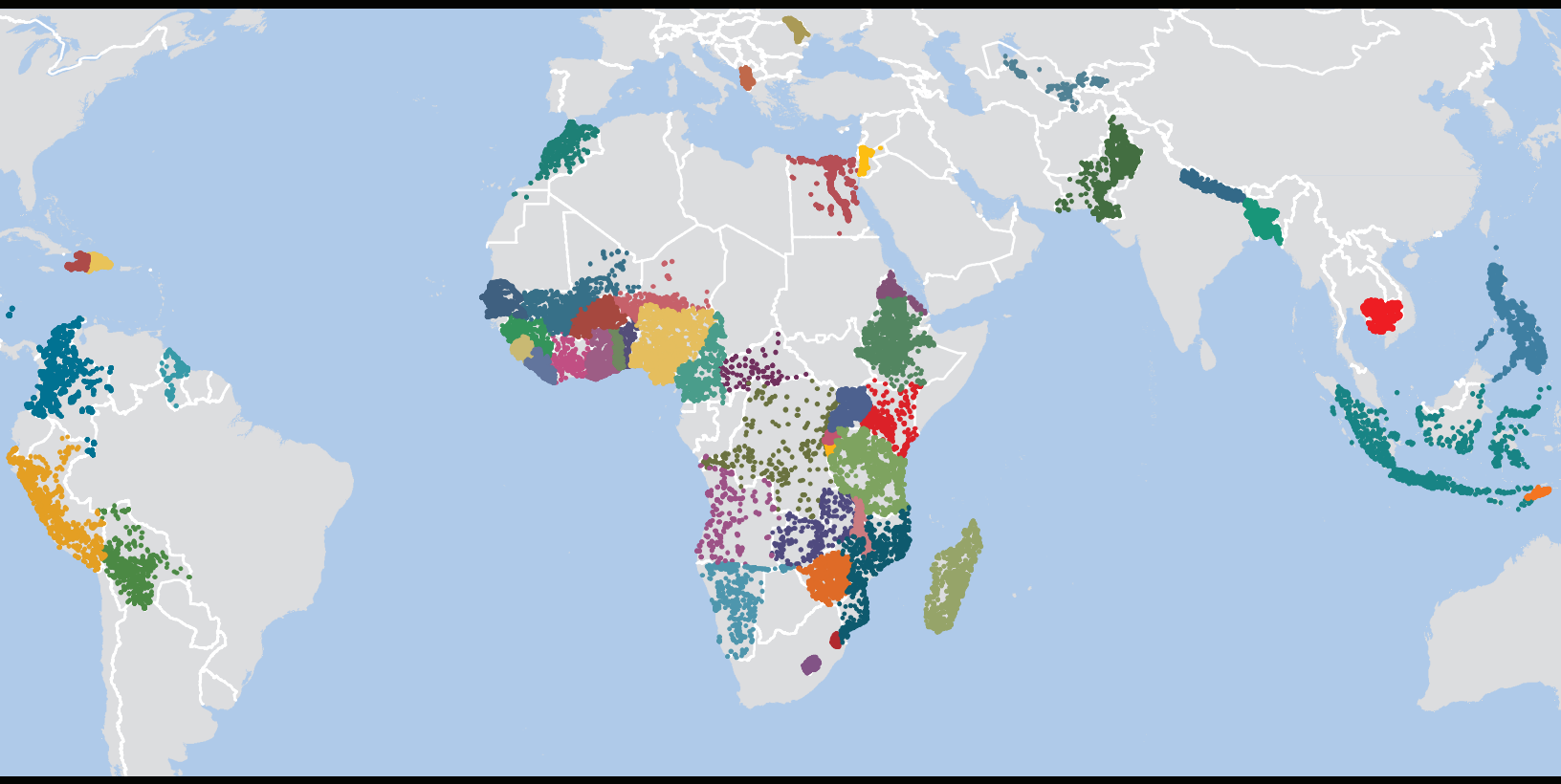
USAID
FROM THE AMERICAN PEOPLE

LINKING DHS HOUSEHOLD AND SPA FACILITY SURVEYS: DATA CONSIDERATIONS AND GEOSPATIAL METHODS

DHS SPATIAL ANALYSIS REPORTS 10



SEPTEMBER 2014

This publication was produced for review by the United States Agency for International Development (USAID). The report was prepared by Clara Burgert and Debra Prosnitz of ICF International, Rockville, MD, USA.

DHS Spatial Analysis Reports No. 10

Linking DHS Household and SPA Facility Surveys: Data Considerations and Geospatial Methods

Clara R. Burgert

Debra Prosnitz

ICF International

Rockville, Maryland, USA

September 2014

Corresponding author: Clara R. Burgert, International Health and Development, ICF International, 530 Gaither Road, Suite 500, Rockville, Maryland, USA; phone: 301-572-0446; e-mail: clara.burgert@icfi.com

Acknowledgment: The authors would like to acknowledge Paul Ametepi, Wenjuan Wang, Martha Skiles, Marc Cunningham, Andrew Inglis, and Blake Zachary for their review of this report.

Editors: Sidney Moore
Document Production: Yuan Cheng

This study was carried out with support provided by the United States Agency for International Development (USAID) through The DHS Program (#GPO–C–00–08–00008–00). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. For additional information about The DHS Program contact: DHS Program, ICF International, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA. Phone: 301-407-6500; Fax: 301-407-6501; Email: reports@dhsprogram.com; Internet: www.dhsprogram.com.

Recommended citation:

Burgert, Clara R., and Debra Prosnitz. 2014. *Linking DHS Household and SPA Facility Surveys: Data Considerations and Geospatial Methods.* DHS Spatial Analysis Reports No. 10. Rockville, Maryland, USA: ICF International.

Contents

Tables ..... **v**

Figures ..... **v**

Abbreviations ..... **vi**

Preface ..... **vii**

Abstract ..... **ix**

Executive Summary ..... **xi**

1. Introduction **1**

1.1. Context: Why Link Household and Facility Surveys? *1*

1.2. Overview: Linking DHS and SPA Data *2*

2. Data Description **3**

2.1. DHS Household Surveys *3*

2.2. SPA Surveys *5*

3. Topics That Can be Linked between DHS and SPA Surveys **7**

3.1. Relevant DHS Data Topics to Link with SPA Data *7*

3.2. Relevant SPA Data Topics to Link with DHS Data *8*

3.3. Operationalization of Data for Analysis *9*

4. Considerations before Linking DHS and SPA Surveys **11**

4.1. Data Considerations *11*

4.2. Country Considerations *13*

5. Geographic Linkage Methods **15**

5.1. Recommended Linkage Methods *18*

6. Case Study: Haiti Birth Delivery Services **21**

6.1. Methods *21*

6.2. Results *23*

6.3. Discussion *25*

7. Drawing Conclusions from DHS and SPA Linkage **27**

References **29**

Tables

Table 1. Summary of methods appropriate for linking DHS and SPA data 16

Figures

Figure 1. Illustration of DHS GPS data displacement 5

Figure 2. Summary of DHS data collection relevant for SPA linkage 8

Figure 3. Example of linkages made in literate between DHS and SPA data 10

Figure 4. Example of timing of DHS and SPA surveys 11

Figure 5. Linking DHS and SPA data considerations 14

Figure 6. Example of possible distance error in direct distance measurement with displaced location 15

