections are estimates extrapolated from census data. As such they have a margin of error associated with them. The document used in this example presented three variants of the estimates - low, medium and high – based on varying assumptions about the future behavior of the levels of fertility and mortality (Central Bureau of Statistics, 2006). The medium variant estimates were used for this analysis.
- **Data on service utilization:** Data on ANC utilization by DHS region was available from the report for the 2006-07 Namibia DHS (Ministry of Health and Social Services (MoHSS) and Macro International Inc 2008). The report presents several ANC indicators relating to the provider, the place (i.e. at a facility or at home), the components and the number and timing of ANC visits. For simplicity, this example will only use rates of ANC utilization by provider. These are presented in columns 7-10 of Table 1. As can be seen from Table 1, Kunene region had the largest coverage gap with 18.2% of pregnant women not receiving ANC from a skilled provider (column 10). It is informative to note that this figure includes the 10.8% of women who received ANC but not from a skilled provider. This might suggest either that there is a barrier to accessing skilled providers in that region or that women are opting to use traditional birth attendants when giving birth. However, it would require further, more detailed quantitative and qualitative research in that region to determine whether the lack of skilled ANC providers is a significant barrier to utilization in Kunene.

Raster population layer: A raster dataset of the population density of women of childbearing age in Namibia by 100m^2 was available from the "WorldPop" website (Tatem and Linard 2013).

Gazetteer point data: A gazetteer dataset was available from the Humanitarian Response Web site giving the latitude and longitude coordinates of 607 settlements in Namibia. The attribute table included the name of each settlement but had no information about population (Humanitarian Response, 2002).

Question 3: Where Should the Government Target Future Investments in ANC Programs?

This section will consider two possible methods for determining where future investments in facility-based programs in Namibia should be made. These two methods correspond to the two types of geographic identifiers that health facility datasets may have – administrative unit identifiers and location identifiers - and are the two possible answers to question 1 in figure 1.

Mapping Accessibility Ratios

If data on health facilities is only available aggregated to administrative unit level, then accessibility ratios can be a useful, though crude, indicator of health service access. They express the number of health facilities in an administrative unit for every 1,000 eligible residents of that unit so the higher the value of the ratio, the greater the health service access. The regional ANC accessibility ratios in column 6 of Table 1 were calculated by dividing the number of health facilities offering ANC in each region (column 2) by the eligible population (column 4) and multiplying the result by 1,000.

$$
\begin{aligned}
&ANC\ accessibility\ ratio \\
&= (Number\ of\ health\ facilities\ offering\ ANC / Number\ of\ women\ aged\ 15-49) \\
&\times 1{,}000
\end{aligned}
$$

The resulting ratios (column 6 of table 1) were imported into a GIS as a table and *joined* to the region boundary layer along with the coverage gap statistics in column 10.

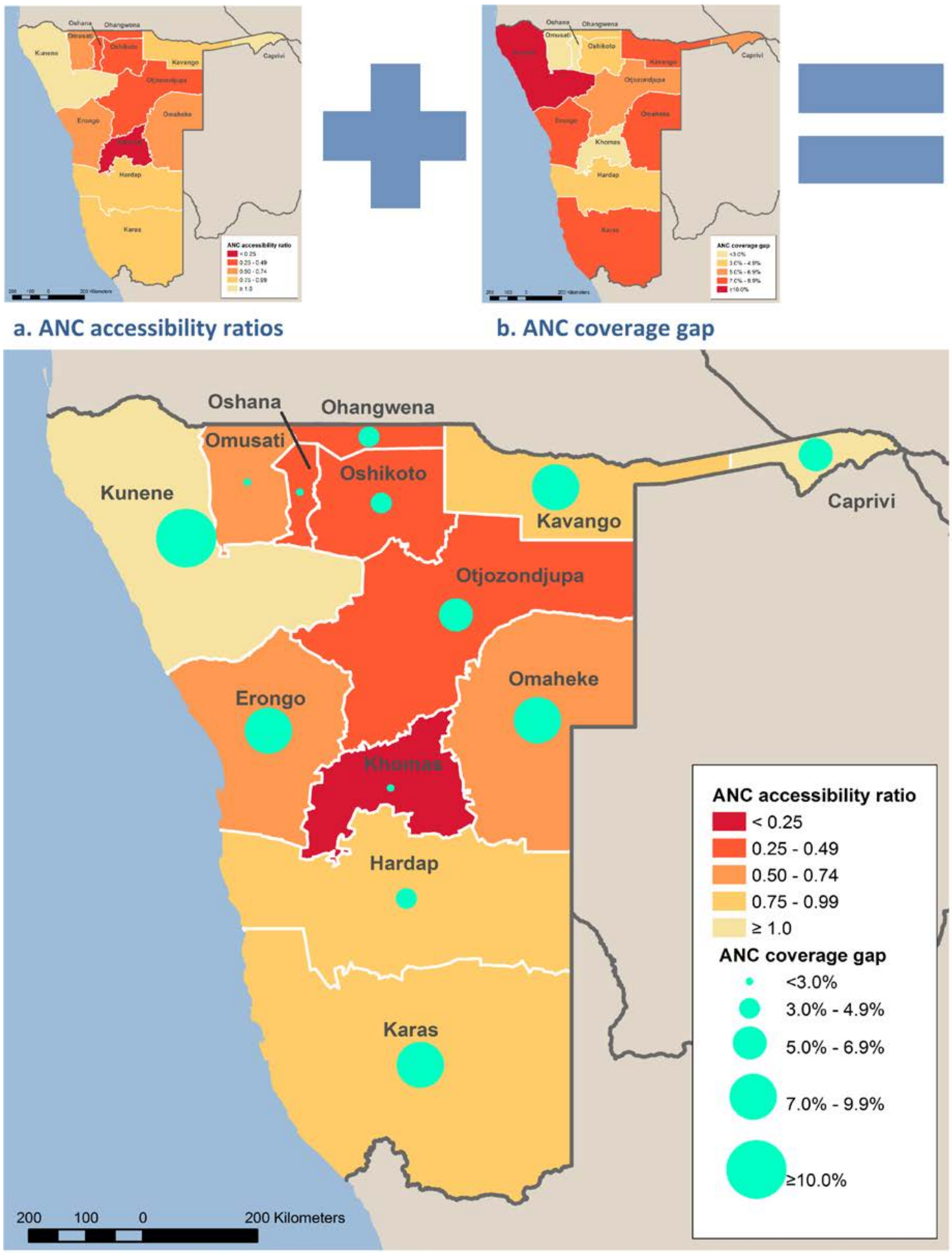
Data: The following data were used to map the regional level accessibility ratios for Namibia:

- ***Health facility data:*** The number of health facilities offering ANC by region (from Ministry of Health and Social Services, 2010).
- ***Eligible population data:*** Women of childbearing age – total female population aged 15-49 by region (from Central Bureau of Statistics, 2006, median variant population projections for 2011).
- ***Utilization data:*** Percentage of women who reported not receiving ANC from a skilled provider during their last pregnancy (from Ministry of Health and Social Services & Macro International Inc, 2008)

Summary: In the example map of Namibia (figure 4c), the polygon layer is shaded according to the ANC accessibility ratios given in column 6 of table 1. The utilization data (the ANC coverage gap) was mapped over the top of this as a second layer. The size of the superimposed circle symbols on the map corresponds to the percentage of women who reported not receiving ANC from a skilled provider during their last pregnancy.

Steps taken to map ANC accessibility ratios are the following:

1. Map accessibility ratios:
   a. Make a point layer of the coordinates of only those health facilities that offer ANC.
   b. Spatially join the coordinates to a polygon layer of the region boundaries to assign region identifiers to each facility.
   c. In a spreadsheet, compile data on total women of childbearing age with the number of health facilities per region to calculate accessibility ratios.
   d. Import the table of accessibility ratios into the GIS and join it to the polygon layer.
   e. Categorize the regions with a graduated color ramp based on their accessibility ratios (figure 4a).
2. Utilization/coverage gap data:
   a. Calculate coverage gaps for ANC by region in a spreadsheet.
   b. Import the table into the GIS and joined it to the polygon layer (figure 4b).
   c. Categorize the regions with a proportional symbol based on their ANC coverage gaps (Figure 4c).



c. *Thematic map* of ANC access and coverage

Figures 4: ANC accessibility ratios (a) and ANC coverage gaps are combined to show ANC access and cover (c).

Figure 5 shows how these steps relate to the route taken through the flowchart in this example.

Note: In this particular DHS, the survey regions are *coterminous* (they overlap precisely) with the administrative regions — the first level divisions — of Namibia, so it was not necessary to *apportion* or adjust any of the aggregated population statistics.

Findings from Using Aggregated Data to Measure Accessibility Ratios

The map produced by following the above steps shows that Khomas region, where the capital city of Windhoek is located, has the lowest ANC accessibility ratio, with just nine health facilities providing ANC to its 113,350 women of childbearing age. This suggests that it could be a potential candidate for targeting facility-based interventions. Khomas does, however, have one of the lowest ANC coverage gaps, with 2.8% of eligible DHS respondents reporting not having received ANC from a skilled provider during their last pregnancy. This might suggest that low levels of access are not a major barrier to utilization in that province or that, in the other provinces, different barriers to utilization are more significant. Kunene region has both a high accessibility ratio and low rates of utilization, suggesting that, in that area, the presence of ANC services is not in itself a strong determinant of utilization. Both of these hypotheses would require further analysis.

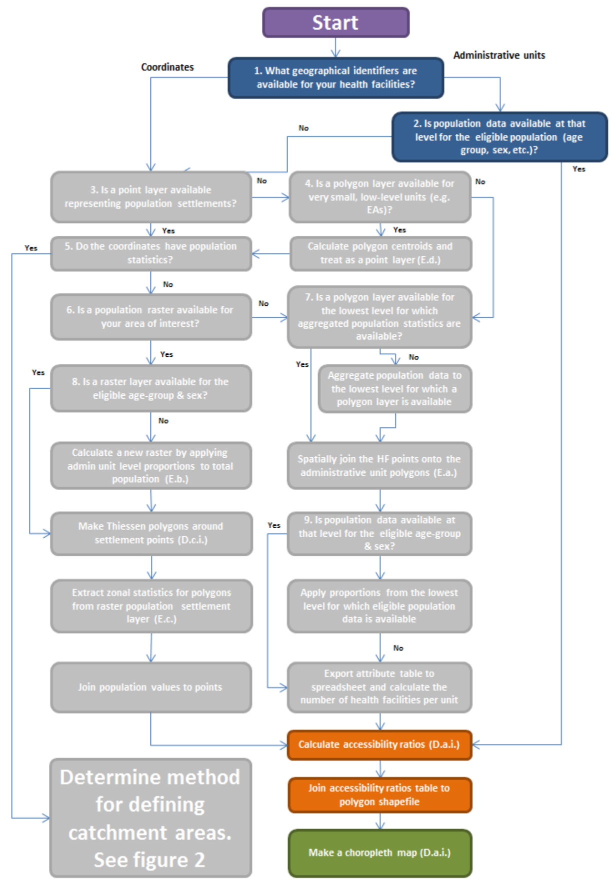


Figure 5: Route taken through the flowchart in this example.