Figure 7. Example sample domain linkage 18

Figure 8. Visualizations of linkage methods 20

Figure 9. Haiti DHS and SPA survey locations 21

Figure 10. Illustration of linkage steps 23

Figure 11. Percentage of DHS clusters having a BEmOC facility within 10 km, by residence 24

Figure 12. Percentage of women delivering at health facility in the past five years who live within 10 km of at least one facility providing BEmOC, by residence 24

Abbreviations

ACT	Artemisinin-based Combination Therapy
AIDS	Acquired Immunodeficiency Syndrome
AIS	AIDS Indicator Survey
ANC	Antenatal care
BEmOC	Basic Emergency Obstetric Care
DHS	Demographic and Health Surveys
EA	Enumeration Area
FP	Family Planning
GIS	Geographic Information System
GPS	Global Positioning System
HIV	Human Immunodeficiency Virus
IUD	Intrauterine Device
KDE	Kernel Density Estimation
MFL	Master Facility List
MIS	Malaria Indicator Surveys
NGO	Nongovernmental Organization
SARA	Service Availability and Readiness Assessment
SPA	Service Provision Assessment
STI	Sexually Transmitted Infections
USAID	United States Agency for International Development
VCT	Voluntary Counseling and Testing
WHO	World Health Organization

Preface

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

The DHS Spatial Analysis Reports supplement the other series of DHS reports to meet the increasing interest in a spatial perspective on demographic and health data. The principal objectives of all DHS report series are to provide information for policy formulation at the international level and to examine individual country results in an international context.

The topics in the DHS Spatial Analysis Reports are selected by The DHS Program in consultation with the U.S. Agency for International Development. A range of methodologies are used, including geostatistical and multivariate statistical techniques.

It is hoped that the DHS Spatial Analysis Reports series will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor

Director, The DHS Program

Abstract

This report examines the potential benefits of linking Demographic and Health Surveys (DHS) data and Service Provision Assessment (SPA) survey data to answer research questions. A case study using Haiti DHS and SPA datasets demonstrates the practical application of linkage to delivery of obstetric services. Individually, the DHS household survey and the SPA facility-based survey are able to answer many relevant questions on access, utilization, and quality of health services. However, being able to link data from the two sources may offer greater insight into 1) how and why individuals access health care and 2) how the type and quality of care at facilities in their service environment can influence care-seeking behavior. The DHS and SPA survey datasets are ideal sources of information on the targeting and impact of health programs. However, for a long time, issues of sampling, geographic data accuracy, access, and operationalization of methodology have prevented effective linkage of DHS and SPA datasets. In 2013, Skiles et al. presented an in-depth methodological study of the issues surrounding geographic displacement of DHS clusters and the sampling of SPA facilities. The work presented here builds on the research carried out by Skiles et al. and outlines specific considerations that need to be examined before linkage of DHS and SPA survey data can take place. The data and linked analyses discussed here may be a starting point for further investigation of care-seeking behaviors, access to health services, and issues of quality of services that require additional quantitative and qualitative study.

Executive Summary

Separately the Demographic Health Survey (DHS) and Service Provision Assessment (SPA) surveys can answer many relevant questions on access, utilization, and quality of services, but being able to link them together may offer even greater insight into why and how individuals access care and how the type or quality of care at the facilities near their homes might influence their actions. Together the DHS and SPA are ideal tools to help researchers, program managers, and decisionmakers answer key questions around program impact and targeting. Geospatial analysis is one tool with which to identify gaps in access and potential barriers to access, all to better target efforts to promote care-seeking and increase use of services. However, for many years several issues including sampling, geographic data accuracy, data access, and problems with the operationalization of linking these datasets prevented the DHS and SPA data from being used in this manner. In 2013, Skiles et al. presented an in-depth methodological study examining the issues around geographic displacement of DHS clusters and the sampling of SPA facilities, to better understand the extent of misclassification and error associated with making linkages between these two datasets (Skiles et al. 2013). The work presented here builds on that extensive methodological study and outlines a number of specific considerations that need to be understood before a linkage between the DHS and SPA surveys can take place.

The technical linkage between the DHS and SPA surveys is easily done in a GIS, but before this is done, it is essential to consider if it is appropriate to make such a linkage at all. Often problems with the linkage can be overcome by reframing the question of interest to more appropriately fit the data or country context. For example, moving from a smaller geographic scale analysis to a larger service environment analysis can make linkage more appropriate. Primarily, if the SPA survey is not a census of facilities or a census of facilities relevant for the question of interested, then the surveys should not be linked at geographic scales below the sample domain. Linking a SPA that is a sample of facilities in the country to a DHS survey causes high misclassification in the subsequent analysis (Skiles et al. 2013). The other areas for consideration are separated into two main groups: data and country context. Data include timing of surveys in relation to one another, DHS cluster displacement, missing data, indicators of rare events and indicator reference period. Country context includes health system structure and geographic barriers to care.

Many geographic linkage methods could be used for linking the DHS and SPA datasets but not all methods should be used. Some examples of appropriate methods include: sample domain linkage, service environment linkage, estimated surfaces, and catchment area. Regardless of the method used for linkage, it is important to note that DHS data does not specify the facility where a person sought or received care; these data are not recorded in the final dataset. The considerations and linkage methods are demonstrated in a short case study using data from Haiti DHS and SPA surveys.

There are many benefits that can be achieved by linking DHS and SPA survey data. Such linkage has the potential of generating answers to key programmatic and policy questions. However, many considerations have to be taken into account when linking these data, particularly geographic limitations. Framing the question of interest and appropriately operationalizing DHS and SPA variables are key steps in creating an appropriate linkage between the two survey databases. It is important when interpreting the results of such analyses to be cognizant of limitations to the questions posed and the data used. These data and linked analyses may be a starting point for further investigation of care-seeking behaviors, access to health services, and issues of quality of services that require other quantitative and qualitative studies.

1. Introduction

There is a need to better identify gaps in access and utilization of care to target services and programs and ensure more equitable coverage of health services. Program managers need to identify the populations in greatest need and, with researchers, explore new and better ways to provide equitable access to quality services. To do this, valid information is needed about individual care-seeking behaviors, access to these services, and the quality of services offered.