One obvious limitation of using accessibility ratios to estimate access is that they do not take into account differences in the relative sizes of the units and the distribution of the population within them. One region may have fewer health facilities than another, but if it also has a much smaller area, then the average distance that its inhabitants have to travel to access them will be less. Similarly, if most of the population and the health facilities are concentrated in a single part of the region, then access is better there than in a region where both are evenly dispersed over a large area. However, both of these scenarios may result in the units having an equal accessibility ratio. This illustrates the fact that accessibility ratios are very crude indicators of true access and that analysis of aggregated statistics are subject to the *modifiable areal unit problem* (see annex D, section a.i). Because of this, it is recommended to conduct analysis only at the level of the administrative unit in two situations:

- to identify very broad trends or patterns to be verified by more detailed analysis; and
- if there are no more detailed data available.

Where coordinate and raster data are available, catchment area analyses are preferred over aggregated analysis.

Mapping Theoretical Catchment Areas

The first part of this illustrative example used aggregated data, which identified Khomas as the region with the lowest ANC accessibility ratio. With this information, it became possible to do a more detailed geospatial analysis of ANC access in that region using the health facility coordinates. These coordinates were used to map theoretical catchment areas — the geographic area from which a health facility draws clients. By knowing a facility's catchment area, it is possible to:

- link the underlying population to that facility; and
- identify the population that has access to that facility.

There are several methods for mapping theoretical catchment areas which are discussed in annex D, section c. The choice of method depends on the kind of data that are available as well as the kind of analysis being conducted; this is illustrated in figure 2. In this example, the aim is both to link the population to facilities and to define the population that lacks access to their linked facility (question 10 in figure 2). A categorical variable for access will be used (question 11 in figure 2). The analysis will be based on straight-line distance rather than network distance (question 12).

Thiessen polygons, shapes defining an area of influence around a point, can link a *population settlement* to a particular facility by assigning any given location within the area of interest to its nearest facility. *Buffers* on the other hand can determine whether that settlement's residents have access to that facility (where access is arbitrarily defined as being located within a threshold distance of a health facility). Buffers divide the population into two categories — those lacking access to a facility, and those with access to one or more facilities (since buffers may overlap). The two approaches (buffers and Thiessen polygons) can be combined in such a way that the population lacking access to a given facility would then be defined as those living within that facility's Thiessen polygon but outside its buffer (figure 6). To see the route taken through the flowchart in this example, see figure 7.

The following data were used to define theoretical catchment areas for Khomas region:

- **Health facility data:** Global positioning system (GPS) coordinates for the locations of health facilities offering ANC. Source – Namibia 2009 SPA (Ministry of Health and Social Services (MoHSS) [Namibia] and ICF Macro 2010)
- **Population data:**
  - Point layer of populated settlements from the Humanitarian Response Web site (Humanitarian Respose, 2002).
  - Population raster of women of childbearing age from the WorldPop project's demography Web site (Tatem & Linard, ND).

The following steps are taken to map theoretical catchment areas:

1. Create Thiessen polygons based on the health facility coordinates (figure 6.a).
2. Use zonal statistics to calculate the eligible population of each polygon from the raster layer (figure 6.b).
3. Join the population statistics to the polygon layer then categorize the polygons with a graduated color ramp based on their total eligible populations (figure 6.c).
4. Generate 30km buffers around the facilities that offer ANC (figure 6.d).
5. Import a point layer of settlements (figure 6.e).
6. Import a point layer of all health facilities with symbols categorized by whether or not they offer ANC (figure 6.f).

The map in figure 6.g shows buffers and Thiessen polygons generated based on the locations of the health facilities in Khomas region that offer ANC. The polygons are colored according to their eligible populations (extracted from a population raster using *zonal statistics* (see annex E, section c).

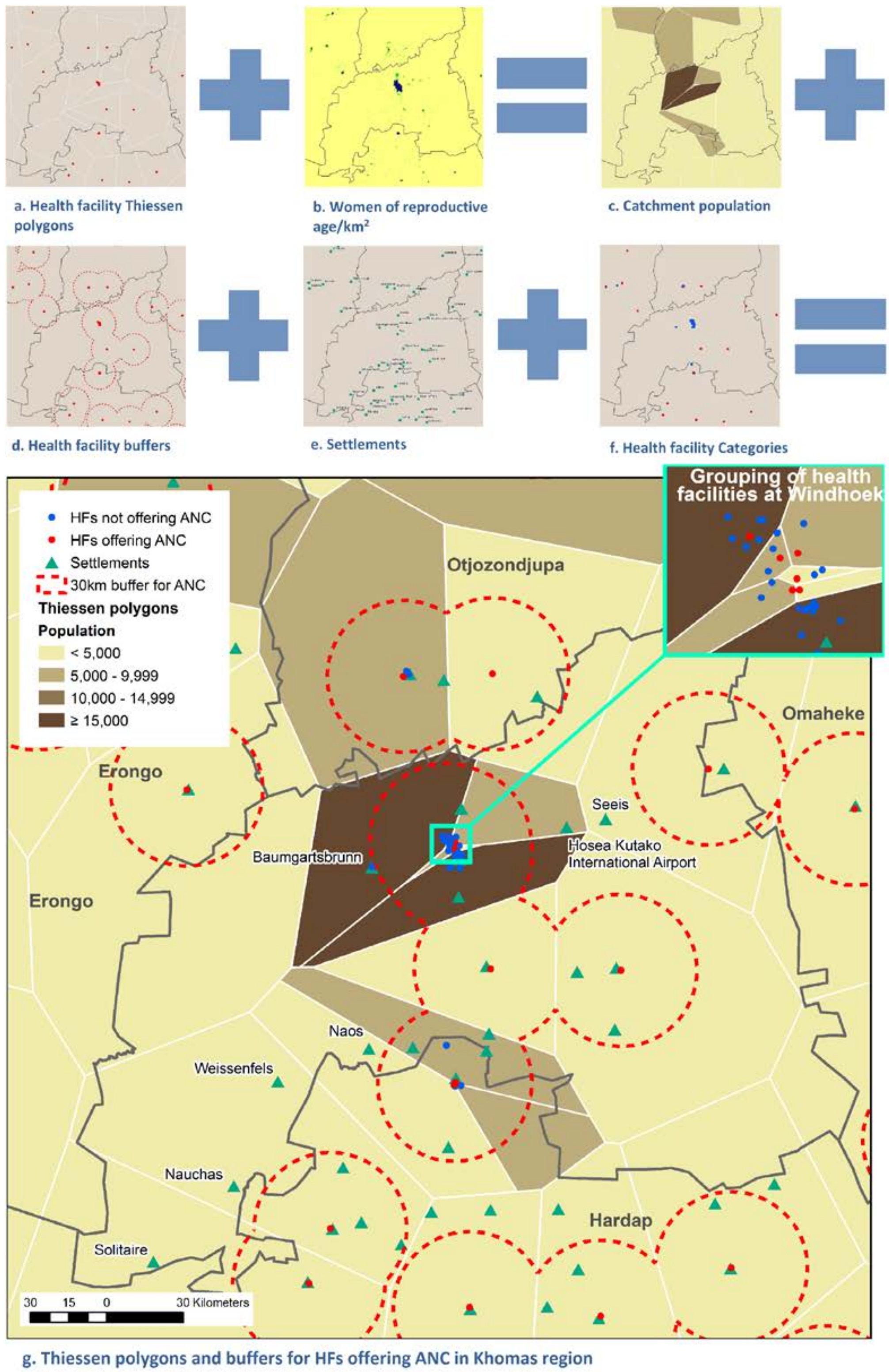


Figure 6: **Steps for creating Thiessen polygons and buffers for health facilities offering ANC in Khomas region.**

Findings from mapping theoretical catchment areas: The two dark Thiessen polygons in the center of the map (figure 6) contain a far greater eligible population than their neighbors. This means that their associated health facility may be serving a larger <u>*catchment population*</u>. The health facilities associated with these two polygons are in the capital city of Windhoek, which is located at the grouping of health facilities in the center of the map (labeled), many of which do not offer ANC.

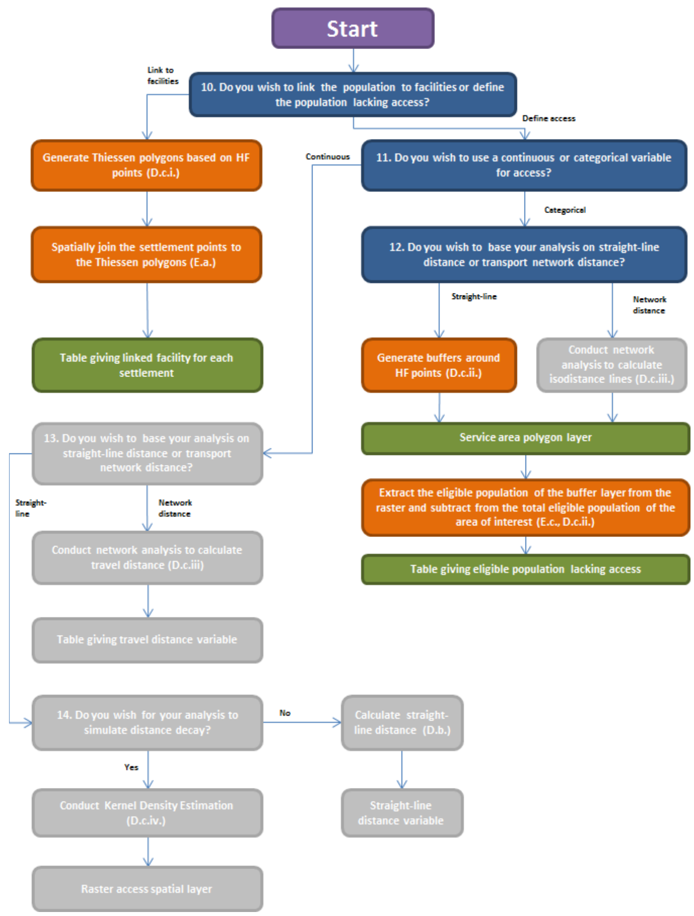


Figure 7: Route taken through the flowchart in this example.

The settlements that fall inside Khomas but outside of the 30km buffer of their linked health facility are labeled in figure 8. One settlement, Baumgartsbrunn, falls within the Thiessen polygon for one of the Windhoek facilities, but just outside its buffer. There is a health facility at this settlement that is not currently offering ANC and therefore might be a potential location for introducing ANC services. There are four settlements along the south western border of the region — Naos, Weissenfels, Nauchas and Solitaire — that fall within the Thiessen polygons of health facilities in Hardap region and outside the facility buffers. This area of low ANC access could also be the target for investments in new facilities offering ANC.

At this point it is worth noting the limitations of Thiessen polygons that are listed in annex D. For example, these polygons classify populations according to their direct distance proximity, but they do not reflect the actual route that clients have to take along the real transport networks in order to access the facilities.