Household and facility surveys are key inputs for a program manager or decisionmaker to use in understanding the health seeking behaviors, health service utilization, and health service preparedness for a given country. Data from The Demographic and Health Survey (DHS) Program's household and facility-based surveys are two trusted sources of information on these topics. The DHS Program's household surveys, also called Demographic and Health Surveys (DHS), provide extensive demographic data as well as data on health care seeking knowledge and behaviors for a country's population. The DHS Program's Service Provision Assessment (SPA) is a health facility survey that provides a comprehensive assessment of a country's health service delivery.

1.1. Context: Why Link Household and Facility Surveys?

Determinants for care-seeking and utilization are context specific and range from behavioral, cultural, political, and geographic. Among geographic determinants, distance from a health facility is often an important barrier to seeking and accessing care. Distance is not the only geographic barrier to access; rivers, mountains, and poor roads make a household or community's distance from a health facility non-linear or seasonably variable. When an individual, care-giver, or family makes the decision to seek care at a health facility, the nearest facility is not always the facility of choice (Montana et al. 2001). Not all health facilities are equal, and the differences affect care-seeking; these include, but are not limited to:

- *Cost.* Government facilities in many countries provide some or most services free of charge, whereas most private facilities charge user fees. In cases where accessing care at the closest facility incurs a fee, some care-seekers may choose to go to another facility (i.e. government or non-government organization (NGO) facility) further away to receive services that are free or less-expensive (Bourne et al. 2010; Ozawa and Walker 2011; Yanagisawa et al. 2004) .
- *Quality.* Perception of quality of care, including attitudes and behavior of health facility staff toward patients, can either deter or encourage care-seeking generally (Hong et al. 2006; Kyei et al. 2012; Mensch et al. 1996) or at a particular facility.
- *Type of care sought and type of services provided.* In most contexts, different levels of facilities (e.g. health center, hospital) provide a different range of services. For example, emergency obstetric services including delivery by caesarian section is typically not available at the lowest level of facility in many countries, and in some countries may only be available in hospitals. Some women and families may go directly to a higher level facility that is further away (Baker and Liu 2006) for an event such as birth of a child. Similarly, individuals may choose to seek care at facilities further from their home to in order to access specific long-term care or services that may incur stigma, to offer comfort of greater confidentiality via distance from their community. This kind of care-seeking may include accessing long-acting contraceptives (e.g. insertion of intrauterine device (IUD)) or anti-retroviral treatment. Whereas for acute care, such as treatment of wounds or febrile illness, it is more likely that care-seekers will go to the closest health facility.

Geospatial analysis is one tool with which to identify gaps in access and potential barriers to access, all to better target efforts to promote care-seeking and increase use of services. Geographic information systems

(GIS) can help determine where existing services are located and identify the locations of eligible populations that do not have access to, or are not utilizing, services. By geographically linking data from a variety of sources, such as household and facility surveys, we can examine the relationship between individual care-seeking behaviors and the facility service environment. Linking datasets can provide critical information such as access to services of different types and quality, whether services sought are of high quality, and the importance of various characteristics of choices in care-seeking behaviors (Institute for International Programs 2014). Program managers and decisionmakers can use these data to analyze which health facilities need to be improved, where referral services and systems need to be strengthened, and which communities may need to be targeted for outreach or community-based services, etc. (Bailey et al. 2011; Mwaikambo et al. 2014). Further, linked data can be used to monitor (e.g. quality of services, changes in utilization) and evaluate (e.g. did the improvements in quality result in greater utilization or did the intervention reach populations further from facilities) programs.

1.2. Overview: Linking DHS and SPA Data

Separately the DHS and SPA surveys can answer many relevant questions on access, utilization, and quality of services, but being able to link them together may offer even greater insight into why and how individuals access care and how the type or quality of care at the facilities near their homes might influence their actions. Together the DHS and SPA are ideal tools to help researchers, program managers, and decisionmakers answer key questions around program impact and targeting. However, for many years several issues including sampling, geographic data accuracy, data access, and problems with the operationalization of linking these datasets prevented the DHS and SPA data from being used in this manner. In 2013, Skiles et al. presented an in-depth methodological study examining the issues around geographic displacement of DHS clusters and the sampling of SPA facilities (Skiles et al. 2013). Key conclusions from that paper include:

- Using a SPA dataset that is a sample of facilities will result in high misclassification error.
- DHS cluster displacement will result in misclassification if the nearest facility is used for linkage instead of the service environment.
- Service environment linkages using various geospatial techniques are appropriate for use when linking the DHS and SPA data.

The work presented here builds on that extensive methodological study and outlines a number of specific considerations that need to be understood before a link between the DHS and SPA surveys can take place.

- Section 2 outlines the DHS and SPA datasets their structure, sampling, and geographic data.
- Section 3 focuses on the topics within the DHS and SPA that are relevant for linkage and proposes some methods for operationalizing indicators for ease of analysis.
- Section 4 outlines key considerations that need to be examined before proceeding with linking the two datasets.
- Section 5 proposes some geographic linkage methods that are appropriate for use in linking the DHS and SPA.
- Section 6 presents a short case study using the Haiti 2012 DHS and Haiti 2013 SPA datasets.

2. Data Description

Understanding the structure and content of DHS and SPA surveys, as well as the sampling and geographic components of each, is important in identifying how the two can be linked. This section outlines the structure, sampling, and geographic data in the DHS and SPA surveys.

2.1. DHS Household Surveys

There are three DHS household surveys: Demographic and Health Survey (DHS), AIDS Indicator Survey (AIS), and Malaria Indicator Survey (MIS). Since the data structure, sampling, and geographic data are similar across these surveys, they will be referred to generically as DHS household surveys in this document. The DHS surveys provide nationally representative data on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria, and nutrition. The MIS and AIS are smaller surveys with more of a topic focus. The questionnaires are updated every 5 years to reflect new international indicators and trends. Data from the DHS household surveys are widely used to advance global understanding of health and population trends in developing countries and for planning and monitoring development programs.

2.1.1. DHS survey structure and data

The DHS household surveys have three main sections:

1) *Household questionnaire* with household roster

2) *Individual questionnaire*

   - Women: typically all women in the household between the ages of 15-49, in some countries limited to women who have ever been married.
   - Men: typically a sub-sample of men in the household between the ages of 15-49, up to 59 or older in some countries.

3) *Biomarker* data collection varies by country but can include:

   - Anthropometric data: height and weight measurements usually from children 0-5 years old and women 15-49 years old,
   - Hemoglobin measurement to assess anemia status: usually children 6 months to 5 years old and women 15-49 years old,
   - HIV testing: women and men 15-49 years old,
   - Malaria parasite testing: children 6 months to 5 years old, and
   - Other biomarkers including micronutrients, blood pressure, etc.