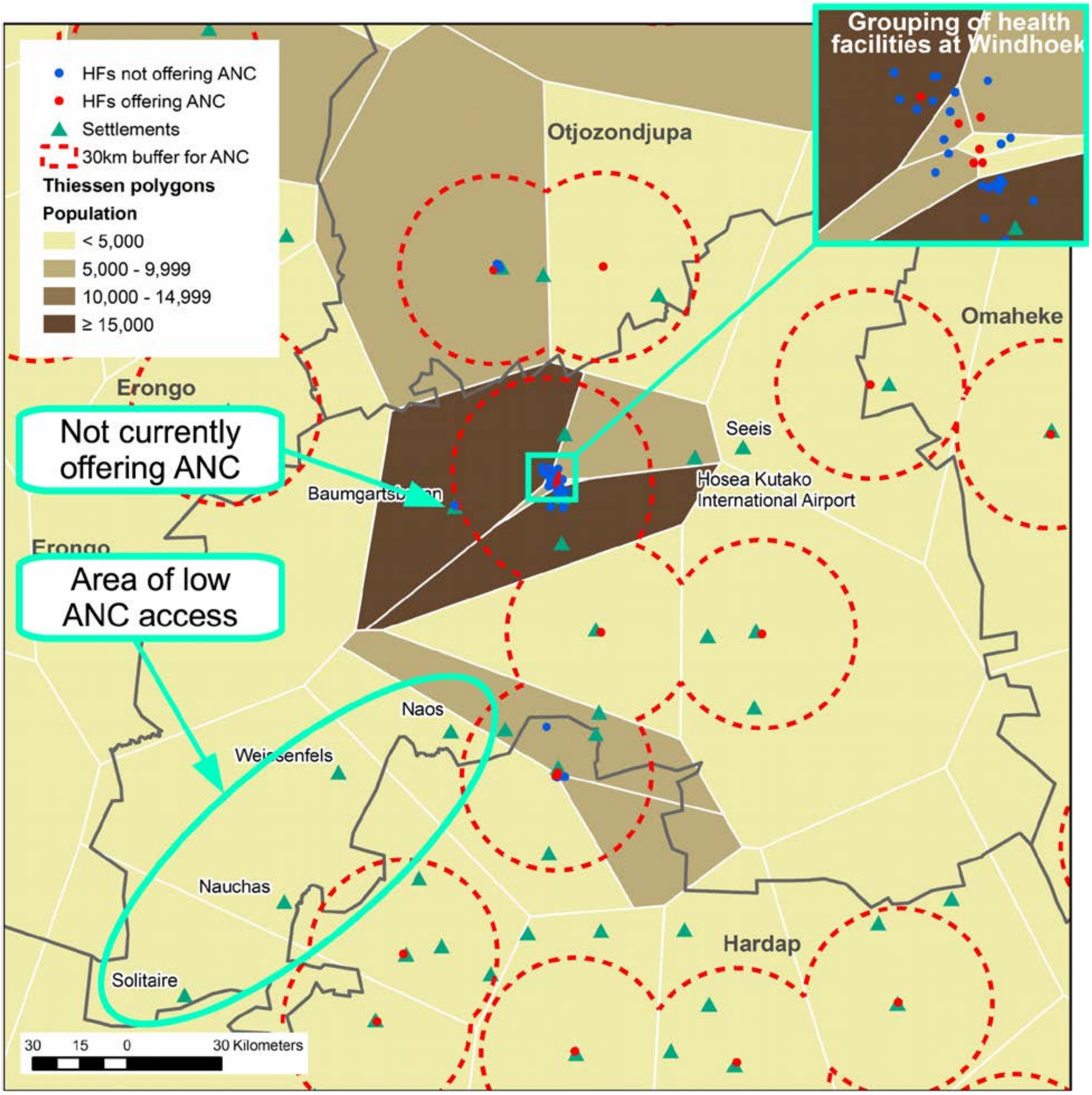
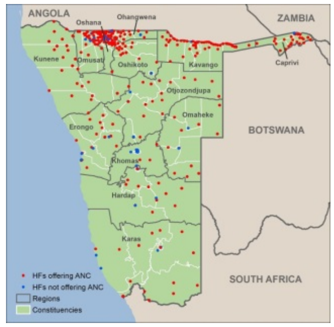


Figure 8: Areas for potential targeting of investments in ANC.

Where are existing services currently located?



Map facilities

Where is the eligible population that lack access to ANC?

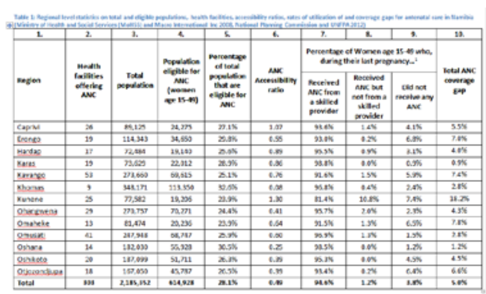
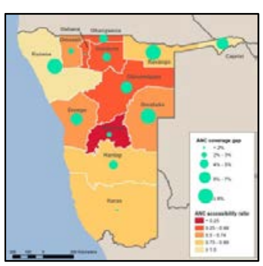


Table 1: Regional level statistics on total and eligible populations, health facilities, accessibility ratios, rates of utilization of and coverage gap for antenatal care in Namibia (Ministry of Health and Social Services (MoHSS) and Macro International Inc 2008, National Planning Commission and UNFPA 2012)

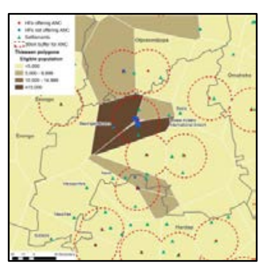
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Region	Health facilities offering ANC	Total population	Population eligible for ANC (women age 15-49)	Percentage of total population that are eligible for ANC	ANC Accessibility ratio	Percentage of Women age 15-49 who, during their last pregnancy...[1]			Total ANC coverage gap
						Received ANC from a skilled provider	Received ANC but not from a skilled provider	Did not receive any ANC	
Caprivi	26	89,125	24,275	27.1%	1.07	93.6%	1.4%	4.1%	5.5%
Erongo	19	114,343	34,650	25.8%	0.55	93.0%	0.2%	6.8%	7.0%
Hardap	17	72,484	19,140	25.6%	0.89	95.5%	0.9%	3.1%	4.0%
Karas	19	73,625	22,012	28.9%	0.86	98.8%	0.0%	0.9%	0.9%
Kavango	53	273,660	68,615	25.1%	0.76	91.6%	1.5%	5.9%	7.4%
Khomas	9	348,171	113,350	32.6%	0.08	95.8%	0.4%	2.4%	2.8%
Kunene	25	77,582	19,206	23.9%	1.30	81.4%	10.8%	7.4%	18.2%
Ohangwena	29	273,757	70,271	24.4%	0.41	95.7%	2.0%	2.3%	4.3%
Omaheke	13	81,474	20,236	23.9%	0.64	91.5%	1.3%	6.5%	7.8%
Omusati	41	287,948	68,747	25.9%	0.60	96.9%	1.3%	1.5%	2.8%
Oshana	14	182,030	55,928	30.5%	0.25	98.5%	0.0%	1.2%	1.2%
Oshikoto	20	187,099	51,711	26.3%	0.39	95.3%	0.0%	4.5%	4.5%
Otjozondjupa	18	157,650	45,787	26.5%	0.39	93.4%	0.2%	6.4%	6.6%
Total	308	2,185,352	614,928	28.1%	0.49	94.6%	1.2%	3.8%	5.0%

Population table and ANC utilization

Where is the eligible population that lack access to ANC?



Map accessibility



Map theoretical catchment areas

Figure 9: **Schematic of the main steps taken in the illustrative Namibia example.**

Conclusion

A geospatial analysis of health service access is prone to many inherent limitations. The limitations of some of the specific methods recommended in this document are discussed in annex D. One of the general limitations of using GIS in this way is that lack of physical access to a facility offering a health service is just one of many barriers to utilizing that service. There may be numerous reasons why a pregnant woman who lives close to a health facility offering ANC may fail to utilize that service, including cost and quality. Utilization data from household surveys can suggest other possible barriers; however quantitative data can only go so far. To identify other barriers it would be necessary to use qualitative research methods such as focus groups or in-depth interviews with potential clients.

Aspects of the type of service under analysis may influence how accurately access can be estimated. ANC is a non-emergency service that has little stigma attached to its use. Consider the case of voluntary counseling and testing for HIV, which is also often a health facility-based service, but may have considerable stigma attached to its use. Clients may choose not to go to their nearest facility offering that service in order to avoid being seen by their neighbors. If this is the case, then it may not be appropriate to link population to facilities based on geographic proximity. Also, clients may be prepared to travel further to utilize non-emergency services — such as ANC or family planning — if they know that the facility will offer better quality of service. For an emergency service such as obstetric care or treatment for malaria, accessibility may be a more pressing concern than quality. These factors would influence the size of the buffer that would need to be used to define the population lacking access.

The decision-making process outlined here makes no assumptions about the quality or accuracy of the data being used. There are numerous factors that can affect this. Administrative units can be renamed or split into two rendering administrative unit classifications that are out of date. Coordinates may be truncated to a certain number of decimal places, which may introduce errors in accuracy. Data quality and accuracy is something that the user will need to determine before starting the analysis. A forthcoming MEASURE Evaluation document will present a framework for GIS decision making that includes issues of data quality and accuracy (Moise, Cunningham & Inglis, ND).

Despite these limitations, geospatial analysis is an extremely informative tool for identifying gaps in access, coverage, and utilization. In this illustrative example, data aggregated at high-level administrative units (regions) was used to broadly identify areas of low access to ANC. Then, focusing in on one area, detailed coordinate data were used to generate theoretical catchment areas and link population settlements to particular facilities. The settlements were then categorized as either having or lacking access to their linked facility. Based on this exploratory analysis it was possible to identify several settlements that might benefit from investment in new health facilities, and one health facility that might better serve its catchment population if it were allocated resources to offer ANC. This example shows that geospatial data and analysis methods can reveal a great deal about access to health services, but only if these limitations and considerations are taken into account when interpreting the results.

Annex A: Glossary

Access	The physical presence of a facility offering that service within a given distance or catchment area of its potential clients.
Accessibility	The distance or travel-time (the travel impedance) between the client's location and the facility (Guagliardo 2004). It is just one of the things that may prevent an eligible individual from utilizing a service.
Accessibility ratio	The number of health facilities in an administrative unit divided by the eligible population of that unit. Often it as expressed as facilities per 1,000 inhabitants.
Administrative unit	A defined area of a country used for the purposes of government or administration, such as a state, county, province or district.
Aggregated statistics	Data that is summed and presented as totals for each of a list of administrative units.
Apportionment/apportion	A geospatial method for splitting up statistics (such as population totals) from one polygon according to the fraction of another, non-coterminous polygon that falls within the first polygon.
Area of interest	The geographic extent of the geospatial analysis.
Attribute table	A table attached to a set of geographic features, in which the rows represent each feature in the set, and the columns represents some attributes of those features such as names, area or categories or the result of a field calculation.
Boundary polygon	A closed shape, the border of which represent the boundaries of adjacent political entities such as countries or districts.
Buffers	Circular polygons with a point at the center and with a radius defined by the buffer distance from that point.
Catchment area	The geographic area from which a health facility attracts clients.
Catchment population	The population of a health facility's catchment area.
Centroid	The geographic center point of a polygon or other feature.
Choropleth map	A thematic map in which administrative areas are colored or shaded according to the range in which the aggregated statistic of interest falls.

Contour — A line that connects points of an equal value (e.g., of a raster).

Coterminous — Two administrative units or polygons are coterminous if their boundaries align and they overlap precisely.

Coverage — The proportion of people that are eligible to receive an intervention or utilize a service that actually do receive or utilize it.

Coverage gap — The proportion of people that are eligible to receive an intervention or utilize a service that do not receive or utilize it.

Distance decay — A term that describes the effect of distance on spatial interactions. The interaction between two locations declines the greater the distance between them. For example, the price of properties may decrease the further they are from the center of a city.

Eligible/eligibility — The eligible population is those individuals that have the capacity to benefit from an intervention. Eligibility can be defined according to a norm (e.g. women of childbearing age are eligible for contraceptive counseling) or by virtue of having a certain health condition (e.g. children with diarrhea are eligible for oral rehydration therapy).

Enumeration area — The lowest level area defined for the purposes of enumerating population during a census. An enumeration area (EA) can be a city block or apartment building in urban areas, while in rural areas it is typically a village or group of villages.

Gazetteer — A list of geographic place-names and their coordinates.

Geographic identifier — A field in a table or database that gives the geographic location of a record. This could be the name or code of an administrative unit, or the latitude and longitude values for the exact location.

Grid/gridded data — See raster.

Draw (of a health facility) — The extent to which a health facility attracts clients from the population it serves.

Health system performance — The degree to which a health system carries out its functions to achieve its goals.

Impact — The net improvement in population health status that can be attributed to an intervention.

Joined/joining data	Merging two data sets based on a common field or identifier value.
Location identifier	An attribute in a data set that gives the specific location of a feature such as a health facility or settlement (as opposed to just the administrative unit in which it is located). A location identifier can be an address or a set of coordinates.
Modifiable areal unit problem	A source of bias in analyses performed on aggregated data. Administrative unit level statistics depend on how the boundaries of the unit happen to be drawn. Since the boundaries of administrative units rarely correspond to real differences in the attributes of their population, aggregated analysis cannot pick up smaller scale patterns or variations in the variable of interest.
Polyline	A feature in a GIS that has length but not area and often represents a long, thin geographic feature, such as a river or road.
Population settlement	A populated place such as a community, village, town, or city.
Raster	A gridded surface layer for a specified variable in a GIS that assigns a value to every grid square.
Service area	An area, the extent of which is determined by a specified travel impedance (i.e., time or distance) along a transport network.
Settlement	A discrete, populated place such as a village, town, or city.
Spatial join	A basic GIS tool that takes data from a polygon layer and assigns it to a point layer based on the spatial location of the points.
Readiness score	An index that assigns a numerical value to a health facility that corresponds to how well-equipped the facility is to deliver a certain service. The index is calculated based on health facility assessment data.
Shapefiles	A GIS file format for storing data on the location, shape, and attributes of geographic features.
Spider-diagram	A diagram generated by drawing lines connecting points in one layer to their linked points in another layer. These diagrams resemble spiders because the lines radiate out of a central point. The method is also known as “desire-line analysis”.
Thematic map	A map created to display information about a topic or theme.

Thiessen polygons	Polygons generated from a set of sample points. Each Thiessen polygon defines an area of influence around its sample point, so that any location inside the polygon is closer to that point than any of the other sample points.
Utilize/utilization	Actual attained use of a service by a client.
Zonal statistics	A GIS tool that calculates statistics based on the values of a raster layer for each polygon within a polygon layer. The statistics include among others a minimum, maximum, mean, and median.

Annex B: Health Service Environment Data

This annex summarizes important considerations regarding data about the health service environment. These include the type of geographic identifier that the data has, the health facility attributes that are included in the data, and data sources.

a. Geographic Identifiers

There are two types of geographic identifier that health data may include and that can be used to inform an analysis:

- **Administrative unit identifiers:** Countries tend to be divided up into smaller areas for the purposes of government or administration. There can be multiple levels of division, with the smaller (low level) units being aggregated up into the larger (higher level) units. Table 2 illustrates this idea by presenting the administrative hierarchy of Namibia. Namibia is divided at the first level into 14 regions, which in turn are divided into 121 constituencies. The lowest level administrative unit in Namibia is the *enumeration area*.

Table 2: Administrative Hierarchy of Namibia

Administrative Level	Divisions	Number of Units	Level
0	National	1	High
1	Region	14	…
2	Constituency	121	…
3	Enumeration area	6,103	Low

Health facility data sets with administrative unit identifiers could come in the following forms:

a. Identifiers could be in a table with health facility names or identification codes in one column and the subnational administrative unit (such as the state, region, district or province) in which it is located in another column. An administrative unit identifier can be a name or an official identification code. Table 3 is an example:

Table 3: Example of Identifiers in a Table

Facility ID	Administrative unit name	Administrative unit code
479	Oshana	WA37
355	Kunene	WA32
150	Erongo	WA29
327	Khomas	WA21
364	Kunene	WA32

b. Or a table could provide the administrative units in one column and the number of health facilities located in each unit in another column. Table 4 is an example:

Table 4: Example of Number of Facilities by Administrative Unit

Administrative Unit Name	Number of Facilities Offering Service
Kunene	25
Caprivi	26
Hardap	17

Location identifiers – coordinates: A table could give the latitude and longitude values for the location of each facility. These could be the exact coordinates of each facility taken using a global positioning system device or the coordinates of a city or settlement where the facility is known to be located, taken from a gazetteer.

b. Health Facility Attributes

In addition to geographic location, other attributes of the health facilities can be included in the analysis. For example:

1. **Service availability:** Information on which facilities are currently providing a particular service may be available from health facility censuses and/or surveys.
2. **Type of facility:** Health facilities vary in scale from community dispensaries to national hospitals; they may be at primary, secondary or tertiary level in the referral network; and they can be private or public, all of which will influence the *draw* that they exert on their potential clients (see annex D, section c.iv) and, if the data are available, can be taken into account to add more detail to an analysis.
3. ***Readiness scores:*** can be calculated for specific services (e.g., family planning, voluntary counseling and testing) based on health facility assessment data. These assign a numerical value to each facility corresponding to how well equipped they are to deliver the service (Wang, Wang, Pullum & Ametepi, 2012).
4. **Other:** Other relevant information might include the number of inpatient beds per facility, staffing (e.g., number of doctors, nurses, etc.), and equipment or pharmacy stock-outs.

c. Sources of Geo-referenced Health Service Data

There are a number of organizations that make available to the public geo-referenced data on health services, including the following:

1. **The DHS Program:** The DHS Program's Service Provision Assessments (SPA) surveys include detailed facility-based surveys on nationally and regionally representative samples of health facilities. The final reports present the findings broken down by survey region. The GPS coordinates of the surveyed facilities are available for download upon request from the Web site http://www.dhsprogram.com/data/available-datasets.cfm, for SPAs since 2009. The boundaries of the survey regions at which the results of the SPAs are representative can be downloaded as shapefiles from the DHS Spatial Data Repository http://spatialdata.dhsprogram.com/boundaries.html

2. **Humanitarian Response:** This organization's Web site has point data of health facilities for various countries including Nepal, Burkina Faso, Senegal and Jordan. They vary in their completeness. https://cod.humanitarianresponse.info/search/field_data_type/points-249?search_api_views_fulltext=
3. **The World Health Organization (WHO):** The WHO's Geonetwork Unit gives users access to geo-referenced databases, cartographic products and related metadata from a variety of sources http://apps.who.int/geonetwork/srv/en/main.home. The WHO's Service Availability and Readiness Assessment (SARA) and Service Availability Mapping projects use specially designed tools to assess and monitor the service availability and readiness of national health sectors and to generate evidence to support the planning and managing of a health system. Data are made available in reports at http://www.who.int/healthinfo/systems/sara_introduction/en/.
4. **National government agencies:** Web sites of the ministries of health or national statistical authorities of individual countries may make health statistics or routine health information system data available to the public.

Annex C: Population Data

There are two types of population-based data that can inform an analysis: data on the geographic distribution and age and sex structure of the population; and data on the utilization of services by the population.

a. Data on Population Distribution and Structure

Georeferenced data on population distribution and structure can come in three formats:

i. Aggregated Population Data

Aggregated data on population is most commonly presented as sum totals of the inhabitants of a country, region or district.

Description: Aggregated population data is summed and presented as the total number of inhabitants of each of a list of administrative units. The most complete sources of aggregated population statistics are censuses. Census reports often summarize population statistics for subnational administrative units, which can be joined in a GIS to a *boundary polygon*.

Advantages:

- **Completeness:** Censuses aim to give a complete picture of the population and age- and sex-structure of a country.

Limitations:

- **Timing:** Censuses tend to take place at intervals of 10 years or more so population estimates may not be up-to-date and may have to be adjusted.
- **Level of detail:** Often published data is not broken down into age, sex or other groupings at a low enough administrative level. In these cases, it may be necessary to apply the proportion of the population in a particular group at a higher administrative level (such as national level) to the total population at lower levels to arrive at estimates. This approach makes assumptions about the distribution of the population and its age structure that may introduce misclassification error or bias into the analysis.

Sources: The first place to look for census data or population projections for any given country is the website of that country's statistical authority. In addition, the Integrated Public Use Microdata Series, International project is an online repository of census microdata available for social and economic research (https://international.ipums.org/international/). The website also has available first level GIS boundary polygon files.

ii. Raster Population Layers

Description: Population rasters are continuous *gridded* surface layers that assign an estimated population density value to every square in their grid. Various gridded population layers are described and compared in an article by Tatem and others (Tatem et al., 2012).

Advantages:

- **Detail:** The detail and versatility of raster population data sets make them a valuable resource for geospatial health research. High resolution rasters can give estimates down to the level of the $500m^2$ pixel. They can be used with point data to extract population density statistics at a specific location, or with polygons to extract average or summary zonal statistics (see annex E, section c).

Limitations:

- **Aggregating raster data:** Raster population layers are intended to give accurate estimates for particular locations, small zones, or low level administrative units such as wards or EAs. Aggregating the raster data up to a higher level administrative unit (e.g. using zonal statistics), will not necessarily give results that match aggregated census data from other sources.
- **Age/sex composition:** Raster population data is not usually broken down by age or sex composition, although the WorldPop project started to address this in 2013. Annex E, section b, discusses how to adjust a population raster by age-structure statistics.
- **Uncertainty:** Population estimates derived from rasters vary in their degrees of uncertainty, depending on the resolution of the raster and the data that was used as inputs into the model. Put another way, raster population layers are only as good as their inputs. In areas where the input data, is incomplete or out of data, the population distribution predicted by the raster may be unreliable.

Sources: Several Web sites make population data available as rasters.

- **WorldPop:** The WorldPop project has detailed, high-resolution population density maps for the whole of Africa, Asia, and Latin America and the Caribbean. The maps combine satellite land cover and settlement imagery to reallocate contemporary census-based spatial population count data. The project also aims to provide similar data sets showing age composition and five-year groupings and gender (Tatem et al., 2014). So far, they have children under the age of five and women of childbearing age for 2000, 2005, 2010 and 2015: http://www.worldpop.org.uk/
- **LandScan**: These raster layers have a finer resolution and are adjusted to Central Intelligence Agency estimates: http://web.ornl.gov/sci/landscan/
- **Gridded Population of the World (GPW):** GPW data is adjusted to the United Nations Development Programme (UNDP) estimates: (http://sedac.ciesin.columbia.edu/data/collection/gpw-v3).

- **Global Rural Urban Mapping Project (GRUMP):** This project's Web site offers finer resolution layers compared to GPW that include urban extents and are adjusted to UNDP estimates: http://sedac.ciesin.columbia.edu/data/collection/grump-v1.

iii. Gazetteer Point Data

Description: Gazetteer data sets are tables or point shapefiles giving the coordinates of settlements and populated places. A low-level administrative unit polygon layer with population data in its attribute table (such as census enumeration areas) can be converted to a point dataset by taking the coordinates of the polygon centroids (see annex E, section d).

Advantages:

- **Precision:** The advantage of working with coordinates is that they represent exact locations of settlements and thus lend themselves to finer-grained spatial analyses.

Limitations:

- **Geographic area of settlement is unknown:** Because the settlement is represented by a single point, it is not possible to judge its true size.
- **Lack of population statistics:** Gazetteer data sets do not always include population statistics in their attribute tables, and so must be joined with data from other sources. One approach is to generate Thiessen Polygons around the points (see annex D, section c.i), extract zonal population statistics from a raster population layer (see annex E, section c), and join the resulting table to the point layer's attribute table.