The component and specific questions included in each questionnaire section depends on the type of survey that is being implemented: DHS, AIS, or MIS. Additionally, a country may decide to include a number of country-specific questions or biomarkers in the survey. The data for each survey is organized

in a number of "recode[1]" datasets that include household file, women's individual file, children's file from the women's questionnaire, and men's individual file[2].

2.1.2. DHS sampling

The DHS household surveys utilize a two-stage cluster sampling design within each sample domain. DHS household survey samples are designed to provide indicator estimates that are nationally representative, as well as representative at the lower level of DHS regions and urban/rural residence. DHS "regions" are subnational units defined for purposes of the survey that usually correspond to existing administrative units or groupings of these units. Increasingly, DHS household surveys are representative at lower levels of administrative units but to date these remain either the first or the second administrative level. The urban/rural "residence" of clusters, as defined by each country's census bureau, is usually also part of the sampling domain. Clusters are preexisting, geographic groupings within the population of interest. These are also called primary sampling units and in the majority of DHS household surveys, census enumeration areas (EAs), as defined by the country's census bureau, delimit the survey clusters. An 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. The population and size of sampled clusters vary between and within countries; typically, clusters contain 100 to 300 households, of which 20 to 30 are randomly selected for survey participation.

2.1.3. DHS geographic data

The DHS Program started georeferencing coordinate data for cluster locations in the late 1980s, but only routinely incorporated this data in surveys beginning in the early 2000s. The DHS Program began making georeferenced GPS (Global Positioning System) datasets available in the public-domain in 2003. The georeferenced datasets can be linked to individual records in DHS household surveys through unique survey identifiers. In addition, as part of the confidentiality process the GPS coordinates of the cluster are randomly displaced (Burgert et al. 2013). The displacement is as follows:

- Urban clusters are displaced a distance up to two kilometers.
- Rural clusters are displaced a distance up to five kilometers, with a further, randomly selected one percent of the rural clusters displaced a distance up to ten kilometers.

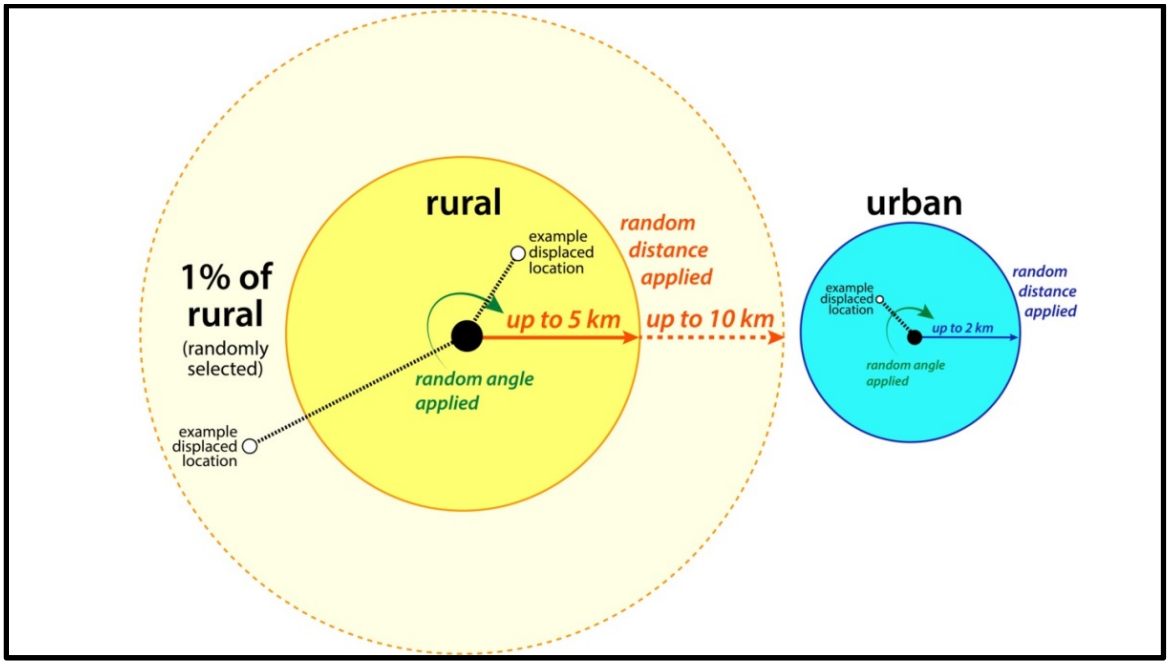
Figure 1 shows an illustration of the displacement. Details on the DHS georeferenced data displacement process and the spatial variability of the resulting data can be found in *Spatial Analysis Report 7* (Burgert et al. 2013). This displacement affects the use of the data in geospatial analysis in several ways which are described in Section 4.1.

---

[1] A recode data file uses a standardized data definition in order to facilitate comparisons across surveys - See more at: http://dhsprogram.com/data/Data-Variables-and-Definitions.cfm

[2] More information on the DHS household survey dataset naming conventions can be found at: http://dhsprogram.com/data/Dataset-Types.cfm

Figure 1. Illustration of DHS GPS data displacement



2.2. SPA Surveys

The SPA examines availability of health services at different types of facilities including human resources, infrastructure, and supplies; the types of support services and processes available; quality of services as compared with accepted standards; and client understanding, satisfaction, and perceptions of care (ICF International 2013). The standard SPA questionnaires were last updated in 2012, and address 10 categories of key services: infrastructure, resources, and systems; child health; maternal and newborn health; family planning; HIV/AIDS; sexually transmitted infections (STIs); malaria; tuberculosis; basic surgery; and non-communicable diseases.

The World Health Organization's Service Availability and Readiness Assessment (SARA) uses a methodology similar to the SPA to collect data on availability of key human and infrastructure resources, basic equipment, medicines, and diagnostics, and on the readiness of health facilities to provide a broad range of key health interventions (WHO 2014). All SARA questions are included in the SPA inventory modules.

2.2.1. SPA survey structure and data

The SPA survey uses four questionnaire types that reflect accepted standards of care and are adapted to fit the information and health needs of each surveyed country. The four questionnaire types are:

1) *Inventory*. The inventory questionnaire is comprised of three modules that collect information on facility-level preparedness to provide services and service-specific preparedness. Specifically, this includes information on infrastructure, including water, electricity, and waste management; service delivery environment, such as staff coverage; logistics and management; and availability, location, and functional status of specific supplies and commodities including drugs and diagnostics.

2) *Observation Protocols for ANC, FP, and sick child consultations.* Using observation protocols, interviewers observe sick child, antenatal care (ANC), and family planning (FP) consultations to assess the extent to which providers adhere to accepted standards of care.

3) *Health Worker (or Provider) Questionnaire.* This questionnaire collects information on qualifications of individual providers, types of services they provide and in-service training they have received.

4) *Exit Interviews for ANC, FP, and sick child consultation.* The "Client Exit Interview" questionnaires are administered after the observed consultation but before the client leaves the facility to assess the client's understanding of their consultation, their recall of instructions and treatment, and perception of the service delivery environment.

The SPA recode datasets are available for surveys conducted after 2004. Each country has a slightly different data structure but generally each of the four survey questionnaires are organized into different datasets that include unique facility identification and provider identification (where appropriate) to link between datasets.

2.2.2. SPA sampling

SPA surveys are conducted among a nationally representative sample of health facilities, a census of health facilities, a census of public and sample of private, or some other combination of census/sample of facilities. A complete list of facilities is necessary for any of these methods, and ideally uses a pre-existing "Master Facility List[3]" developed and maintained by the country's Ministry of Health (or similar agency). Samples are selected taking into account survey sample domain and type of facility, to enable comparison of different operation authorities (e.g. public and private). The inclusion of NGO, faith-based, private, and other non-public facilities varies by country. Survey sample domains are usually the first administrative boundary level (i.e. region, province). In countries where a sample (instead of a census) of facilities is used, the survey typically includes a complete census of hospitals (highest level of care in the country), and then a sample of the subsequent levels of care depending on survey funds and focus.

2.2.3. SPA geographic data

Since 2009, the geographic position of each facility in a SPA survey is released as part of the standard SPA data. The latitude and longitude information for each facility is verified to be within the correct administrative boundary listed in the sample frame. Unlike the DHS data, the SPA GPS locations are not displaced. No GPS data are available for SPA surveys prior to 2009.

[3] A Master Facility List (MFL) is a full national list of public (and in some cases private) health facilities that uniquely identifies each health facility by code, name, and geopolitical hierarchical unit. The list may also contain the GPS location of the facility. An MFL also includes a system for maintaining, updating, and adding to the list over time.

3. Topics That Can be Linked between DHS and SPA Surveys

The framing of the research or policy question of interest is key to determining whether it is appropriate to use data from the DHS and SPA surveys to provide an answer and to determine whether linkage between the two surveys is possible. Some questions may simply be "yes" or "no" questions such as "does an individual live in an area where a certain type of health service (e.g. parasite testing) is offered?" Another might be "is the likelihood of an individual having accessed that type of service higher for those living in that area?" Other questions might be more complex, such as "does the quality of health services at a facility influence the likelihood of an individual using that type of service at that facility?"

An important consideration when framing these questions is to remember that the DHS dataset may provide information on the *type* of facility an individual used but not the *specific facility's name* or *location*. This means that all linkages are probabilistic—based on assumptions about the service environment where an individual lives and the influence that environment may have on their behavior and knowledge. This section summarizes the general topics and creation of variables necessary for linking between the DHS and SPA surveys.

3.1. Relevant DHS Data Topics to Link with SPA Data

There is a wealth of information collected in the DHS that can be examined in relationship to health facilities for programmatic decisionmaking and for research. Generally, information collected can be grouped into three broad categories:

1) *Service Related.* This refers to receipt of specific services and interventions, or receipt of specific commodities. This includes being tested for HIV; being tested for malaria; skilled attendance at birth; health system factors that make it difficult for a woman to seek care for herself or her child(ren) such as distance or cost;
2) *Behavior.* This includes care-seeking for respondent's own health and health of children, the timing of care-seeking, whether or not care was sought at a facility, and whether the first source of care sought was at a facility or elsewhere (e.g. community health worker, traditional birth attendant); man's involvement in maternal and child health care seeking and man's presence at ANC visits; condom use; use of insecticide treated nets; and household behaviors such as handwashing at appropriate times and decisionmaking for care-seeking.
3) *Knowledge.* This ranges from knowledge of disease transmission, prevention, and treatment including knowledge of where to get condoms or long-acting methods of contraception; and knowledge that mosquito bites cause malaria.

Information from both the woman's and man's questionnaire can be used to assess the relationship between distance to a health facility and service related issues, behaviors, and knowledge. Figure 2 summarizes the relevant data collected in the individual woman's and man's questionnaires. Of note, the survey enumerator does collect the name of the facility visited by the survey respondent. This information is used to help properly identify the type and managing authority (public/private etc.) of the facility and to assist in quality control during the survey data collection. The facility name is not entered into the final data entry program due to difficulty in standardizing and verifying the response in this type of open field entry.

Figure 2. Summary of DHS data collection relevant for SPA linkage

Data collected in the DHS Woman's questionnaire:

- Use of contraceptive by type (e.g. sterilization, IUD, injectables, and implants may be most relevant to examine in proximity to facility)
- ANC attendance (yes/no; number of visits; timing of first visit)
- Facility delivery (yes/no; length of stay at facility after birth)
- Postnatal check-up (yes/no; timing after birth; location)
- Newborn check-up (yes/no; timing after birth; location)
- Child fully vaccinated
- Advice or treatment seeking for child with diarrhea (yes/no; first source)
- Treatment for child with diarrhea
- Advice or treatment seeking for child with fever (yes/no; first source; time since onset)
- Tested for malaria (yes/no; location (community vs. facility))
- Treatment for child with fever (Artemisinin-based Combination Treatment (ACT) vs. other)
- Advice or treatment seeking for child with cough and rapid breathing (yes/no; first source)
- Knowledge of where to get condoms or other contraceptives
- Been tested for HIV
- Sought advice or treatment for STI
- Problems with factors that make it difficult to seek care

Data collected in the DHS Man's questionnaire:

- Man's presence at ANC visits
- Condom use during last sex
- Been tested for HIV
- Sought advice or treatment for STI
- Knowledge of where to get condoms (or other contraceptives)

3.2. Relevant SPA Data Topics to Link with DHS Data

SPA datasets are extensive and the variables often benefit from transformations to operationalize their use. Indices and scores can be created using data from one questionnaire or across multiple questionnaires, thus creating variables that are more useful than individual SPA variables. For example, the ANC service readiness indicator is made up of a facility offering ANC services (inclusive of iron supplementation, folic acid supplementation, intermittent preventive treatment for malaria, tetanus toxoid vaccination, and monitoring for hypertension in pregnancy), having ANC guidelines, trained staff, laboratory capacity for hemoglobin and urine testing, and appropriate corresponding medicines (WHO and USAID 2012). Each of the elements necessary to determine whether these indicator components are met is captured in the SPA inventory questionnaire.