Sources: There are several Web sites that make gazetteers of settlements available, including:

- Humanitarian Response initiative https://cod.humanitarianresponse.info/
- DIVA-GIS http://www.diva-gis.org/Data
- National Geospatial-Intelligence Agency GEOnet Names Server http://earth-info.nga.mil/gns/html/namefiles.htm

b. Data on Service Utilization

Information on the extent to which eligible individuals are actually utilizing the services can also be important in analysis.

Description: The rate of utilization of a particular service refers to the proportion of the eligible population that actually used the service. Box 3 gives some examples of indicators of low service utilization. The main sources of data on these indicators for developing countries are household surveys, although routine health information systems may also serve as a source of utilization data.

Advantages:

- **Comparing access with utilization:** The advantage of bringing utilization data into an analysis is that data on access can be compared with rates of actual service-use to see if the two are correlated (see the Namibia example section). If access is high but utilization is low, this might suggest that factors other than the physical presence of the service are influencing whether the service gets used.

Limitations:

- **Availability:** Utilization data are usually only available for a sample of the population.
- **Representativeness:** Due to the constraints of sample sizes, statistics from household surveys are usually only representative at the level of high level administrative units.

Examples of indicators of low utilization of health facility-based surveys include percentage of:

- *currently married women with unmet need for family planning*
- *women who did not receive antenatal care during their most recent pregnancy*
- *women who did not deliver at a health facility for their most recent pregnancy*
- *women whose delivery was not attended by a doctor or other health professional for their most recent pregnancy*
- *children who are not fully vaccinations for their age*

Sources: The main source of data on utilization of health services are population-based surveys such as:

- **Demographic and Health Surveys (DHS):** Data from DHS (as well as AIDS indicator Surveys (AIS) and Malaria Indicator Surveys (MIS)) are available as aggregated polygon shapefiles from the Spatial Data Repository http://spatialdata.dhsprogram.com/. Note that the lowest subnational level at which DHS data is representative is the survey region - subnational geographic entities that are defined for the purposes of the survey sample design. In some cases these correspond to existing administrative units used by the country. In others they are agglomerations of administrative units. DHS regions do not necessarily correspond to the SPA regions or to the first level administrative units of a given country. The displaced GPS coordinates of the surveyed cluster center-points can be request through an online system on this Web site:
  http://www.dhsprogram.com/data/available-datasets.cfm.
  To ensure respondent confidentiality the cluster coordinates from DHS household surveys are displaced. This means that they are randomly "shifted" to a new location under set parameters before the data is publically released. Urban clusters are displaced in a random direction and distance up to 2km. For rural clusters the maximum displacement distance is 5km with a further, randomly selected 1% of the rural clusters displaced a distance up to 10 km.
  More information can be found in these two publications in the DHS Spatial Analysis Reports series:

  Burgert C, Colston J, Roy T, Zachary B. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. In

DHS Spatial Analysis Reports. Calverton, MD: ICF International; 2013. http://dhsprogram.com/publications/publication-SAR7-Spatial-Analysis-Reports.cfm

Perez-Heydrich C., Warren J, C. Burgert C, and M. Emch M. Guidelines on the Use of DHS GPS Data. In *Spatial Analysis Report*. Calverton, MD: ICF International; 2013. http://dhsprogram.com/publications/publication-SAR8-Spatial-Analysis-Reports.cfm

- **Multiple Indicator Cluster Surveys:** The United Nations Children's Fund's MICS surveys collect similar data to the DHSs, however, they do not currently release any geographic identifiers below the region level: http://www.childinfo.org/mics_available.html
- **Living Standards Measurement Surveys (LSMS):** The World Bank releases georeferenced data from these household surveys: http://iresearch.worldbank.org/lsms/lsmssurveyFinder.htm
- **Other:** Surveys of particular countries or regions are sometimes carried out by national or state governments or local nongovernment organizations. Web sites of individual agencies should be consulted to determine the availability of data.

Annex D: Geospatial Methods for Deciding Where to Allocate Resources

This annex describes the types of geospatial methods that can be used to map service access. These methods are organized into three groups:

a. administrative unit-level or aggregated analysis;
b. straight-line distance methods; and
c. methods that involve defining catchment areas.

The types of analysis that are possible to perform with linked health facility and population data depend on the type of geographic identifier that the health facility data has and the format of the population data.

a. Administrative Unit-Level or Aggregated Analysis

Basic geospatial analyses can be carried out at the level of the administrative unit. According to this method, the population living within a particular administrative division is defined as having access to any health facility located within the same unit – a method that has also been referred to as "administrative boundary link" (Skiles, Burgert, Curtis & Spener, 2013).

i. Accessibility Ratios

Description: A unit's accessibility ratio is the number of health facilities in that unit divided by its eligible population. Often it as expressed as facilities per 1,000 inhabitants (Spencer & Angeles, 2007).

Data requirements:

- total eligible population of each administrative unit (e.g. the number of women of childbearing age per province)
- number of health facilities offering the service of interest located in each unit

Steps:

1. In a spreadsheet, calculate the accessibility ratios by dividing the number of health facilities by the eligible population
2. Join the column of ratios to the attribute table of the polygon shapefile
3. Create a *choropleth map (thematic map)*

Outcome: The ratios can be visualized as a choropleth map, in which the units are shaded according to a color progression that corresponds to the ratio, as shown in the Namibia example section. A second variable can be mapped over the top by assigning a symbol whose size is proportional to the other indicator. Thematic mapping of this kind is an example of exploratory spatial data analysis, which can be used to identify or visualize spatial patterns for further analysis. By itself, however, thematic mapping cannot be used for confirmatory analysis or to test hypotheses (Yao, Murary, Agadjanian & Hayford, 2012).

Targeting: Units with the lowest accessibility ratios can be targeted for investment of resources in scaling up services. This can be done outside of a GIS, but to add a geographic component to the analysis, the accessibility ratios must be joined to the attribute table of a polygon shapefile of the administrative unit.

Advantages:

- **Simplicity:** The advantage of working with aggregated statistics is that with very basic, easily available data, one can generate a crude estimation of access and identify areas of a country in which to do more detailed analysis.

Limitations:

- **Requires coterminous boundaries:** In order to join the three datasets — the population, the health facilities and the polygon layer — the administrative unit's boundaries in each data set must correspond at the same administrative level (they must be coterminous). If the boundaries of the areas do not match up, then apportionment methods can be used within a GIS to split up the population according to the fraction of each population unit that falls within each health facility area (Gorr & Kurland, 2011). However, apportioning data in this way is prone to errors since it assumes that the population is uniformly distributed across each polygon. Apportionment is not recommended for large polygons but only for detailed, low-level administrative divisions.
- **Modifiable areal unit problem:** When a spatial phenomenon is aggregated to the level of an administrative unit, bias is introduced to the analysis due to the modifiable areal unit problem. The resulting statistics will depend on how the boundaries of the unit happen to be drawn. Since the boundaries of administrative units rarely correspond to real differences in the attributes of their population, aggregated analysis cannot pick up smaller scale patterns or variations in the variable of interest. Furthermore, access is not evenly spread across a region or district. A person living close to the border of an administrative unit may utilize a service in the adjacent unit. Generally, the smaller the unit, the less prone the analysis will be to this problem, but it can never be fully eliminated.

Further resources:

Cote P. 2013. Effective cartography: mapping with quantitative data. Cambridge, MA: Harvard University Graduate School of Design; 2013. Available from: http://www.gsd.harvard.edu/gis/manual/normalize/.

Dramowicz E, Dramowicz K. Choropleth mapping with exploratory data analysis 2004 [Web page]. Available from: http://www.directionsmag.com/articles/choropleth-mapping-with-exploratory-data-analysis/123579.

Skiles M, Burgert C, Curtis S, Spencer J. 2013. Geographically linking population and facility surveys: methodological considerations. *Popn Health Metrics.* 2013. 11(1):14

Yao J, Murray AT, Agadjanian V, Hayford SR. Geographic influences on sexual and reproductive health service utilization in rural Mozambique. *Applied Geog.* 2012. 32(2):601-607. DOI: 10.1016/j.apgeog.2011.07.009.

b. Straight-Line Distance Methods

Some studies have shown that one of the main factors that determines how likely an eligible individual is to utilize a health service is their geographic proximity to a health facility (Al-Taiar et al. 2010, Yao et al. 2012). Therefore, a potential indicator for health service accessibility could be the distance in a straight-line between a population settlement and a linked health facility. It is often convenient to define this linkage in terms of distance, so that a settlement is linked to whichever health facility is nearest to it in a straight line. However, linkage can be defined in other ways. For example, if in a survey, a respondent has specified a particular health facility as one that they actually utilize, then the linkage can be defined based on reported actual usage, even though this will mean that not all the population is linked to their geographically closest facility (Noor et al. 2003).

Description: In a GIS, a set of points representing population settlements can be assigned a variable corresponding to each point's distance from its linked facility (nearest or otherwise defined). The linkage can be mapped visually using Spider-diagrams. These diagrams are illustrated in figure 10. At the center of each "spider" is a point representing a health facility, while the "legs" represent the shortest distance from the facility to its linked settlements. These diagrams are useful visual tools as it is easy to identify long lines which represent settlements with low access. However, spider-diagrams are not necessary for the purposes of analysis, since the distance variable can be exported to a spreadsheet and analyzed using non-spatial analysis techniques.

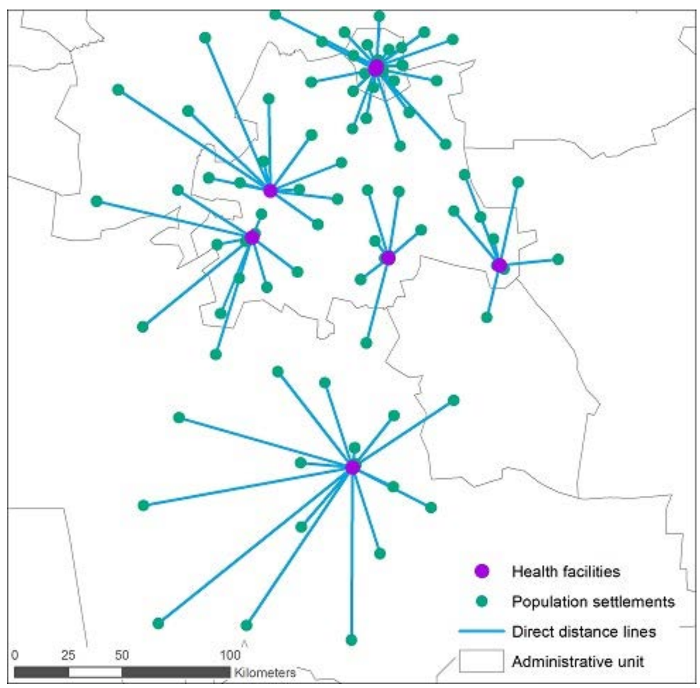


Figure 10: Illustrative example of Spider-diagrams representing the straight-line distances from population settlements to health facilities.

Data requirements:

- a point layer representing settlement locations
- a point layer representing health facility locations

Steps:

1. In a GIS, use a tool (such as "Near (Analysis)" in ArcGIS) to calculate the distance from each settlement location to its nearest health facility. This will generate a second point layer with a field in its attribute table giving the distance from each settlement to its nearest facility.
2. Assign graduated colors to points from the output layer to show the value of their distance variable. Alternatively use a spider diagram tool ("Spider Diagram — Spatial Location" in ETGeoWizards, "Create spider diagram (desire lines)" in ArcGIS) to create a layer of *polylines* representing the shortest direct routes from settlements to their nearest facility.
3. Export the attribute table of the new layer to a spreadsheet in order to analyze the distance variable.