The following are four general categories of topic areas for which SPA data can be transformed for linkage with DHS data:

1) *Services Offered*
    - Level of health facility
    - Intervention area (e.g. HIV, delivery care)

2) *Preparedness*

- Score or indices, based on international standards
- Adequate and functional equipment
- Components of care

3) *Quality of services*

- Score or indices, based on international standards

4) *Environment/Perceived quality*

- Client perception of service environment

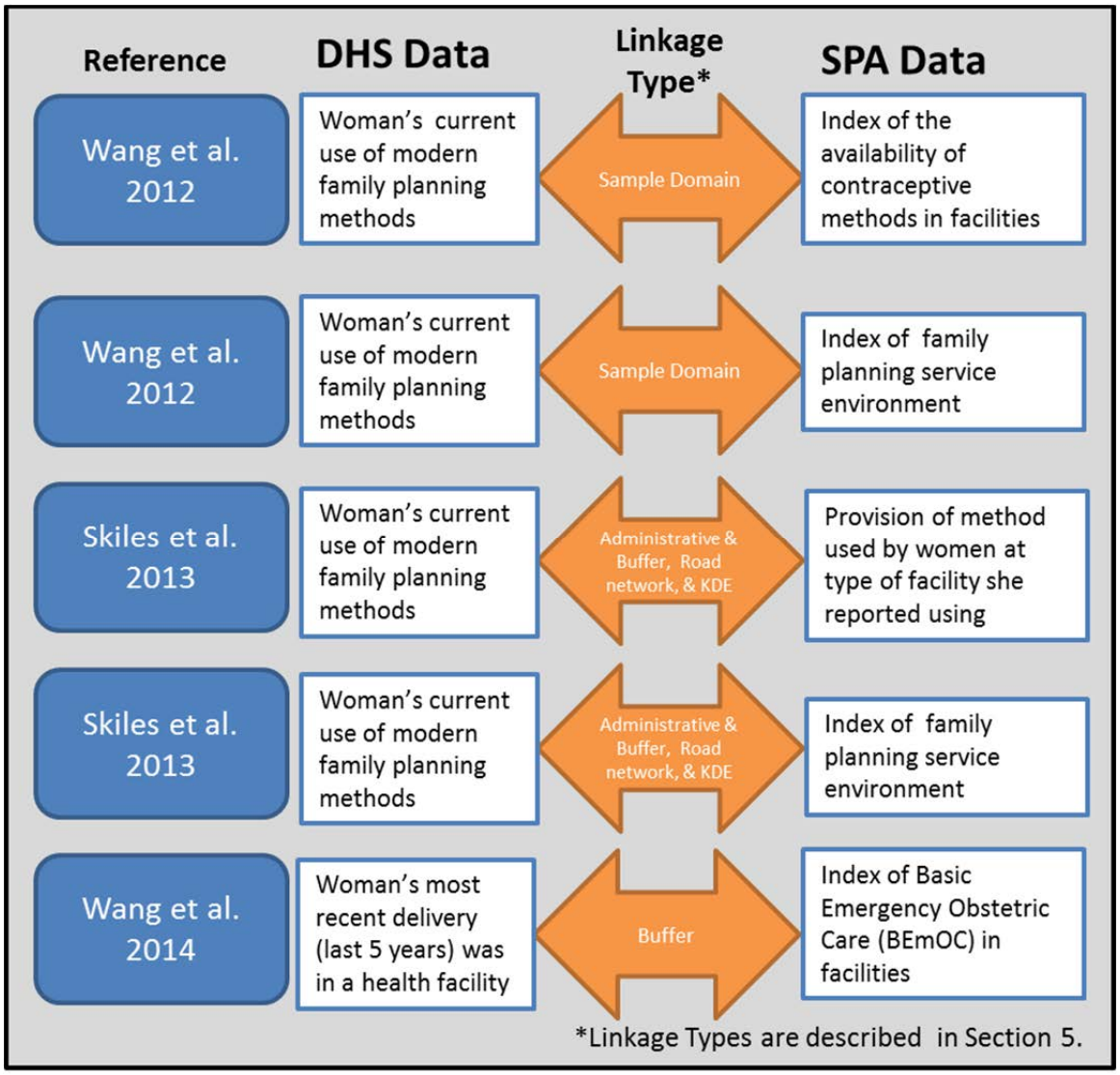
3.3. Operationalization of Data for Analysis

Once the variables for linkage are determined, they need to be operationalized. Parameters can also be set to constrain the linkage to the aggregation of only facilities of the same type or management authority of the one used by the respondent for the service of interest. Two main types of operationalization exist:

1) *Yes/No indicators of availability of a type of service.* Yes/no type indicators can be aggregated by service environment during the linkage process and might also include a count of the number of facilities in the service environment providing that service.
2) *Index or score generation to capture several different indicators together to determine quality or availability of proper materials/services.* For scores or index variables, an average or high/low value can be used.

Three examples of analysis in published journals or reports linking DHS and SPA data are shown in Figure 3. Wang et al. 2012 operationalized SPA data by developing two key independent variables to examine women's use of modern contraceptives using an index of availability of contraceptive methods and family planning service environment score. Data for availability of contraceptive methods comes from the inventory questionnaire, which uses a checklist for each type of contraceptive method and validation of expiration date. Each facility thus has a score that is based on the total number of modern contraceptive methods provided and available. Using DHS sample domains, this is regionally measured as the average number of methods provided and available on the day of the survey (Wang et al. 2012). The service environment score is the total score from a yes/no count (yes = 1, no =0) of 15 variables across four service dimensions: family planning counseling, infection control, pelvic examination, and management practice. The score applies only to facilities that provide family planning services (Wang et al. 2012). Data for these variables come from observations, the inventory questionnaire, and provider interviews. Skiles et al. created a voluntary counseling and testing (VCT) readiness score, based on World Health Organization (WHO) and US Agency for International Development (USAID) service readiness indicators (WHO and USAID 2012). Drawing on data across the SPA questionnaires, seven dichotomous variables were developed to measure counseling and testing, condom availability, and management practices. These were then summed into a composite score (Skiles et al. 2013). Skiles et al. also used the same the Wang et al. 2012 method to create a family planning score. Using data from the Haiti 2013 SPA, Wang et al. 2014 constructed a delivery care readiness score using principal component analysis based on more than 30 basic obstetric care and comprehensive obstetric care indicators defined by the WHO SARA reference manual (WHO 2013).

Figure 3. Example of linkages made in literate between DHS and SPA data



4. Considerations before Linking DHS and SPA Surveys

The technical linkage between the DHS and SPA surveys is easily done in a GIS (see Section 5) but before this is done, it is essential to consider if it is appropriate to make such a linkage at all. This section summarizes key areas of consideration when deciding if a DHS household survey should be linked with the SPA facility survey in a given country. The considerations are dependent on the research question of interest. Often problems with the linkage can be overcome by reframing the question of interest to more appropriately fit the data or country context. For example, moving from a smaller geographic scale analysis to a larger service environment analysis can make linkage more appropriate.