Outcome:

- **Distance analysis:** A point layer representing settlements with a field in its attribute table giving the distance from each settlement to its nearest or linked facility. This distance can be treated as a continuous variable, to compare averages between facilities or settlements grouped by certain characteristics. Alternatively, threshold distance can be specified to divide the population into binary categories: those with access to the service (below the threshold distance) and those lacking access.
- **Spider diagram:** A polyline layer representing a spider-diagram of the shortest direct routes from settlements to their nearest facility. The layer will also have the distance field in its attribute table.

Targeting: Distance is used as an indicator of access, with longer distances representing lower access. Therefore investments in new services can be targeted to areas that have settlements with higher distance variables. Once the distance variable has been generated, it can be analyzed outside a GIS. However, coloring settlement points in a GIS according to their distance variable can help to visually identify pockets of low access that could be candidate locations for scaling up services.

Advantage:

- **Simplicity:** Straight-line distance is simple to calculate and straightforward and intuitive to interpret and analyze as an indicator of access.

Limitations:

- **Does not reflect true travel distance:** Geographic proximity can only approximate actual physical access. It assumes that clients can travel along a straight-line path to get to a service. In reality, this is rarely possible in the real world where they must take indirect routes to avoid geographic obstacles (e.g. rivers, mountains) and use the existing transport network. However, in at least one study based on data from Yemen, straight-line distances were found to be strongly correlated with actual driving distances and times, so in some settings, it may be a reliable proxy (Al-Taiar, Clark, Longnecker & Whitty, 2010).
- **Does not reflect actual usage:** Linking clients to their nearest facility falsely assumes that proximity is the only determinant of utilization, whereas, in fact, a variety of other factors (such as service quality and cost) will affect a person's choice to utilize a certain health facility rather than another or none at all (Burgert et al. 2011). However, one study in Kenya compared access defined by theoretical linkage with access defined by actual usage as reported in a survey and found that the two correlated well (Noor, Zurovac, Hay, Ochola & Snow, 2003).

Further resources:

Al-Taiar A, Clark A, Longenecker JC, Whitty CJ. 2010. Physical accessibility and utilization of health services in Yemen. *Internl J Health Geog.* 2010. 9:38. DOI: 10.1186/1476-072x-9-38.

Noor AM, Gikandi PW, Hay SI, Muga RO, Snow RW. 2004. Creating spatially defined databases for equitable health service planning in low-income countries: the example of Kenya. *Acta Tropica.* 2004. 91 (3):239-51.DOI: 10.1016/j.actatropica.2004.05.003.

Noor AM, Zurovac D, Hay SI, Ochola SA, Snow RW. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Trop Med Intern Health.* 2003. 8(10):917-26.

Noor A, Alegana V, Gething P, Snow R. A spatial national health facility database for public health sector planning in Kenya in 2008. *Intern J Health Geograph*. 2009. 8(1):13.

c. Define Catchment Areas

By knowing the geographic area that a particular health facility serves, it is possible to define and quantify the population that has access to it. In some situations, data may be available on the actual catchment areas of facilities, in which case, this can be used to link the catchment population to each facility (Tanser, Gijsbertsen & Herbst, 2006). In the absence of such data, however, it is possible to calculate theoretical catchment areas based on the spatial distribution of the facilities. Several of these methods will be outlined. All of them require health facility coordinates, but all can be used with population data in point or raster format.

i. Thiessen Polygons

Description: In a GIS, a geographic area can be divided up based on a set of coordinates by generating a layer of *Thiessen polygons*. These polygons are defined by a set of points, such that their sides are lines that are equidistant between each point and its nearest neighboring point. An example of this is illustrated in figure 11. By definition, any location within a given Thiessen polygon is closer to its associated point than to any other point. Thiessen polygons can be used as a crude method for defining theoretical health facility catchment areas.

Data requirements:

- a point layer representing settlement locations or a population density raster
- a point layer representing health facility locations

Steps:

1. In a GIS, generate a Thiessen polygon layer based on the health facility point layer.
2. Use this layer to link population data to particular health facilities in one of the following ways:

   a. The settlement points can be assigned to their nearest facility by spatially joining them to the Thiessen polygon layer (see annex E, section a).
   b. The total population of a health facility's catchment area can be summed by extracting zonal statistics from a raster population layer.
   c. If population data are only available as administrative unit polygons, the statistics can be summed and, where feasible, apportioned from these to the Thiessen polygons (while taking into account the modifiable areal unit problem and the other caveats mentioned previously [see section a. of this annex]).

Outcome: A layer of polygons representing theoretical facility catchment areas, with a field in the attribute table giving the estimated catchment population, is the outcome.

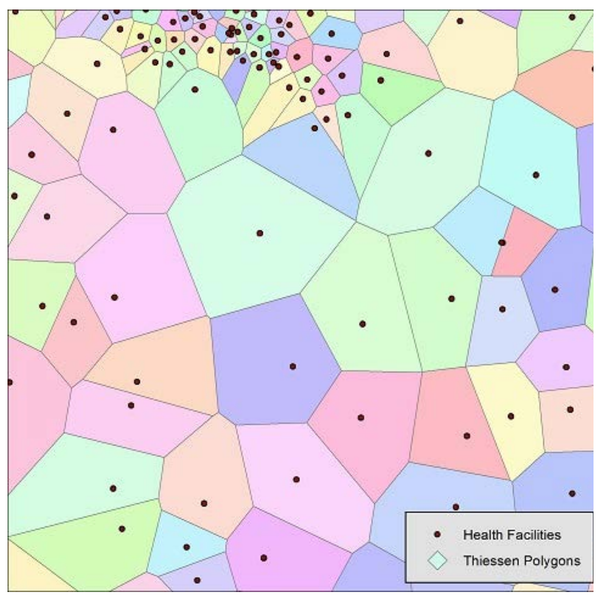


Figure 11: Illustrative example of Thiessen polygons created based on a set of points representing health facilities.

Targeting:

- The more evenly-spaced health facilities are across the area of interest the more their Thiessen polygons will be regular in shape, similar in size and with the health facility point close to the center. This approach does assume that the underlying population will be evenly dispersed. Therefore the population density must be taken into consideration when interpreting Thiessen polygons, or any catchment area. By mapping the polygons over a raster population layer, areas of potential low access can be picked out by identifying polygons that are unusually large or long, or otherwise at variance with the geographic distribution of the population. In the attribute table of the polygon layer, a field can be created giving the area or extent of each polygon. This can be used as a continuous variable to classify catchment areas according to a threshold size.
- By calculating the total population of each catchment polygon using zonal statistics (see annex E, section c), it is then possible to calculate their accessibility ratios (the numerator will always be 1, since, by definition, there is one health facility per polygon). Populous polygons with low accessibility ratios might indicate an area that could benefit from investment.

Advantages:

- **Simplicity:** Thiessen polygons are simple to calculate and straightforward to interpret.
- **Versatility:** Thiessen polygons can be used to assign any given location within the area of interest to a particular facility and can be used with population data in point or raster form.

Limitations:

- **Define linkage, not access:** Thiessen polygons link a population to its nearest facility, but by themselves cannot divide the population within a particular catchment area into those with and those without access to services.
- **Does not reflect true travel distance:** As with all approaches based on direct distance (as opposed to network distance), Thiessen polygons have the limitation of not taking into account local geography, actual transport networks, and other factors that may affect a facility's accessibility.
- **Does not reflect actual usage:** Since Thiessen polygons represent only theoretical catchment areas, they can only approximate actual facility utilization. Based on other factors clients may decide to access services from facilities that are not the nearest to them. In general, within a given polygon, this error will be greatest for settlements that are close to the border of the catchment area than for those closer to their associated facility.
- **Do not reflect variations in population density:** It cannot be assumed that the population of a theoretical catchment area is evenly distributed. Differences in underlying population density within and between Thiessen polygons must be taken into consideration.

Further resource:

Noor AM, Zurovac D, Hay SI, Ochola SA, Snow RW. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Trop Med Intern Health.* 2003. 8(10):917-26.

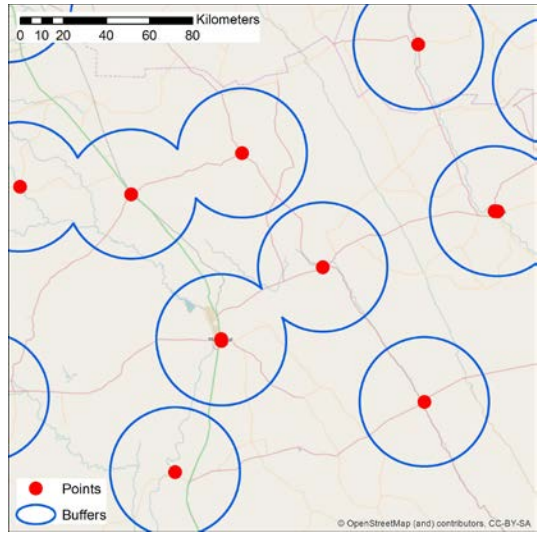


Figure 12: Illustrative example of a set of points with 30km buffers.

ii. Buffers

Description: Buffers are circular polygons with a point at the center and with a radius defined by the buffer distance from that point (figure 12). In the case of points representing health facilities, a buffer can be thought of as a theoretical catchment area demarcating the geographic extent of its coverage or draw.

Data requirements:

- a point layer representing settlement locations or a population density raster
- a point layer representing health facility locations

Steps:

1. Using a buffer tool in a GIS, specify a buffer distance *r* representing the threshold distance by which access to a health facility will be defined. This will create a polygon layer of circular buffers with radius *r*, similar to those shown in figure 12.
2. This layer can be used in a spatial analysis to divide the eligible population into those with and without access in the following ways:
   a. Settlement points (or population polygons) falling outside the polygon can be defined for analysis as lacking access.
   b. The total population of the buffer polygon can be summed by extracting zonal statistics from a population raster layer to calculate the population with access. Subtract this from the total population to arrive at the total population lacking access.
   c. If population data is only available as administrative unit polygons, the statistics can be summed and, where feasible, apportioned from these to the buffer polygons (while taking into account the modifiable areal unit problem and the other caveats mentioned previously — see annex C.a.i).

Outcome: The outcome is polygon layer representing the theoretical catchment area of the health facilities.

Targeting: The population that falls within a health facility's buffer is defined as having access to that facility's services, while the population outside does not. Buffers, therefore, divide the population into binary categories. Multiple buffers at increasing distance intervals can be used to create additional categories; for example, concentric buffer polygons generated at regular increments of the buffer distance (e.g., 5km, 10km, 15km, 20km). These multiple buffers can help to distinguish between populations with different levels of access.

Advantages:

- **Simplicity:** Buffers are simple to calculate and straightforward to interpret.

Limitations:

- **Does not reflect true travel distance:** Buffers suffer from many of the same drawbacks and potential sources of errors as straight-line analysis in that the straight-line distance used to create the buffer is at best a crude indicator for actual travel time.
- **Arbitrary definition of access:** Specifying a buffer distance means setting an arbitrary threshold definition for access. The size of the buffer is selected to reflect a reasonable representation of access. For some services or some locations a 5km buffer is used because it represents what a person may walk in one hour (Skiles et al. 2013), however other situations may require a buffer of a larger or smaller size. Regardless of the size selected the buffer size is an arbitrary indication. There is no reason to assume automatically that individuals just inside the buffer have significantly better access to those just beyond the buffer. Furthermore, a person on the periphery of the buffer is considered to have the same level of access as a person near the center.

Further resources:

Adamou B, Curran J, Wilson L, Apenem Dagadu N, Jennings V, Lundgren R, et al. *Guide for Monitoring Scale-up of Health Practices and Interventions*. Chapel Hill, NC: MEASURE Evaluation; 2013.

MEASURE Evaluation. *Geographic Approaches to Global Health: A Self-Directed Minicourse.* Chapel Hill, NC: MEASURE Evaluation; 2013. Available at: http://www.cpc.unc.edu/measure/publications/ms-12-56.
This course can be taken online to earn a certificate at: https://training.measureevaluation.org/certificate-courses/geo-global-health-en

Skiles M, Burgert C, Curtis S, Spencer J. 2013. Geographically linking population and facility surveys: methodological considerations. *Popn Health Metrics.* 2013. 11(1):14

Spencer J, Angeles G. Kernel density estimation as a technique for assessing availability of health services in Nicaragua. *Health Serv Outcomes Res Meth.* 2007. 7(3-4):145-157. DOI: 10.1007/s10742-007-0022-7.

iii. Network Analysis

Description: A network analysis attempts to model the true journey that clients would have to make to get to a health facility along roads and other transport routes. While Thiessen polygons, buffers, and spider-diagrams define access or linkage based on straight-line distance, geospatial analysis makes it possible to model more realistic travel routes, times, and distances using network analysis. Using GIS data on transport infrastructure, such as a polyline layer representing roads or other transport routes, a population can be categorized based on its distance or travel-time along a route to a health facility (Tanser, Gijsbertsen & Herbst, 2006).

Data requirements:

- a polyline layer representing roads or other transport routes (advanced network models can take into account other road attributes such as road type or surface to account for differences in speed when traveling on different terrains)
- a point layer representing settlement locations or a population density raster
- a point layer representing health facility locations

Steps:

1. Prepare the source data set — a network data set will be created based on a polyline layer representing roads. Sources of GIS road data sets are listed below. All the variables that will be taken into account in a network analysis will need to be present as fields in the source data set. At a minimum, the source data set must have a field representing the network impedance values — distance or travel time — of each stretch of road in the layer. A road-type variable, indicating the hierarchy of roads, can be specified in the database and the network defined such that routes along main roads are preferred over smaller roads. The network may include straight-line distances to represent walking time. For example, one could sum the (straight-line) distance from a settlement to its nearest road, plus the distance along the road to the nearest point to the health facility and then the distance of the facility from the road. This would model the journey a client would have to make by foot from their community to the road, then in a vehicle along the road, and from the road to the facility.
2. Create a network dataset specifying the roads layer as the source.
3. Either:

    a. conduct a *service area* analysis based on the network dataset to arrive at a polygon representing the service area; or
    b. conduct a closest facility analysis to link settlements to their nearest facility along the network and determine the travel time or distance.

Outcome: There are two possible outcomes from a network analysis that are useful for determining access:

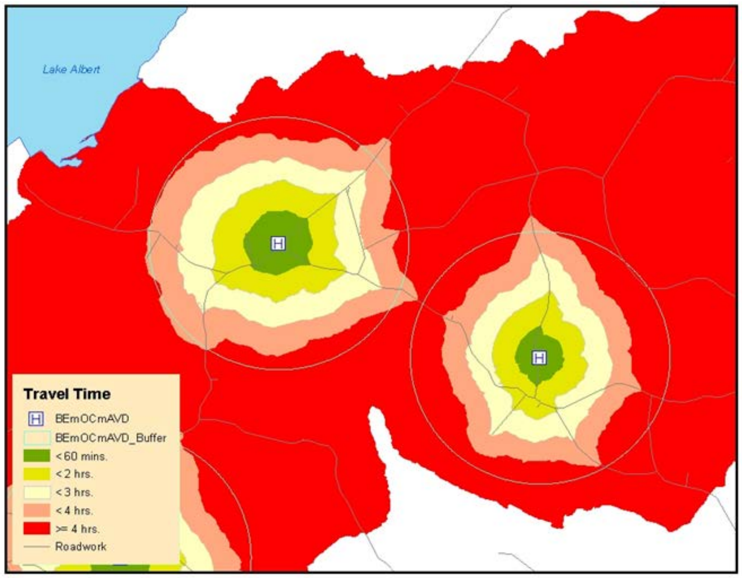


Figure 13: Example of service area polygons defined by network distance.

1. **A catchment area polygon:** There are two equivalents of buffers (see section ii. Buffers above, in this annex) in a network analysis:

    a. Isodistance lines (that connect points of equal distance along the network)
    b. Isochrones (connecting points of equal travel time)

    These lines define the boundaries of a health facility's service area, so the population can be divided up according to whether it falls inside or outside the service area. An example is given in figure 13 and shows service areas based on modeled walking travel for two obstetric care facilities in Kibaale, Uganda (McCracken & Acire, 2013). The equivalents of Thiessen polygons in a network analysis have been referred to in one study as travel time catchments (Tanser, Gijsbertsen & Herbst, 2006).

2. **A network distance variable:** A field that gives, for every population settlement, the distance along the transport network to the facility with which it is linked is an example. It is often convenient to define this linkage in terms of network distance, so that a settlement is linked to whichever health facility is nearest to it along the transport network. However, linkage can be defined in other ways. For example, if in a survey, a respondent has specified a particular health facility as one that they actually utilize, then the linkage can be defined based on actual usage.

Targeting: Network distance can be analyzed in the same way as straight-line distance. A population feature may be assigned a network distance as a continuous variable, or a binary variable according to whether it is located beyond a certain threshold network distance. Service area polygons can be used to classify a population as having or lacking access to a facility, while a population center can be linked with the health facility that is closest to it along the transportation network. Similarly, zonal statistics can be extracted from a population raster based on the service area polygon to give the catchment population (see annex E, section c).

Advantages:

- **Accuracy:** Network analysis gives more realistic and accurate estimation of the true time it takes to travel to a facility than estimates based on straight-line distance.

Limitations:

- **Arbitrary distance:** specifying a service area polygon by a certain distance or time, means setting an arbitrary threshold definition for access. The size of the polygon is selected to reflect a reasonable representation of access. However, it is an arbitrary indication. There is no reason to automatically assume individuals just inside the service area have significantly better access to those just outside. Furthermore, a person on the periphery of the area is considered to have the same level of access as a person near the center.

Source of GIS road data sets:

- **gRoads:** This is a global database that uses a consistent and documented methodology so is good for comparative analysis across multiple countries. However, often information on road type is sparse for developing countries. http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1/docs
- **OpenStreetMap:** Data from the OpenStreetMap project can be downloaded and imported into a GIS. http://planet.openstreetmap.org/
- **Humanitarian Response:** This organization's website has some road datasets available as line *shapefiles* https://cod.humanitarianresponse.info/search/field_data_type/lines-246
- **DIVA-GIS:** This website makes available road datasets from the Digital Chart of the World http://www.diva-gis.org/gdata

Further resources:

Esri. Network analyst tutorial [online training course]. Available at: http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/network-analyst-tutorial.pdf

Esri. ArcGIS Network analyst — creating nework data sets [online presentation]. Available at: http://video.esri.com/watch/1834/arcgis-network-analystcreating-network-datasets

Esri. Using nework analyst in ArcGIS Deskop [online training course]. Available at : http://training.esri.com/Courses/ts_NetworkAnalyst10/player.cfm

iv. Kernel Density Estimation

Description: Kernel density estimation (KDE) is a geospatial method that creates a smooth raster surface from an attribute value of a point data set (figure 14). It "disperses" the variable, in order to approximate more realistically the geographic spread of that variable. In the case of health services, this is to estimate the "draw" that a facility exerts upon its surrounding population.

This draw has _distance decay_ — it gets weaker the further away from the facility location, provided there is no other facility in the opposite direction. Therefore, the population that is situated among a cluster of health facilities will be assigned a higher value than those that only have health facilities in one direction.

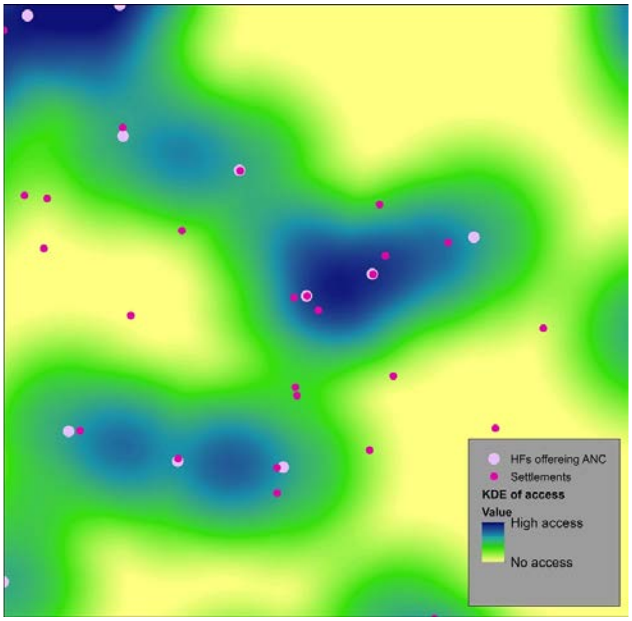


Figure 14: Illustrative example of a KDE raster layer generated based on health facility locations.

Data requirements:

- a point layer representing settlement locations with population attributes.
- a population raster layer
- a point layer representing health facility locations (if the data include relevant attributes about the facilities' staffing or services, this can be incorporated into the analysis)

Steps: Generate a KDE layer based on a set of health facility coordinates. The following parameters will need to be specified:

1. **A density variable:** Named "population field" in the ArcGIS KDE tool, this is the variable that is to be spread across the surface. This can be, for example, presence or absence of the service (in which case, leave the population field as "none"), the number of staff qualified to give the service, or a service readiness score.
2. **The distribution of the variable around the point:** A normal (Gaussian) distribution is standard and assigns greater weight to areas near the center.
3. **Kernel size/bandwidth:** A radius representing a buffer of the extent of the influence of the health facility. This can be tailored to each facility based on characteristics e.g. 10km for hospitals, 5km for health centers and 2.5km for dispensaries.
4. **Grid cell size:** The size of the grid squares produced by the KDE, e.g. 500m^2.

Outcome:

- a gridded raster layer, which assigns an estimated value of the attribute to every grid square in the area of interest (figure 14 shows an illustrative example)

Targeting:

- Points representing population settlements can be compared based on their KDE access values extracted from the raster.

- The resulting kernel density grid can be divided by the population density grid using map algebra to get a raster surface showing people per facility or (accessibility ratios).
- From the resulting raster, a contour line can be created connecting the points of equal accessibility ratio. This can be used to classify population as having or lacking access according to a threshold ratio.

Advantage:

- **Simulates distance decay —** While buffers define access in terms of binary categories (areas with and without access), KDE assigns numerical values to each square in a grid. The further away from a facility a population is located, the lower their access. KDE attempts to more realistically simulate the distance decay in a health facility's "draw". KDE rasters allow for access to be treated as a continuous variable and assigns a value for this variable to every raster square within the kernel.

Limitations:

- **Does not reflect true travel distance —** The kernel size in a KDE is still defined according to straight-line distance making the approach open to many of the limitations of the other straight-line approaches discussed in Annex D.b. For example, it does not take into account the realities of the geographic environment or the constraints of the existing transport network.
- **Raster extent —** The output raster will have the same extent as the point layer on which the KDE was based since the tool cannot interpolate values outside that extent. In order to ensure that the raster covers the entire area of interest, it is necessary to specify a buffer around the point layer so that the KDE calculation does not stop at the extent of the point layer.
- **Difficult to interpret —** The KDE values for a particular point (such as a settlement) can't easily be presented in a table or report because on their own, they are hard to interpret. It is more appropriate to use them as relative measures to compare several points or as inputs into regression type models.