Primarily, if the SPA survey is not a census of facilities nor a census of facilities relevant for the question of interested, then the surveys *should not be linked* at geographic scales below the sample domain. Linking a SPA that is a sample of facilities in the country to a DHS survey causes high misclassification in the subsequent analysis (Skiles et al. 2013). In most SPA surveys, all hospitals are included in the sample, so if the service of interest or the question posed is only relevant for hospitals, then most SPA surveys can be used. Other SPA sampling issues to consider are whether both public and private facilities are included in the sample. Figure 5 outlines key areas of consideration and guiding questions when determining if linking DHS and SPA datasets is appropriate. If the answer to all the questions is "yes" then linkage is possible but limitations may remain. If the answer to one or more of the questions presented is "no" then linkage is not appropriate and the analysis question should be reframed or outcome of interest changed. The areas for considerations are separated into two main groups: data and country context.

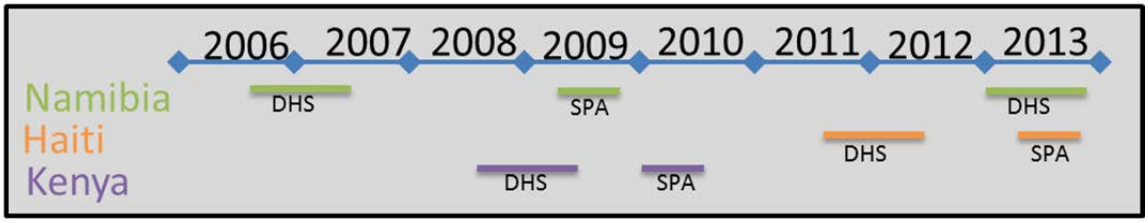
4.1. Data Considerations

Once it is determined that SPA data can be linked (census versus sample of facilities), there remain a number of other data related considerations before deciding that the linkage of the two datasets is appropriate.

DHS/SPA survey dates

The DHS and SPA surveys are usually conducted separately from one another and not during the same time periods. Since service availability, commodity availability, quality of services, and staffing can vary greatly over time, linking DHS and SPA surveys that have occurred at very different points in time will likely introduce a large amount of error into the estimates produced by the analysis. A general rule of thumb might be that surveys that occurred within one year of one another are fairly well matched. Analysis questions that look more generally at availability of services or facility type will likely have less error at different time points. Figure 4 shows three countries and the dates of the DHS and SPA surveys conducted in recent years.

Figure 4. Example of timing of DHS and SPA surveys



DHS data displacement

As described earlier (Section 2.1.3), the GPS locations of the DHS data are geographically displaced so that the survey locations are randomly moved within set parameters of their original location to protect the confidentiality of the respondents. While these geographic datasets are extremely useful for certain types of research, this displacement process makes some types of geographic linkage inaccurate. The amount of error in the linkage depends on urban/rural locations as well as the type of geographical linkage being used (see more on this in Section 5). In general, direct distance linkage (exact distances) or linkages that match a DHS cluster with its closest facility are the most erroneous. In many cases, the question of interest can be restated to look at a larger geographic area or service environment. This results in reduced misclassification error due to the displacement (Skiles et al. 2013).

Missing GPS data

In some surveys, either DHS or SPA, there are missing geographic location data. Usually this is a small amount (less than five percent) of the overall sample. The missing data, however, may not be randomly distributed in the survey; more data might be missing in a specific region, for a specific urban/rural domain (DHS), or for a specific level of facility or management type (SPA). In the case where data is missing from either the DHS or SPA, that location and its associated geographic linkages are lost in the analysis. This may result in misclassification and erroneous conclusions. Missing SPA data is the most problematic since it changes the service environment for the linked DHS cluster. However, sample domain linkage (Section 5.1) can use data with missing GPS.

Indicators of rare events

If the DHS outcome of interest (indicator) is rare, there may not be enough cases in the area to make a statically sound conclusion from the data. Rare events may differ between countries, i.e. an indicator may be rare in one country and common in another (e.g. female genital cutting, malaria prevalence, complications during delivery). The rare event may occur in all clusters but with very few cases so when only 30 households are sampled in a cluster this might not be a large enough sample to find enough cases to give a reliable estimate at this level. The use of statistical significance tests as well as confidence intervals or standard errors will help better understand the error associated with a specific indicator and its appropriateness for linkage.

Indicator reference period

The importance of the reference period and survey timing is specifically relevant for indicators that are temporally bounded or related to the time at which the survey was conducted (e.g. rainy season). Most DHS health seeking indicators have a reference or recall period associated with them. This can vary from a two-week period before the survey to several years in the past. The typical indicator reference period for a SPA readiness variable looks at availability of a commodity or service on the day of the survey. The reference period of the DHS indicator and the relative timing of the surveys is an important limitation in a linkage. However, the linkage analysis can still provide important information.

Locational bias

Locational bias involves attributing an event to one place when in reality it occurred in a different place. This type of bias can be due to an event occurring in a different location from where an individual is currently being interviewed. This bias also occurs if an individual sought services by chance or on purpose outside of their usual service environment. As mentioned in Section 1, this might occur when a service has stigma associated with it (HIV testing or treatment) or when it is an occasional intervention or

non-urgent health need (long lasting contraception or HIV testing), that could be done when visiting a larger town, for example. Though this locational bias may add some bias to an analysis, it is not likely to be high unless the majority of individuals sought care outside of their service environment.

Urban versus rural areas

Urban areas tend to have a much larger number of health facilities and health service options outside of the traditional service provision facilities that are typically included in a SPA survey. These other options might include private medical offices for medical consultation and pharmacies for accessing medicines as well as diagnostics. Urban areas also have many more transportation options that may allow an individual to access health services not in geographic proximity to their home (the location where the survey occurred). The displacement of up to two kilometers in urban areas, likely, also has a larger impact on the service environment than in rural areas. That is that quality of care is likely to change more drastically between neighborhoods in urban areas then between larger rural areas. While rural areas are less likely to have private medical offices, they may have other services such as community health workers who provide basic care in the community.

4.2. Country Considerations

A country's health system and geographic environment can also influence access to and use of health services. It is important that the question of interest in linking the DHS and SPA datasets makes sense within the country context.