Further resources:

MEASURE Evaluation. *Geographic Approaches to Global Health: A Self-Directed Minicourse.* Chapel Hill, NC: MEASURE Evaluation; 2012. Available at: http://www.cpc.unc.edu/measure/publications/ms-12-56.
This course can be taken online to earn a certificate at: https://training.measureevaluation.org/certificate-courses/geo-global-health-en

Skiles M, Burgert C, Curtis S, Spencer J. 2013. Geographically linking population and facility surveys: methodological considerations. *Popn Health Metrics.* 2013. 11(1):14.

Spencer J., Angeles G. Kernel density estimation as a technique for assessing availability of health services in Nicaragua. *Health Serv Outcomes Res Method.* 2007. 7(3-4):145-157. DOI: 10.1007/s10742-007-0022-7.

Annex E: General Geospatial Methods for Converting between Different Types of Data

This annex outlines some general geospatial methods that can be used in many types of geospatial analysis to convert one type of spatial data to another:

a. spatial join
b. calculate a new raster for the eligible population
c. aonal statistics
d. calculating polygon centroids

a. Spatial Join

Description: A spatial join is a basic GIS tool that takes data from a polygon layer and assigns it to a point layer based on the spatial location of the points. If a set of coordinates representing health facilities is available but with no information about what administrative unit they are located in, they can be spatially joined with a polygon layer of administrative unit boundaries, and each facility assigned the attribute values (including unit names or codes) of the unit in which its point falls.

Data requirements:

- a polygon layer of administrative unit boundaries with an identifier field in its attribute table
- a point layer representing the coordinates of health facility locations

Steps:

1. Join the point layer to the polygon layer based on spatial location to generate a second point layer. In this layer's attribute table will be the attributes of the units in which each point falls.
2. Export the attribute table to a spreadsheet for further analysis.

Outcome: The attribute table of the new point layer can be exported to a spreadsheet in order to sum the number of facilities per unit.

Advantages:

- **Simplicity:** Spatial joins are simple to do.

Limitation:

- **Accuracy of borders:** The accuracy of a spatial join depends on how well the boundaries in the shapefile match the true borders of the administrative units. Boundaries may change over time and administrative divisions may get divided into new units, rendering shapefiles out-of-date. Furthermore, the boundaries in the shapefile may have been generalized to make them simpler and more regular, which can also cause misclassification in the results of a spatial join.

b. Calculate a New Raster for the Eligible Population

Description: If the distribution of the total population is available as a raster, but the eligible population is only available at the level of the administrative unit, the population raster can be adjusted by the proportion of the population in each unit that is eligible. Two rasters can be combined within a GIS based on a mathematical operation using map algebra. For example, the raster calculator can take the values for all the grid squares of one raster and multiply each one by values in the corresponding squares of a second raster to produce a third raster of the product of the two rasters.

Data requirements:

- a raster population layer
- population statistics broken down by administrative unit, sex and age group
- a polygon layer of the administrative unit boundaries

Steps:

1. Calculate the proportion of the overall population that is eligible for the service by administrative unit.
2. Join the proportions to the polygon shapefile of the administrative units so that there is a proportion field in the attribute table.
3. Use the Polygon to Raster tool to convert the polygon into a grid using the proportion field as the value field.
4. Use Raster Calculator to multiply the values from the population raster by the proportions from the other raster to generate a third raster, which will be a gridded layer of the eligible population.

Outcome:

- a raster layer giving an approximation of the eligible population per grid square

Advantage:

- **Detail:** The output raster can be used to calculate eligible population statistics for points and polygons (rather than overall population).

Limitations:

- **Modifiable areal unit problem:** Since the proportion of the population that is eligible is summarized at the administrative unit level, this approach is subject to the modifiable areal unit problem. When a spatial phenomenon is aggregated up to the level of an administrative unit, bias is introduced to the analysis. Unit level statistics depend on how the boundaries of the unit happen to be drawn.
- **Uniform distribution:** Age structure and population density vary across administrative units, while this method assumes that it is uniformly distributed across each unit.
- **Accuracy of borders:** The accuracy of the proportions raster will depend on how well the boundaries in the shapefile match the true borders of the administrative units.
- **Grid square size:** Before performing a calculation based on two rasters, the rasters need to be "snapped" to one another so that the grid square align as much as possible.

c. Zonal Statistics

Description: Zonal statistics is a GIS tool that calculates statistics based on the values of a raster layer for each polygon within a polygon layer.

Data requirements:

- a raster surface layer representing the distribution total or eligible population
- a polygon layer of administrative units or health facility catchment areas

Steps:

1. Using a zonal statistics tool within a GIS, choose the polygon layer as the input feature, the population raster as the input value raster and "Sum" as the statistics type. The attribute table of the output layer will contain the total population of each polygon in its "Sum" field.
2. Join the total population data with the attribute table of the original polygon layer.

Outcome: An attribute table giving the eligible population for every polygon in the layer is the outcome.

Advantages:

- **Versatility —** Zonal statistics can be used to derive population estimates for any polygon layer, including layers created from service area analysis (see annex D, section c.iii) or accessibility contour (see annex D, c.iv).

Limitations:

- **Aggregating raster data:** Raster population layers are intended to be accurate at low levels. Aggregating the raster data up to a higher level administrative unit, will not necessarily give results that match aggregated census data.

d. Calculating Polygon Centroids

Description: A polygon layer can be converted to a point data set by taking the coordinates of the polygon centroids.

Data requirements:

- a low-level administrative unit polygon layer with population data in its attribute table

Steps:

1. In the attribute table for the polygon layer, create two new fields and name them "Latitude" and "Longitude".
2. Using Calculate Geometry, calculate the Y coordinate of the centroid in the Latitude field and the X coordinate in the Longitude field.
3. Export the table and then use it to make an XY event layer from the latitude and longitude fields.

Outcome:

- a point dataset representing the geographic center-points of the polygons

Advantage:

- **Versatility:** Point data lend themselves better to many of the approaches explored in straight line distance (see annex D, section b) and catchment areas (see annex D, section c).

Limitations:

- **Prone to error:** This method assigns the geographic location of the unit's centroid to the whole population of the unit. In reality, a population will be distributed throughout the unit. Therefore analyses based on distance will be less accurate for populations that are located near the edges of the unit. The larger the average unit size or the variation in size, the greater will be the error when using the centroid point layer in an analysis. This method is only recommended for very small, low-level units such as wards or enumeration areas.
- **Centroids outside of original unit:** For irregularly shaped units the centroid may fall outside the borders of the original unit.

Further resources:

Ersi. How to: find the centroid of polygons using Calculate Geometry [online training course]. Available at: http://support.esri.com/en/knowledgebase/techarticles/detail/32988

References

Al-Taiar A, Clark A, Longenecker JC, Whitty CJ. 2010. Physical accessibility and utilization of health services in Yemen. *Internl J Health Geog.* 2010. 9:38. DOI: 10.1186/1476-072x-9-38.

Burgert CR, Bigogo G, Adazu K, Odhiambo F, Buehler J, Breiman RF, Laserson K, Hamel MJ, Feikin DR. Impact of implementation of free high-quality health care on health facility attendance by sick children in rural western Kenya. *Trop Med Intern Health.* 2011. 16 (6):711-720.DOI:10.1111/j.1365-3156.2011.02752.x.

Central Bureau of Statistics [Namibia]. Windhoek, Namibia:*Republic of Namibia Population Projections*. National Planning Commission; 2006.

Ensor T, Cooper S. 2004. Overcoming barriers to health service access: influencing the demand side. *Health Pol Planning.* 2004. 19(2):69-79.

Gorr WL, Kurland KS. *GIS Tutorial 1: Basic workbook*. Redlands, CA: Esri Press; 2011.

Guagliardo M. 2004. Spatial accessibility of primary care: concepts, methods and challenges. *Intern J Health Geog.* 2004. 3(1):3.

Humanitarian Response. *Namibia Atlas*. Namibia: Settlements (villages, towns, cities) [database]. Winhoek, Namibia: Atlas of Namibia Project, Directorate of Environmental Affairs, Ministry of Environment and Tourism; 2002.

McCracken S, Acire G. Addressing maternal mortality: using GIS for evaluating access to emergency obstetric care in the SMGL Project in four districts of western Uganda [presentation]. Global Maternal Health Conference, Jan. 15-17, 2013, Arusha, Tanzania.

McLafferty S, Grady S. Prenatal care need and access: a GIS analysis. *J Med Systems.* 2004. 28(3):321-33.

Ministry of Health and Social Services (MoHSS) [Namibia] and ICF Macro. 2010. Namibia Health Facility Census 2009 [data set]. NMGE61.shp. Windhoek, Namibia: MoHSS and ICF Macro [Producers]. ICF International [Distributor].

Moise IK, Cunningham M, Inglis A. Use of geographical information systems for M&E: informed decision-making framework [unpublished manuscript]. Chapel Hill, NC: MEASURE Evaluation.

Namibia Ministry of Health and Social Services (MoHSS), ICF Macro. *Namibia Health Facility Census 2009*. Windhoek, Namibia: MoHSS and ICF Macro; 2010.

Namibia Ministry of Health and Social Services (MoHSS), Macro International Inc. *Namibia Demographic and Health Survey 2006-07*. Windhoek, Namibia and Calverton, MD: MoHSS and Macro International Inc.; 2008.

National Planning Commission, United Nations Population Fund (UNFPA). *Namibia 2011 Population and Housing Census Preliminary Results*. Winhoek, Namibia: National Planning Commission; 2012.

Noor AM, Zurovac D, Hay SI, Ochola SA, Snow RW. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Trop Med Intern Health.* 2003. 8(10):917-26.

Skiles M, Burgert C, Curtis S, Spencer J. 2013. Geographically linking population and facility surveys: methodological considerations. *Popn Health Metrics.* 2013. 11(1):14.

Spencer J., Angeles G. Kernel density estimation as a technique for assessing availability of health services in Nicaragua. *Health Serv Outcomes Res Method.* 2007. 7(3-4):145-157. DOI: 10.1007/s10742-007-0022-7.

Tanser F, Gijsbertsen B, Herbst K. Modelling and understanding primary health care accessibility and utilization in rural South Africa: an exploration using a geographical information system. *Soc Sci Med.* 2006. 63(3):691-705.DOI: 10.1016/j.socscimed.2006.01.015.

Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Pop Health Metrics.* 2012. 10(1):8. DOI: 10.1186/1478-7954-10-8.

Tatem AJ, Campbell J, Guerra-Arias M, de Bernis L, Moran A, Matthews Z. 2014. Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *Intern J Health Geograph.* 2014. 13 (1):2.

Tatem AJ., Linard C. WorldPop project [Web page database]. Southampton, United Kingdom: University of Southampton; ND. Accessed 1 October 2013 at http://www.worldpop.org.uk/.

Wang W, Wang S, Pullum T, Ametepi P. How family planning supply and the service environment affect contraceptive use: findings from four East African countries. In *DHS Analytical Studies*. Calverton, MD: ICF International; 2012.

Yao J, Murray AT, Agadjanian V, Hayford SR. Geographic influences on sexual and reproductive health service utilization in rural Mozambique. *Applied Geog.* 2012. 32(2):601-607. DOI: 10.1016/j.apgeog.2011.07.009.