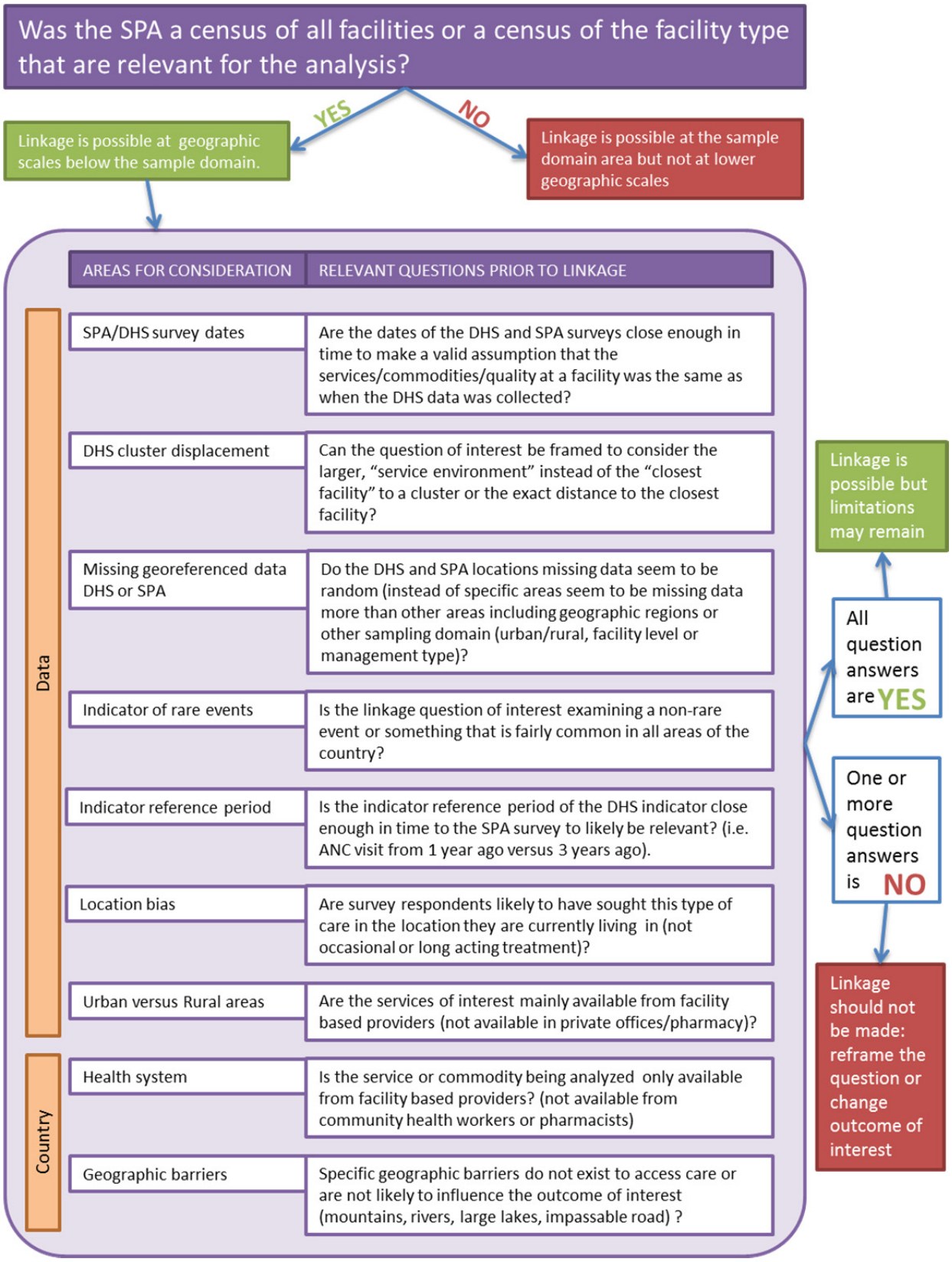
Health system

Several health system factors should be considered when deciding if the question of interest is appropriately answered by the DHS and SPA data linkage. The levels and management authority of the facilities that survey respondents use depends on the services they are seeking. The DHS asks questions on the type of facility someone sought health services at; this can help guide the appropriateness of linking with the data available in the SPA. For example, if the health care providers that are used by the population were not sampled in the SPA then the linkage is not likely to be useful. Similarly, not all health services are available at different levels of facilities, so not all SPA data is necessarily relevant to be linked. A country may have community health workers or other trained individuals who provide diagnostics and drugs outside of a formal health facility, so if the service in question can be found outside of the formal health sector it may not be appropriate to use the DHS and SPA data to draw conclusions about the utilization of those services. Eligibility for services might also be an important consideration in countries with programs such as universal health insurance that dictate where services are available for an individual (e.g. within your district of residence). Finally, cost of care varies within a country's health system both by type of services (maternal and child health services tend to be at low cost or free while other services have a cost) and by management authority of the health facility (public facilities may provide free care while private ones do not).

Geographic barriers

Environmental and manmade barriers can limit individual's access to care. If the outcome of interest in a study is likely to be linked to these factors, it is important to consider external influences that may not be directly linked to DHS or SPA data. Examples might include hills or mountains, rivers and lakes that make it difficult to access care that might be geographically close. Dense forests or national parks may also limit an individual's ability to access a facility. Road networks and seasonal impassability of roads can also play a role in accessing facilities. During rainy season, some roads become impassable making some routes for accessing care difficult or impossible to use.

Figure 5. Linking DHS and SPA data considerations



5. Geographic Linkage Methods

Once the considerations outlined in Section 4 are assessed and the question reframed to be more appropriate for linkage, if necessary, the next step is to determine the appropriate geographic linkage method to use to bring together the DHS and SPA survey data. A more in-depth presentation of methods and steps for linkage of facility location data and population point data is discussed in the guidance document produced by the USAID funded MEASURE Evaluation project "*Using Geospatial analysis to Inform Decision Making in Targeting Health Facility-Based Programs*" (Colston and Burgert 2014). That document is not specific to the DHS and SPA surveys but does give a step-by-step description of key GIS tools that can be used to link population- or household-level data and facility data. Additionally, a systematic review of literature that involves linking population and facility data was done by Brondi et al. in 2014 (Brondi et al. 2014). Both resources may be useful for individuals looking for more information on this topic. This section will provide a general overview of the GIS tools best suited for linking the DHS and SPA datasets.

As mentioned earlier, linkage between the DHS and SPA data should not be done at geographic scales below the sample domain if the SPA is a sample instead of a census of facilities. Linkages that are made using a sample result in upwards of 60 percent misclassification error or bias in the conclusions (Skiles et al. 2013). Additionally, linking a DHS cluster to its closest facility in the SPA may be the easiest and most intuitive linkage, but it is also the linkage method that is most prone to misclassification and error and it *should not* be used. The magnitude of this closest facility misclassification and error is described in Skiles et al. Figure 6 shows an example of the possible different distances measured to Facility A for the original cluster point (5 km) and three possible displaced cluster locations (10, 6.4, and 1 km respectively). In general, the DHS point displacement moves the DHS cluster location to a new location that could be closer or further away from a facility. It can also move the cluster closer to a different facility than it would have been originally linked to had it not been displaced. Key errors in typical DHS and external data (including facility point data) linkages is outlined in *Spatial Analysis Report 8* (Perez-Heydrich et al. 2013). The report outlines basic guidance and tools for overcoming some of the misclassification and errors that exist due to the DHS cluster displacement.

Figure 6. Example of possible distance error in direct distance measurement with displaced location

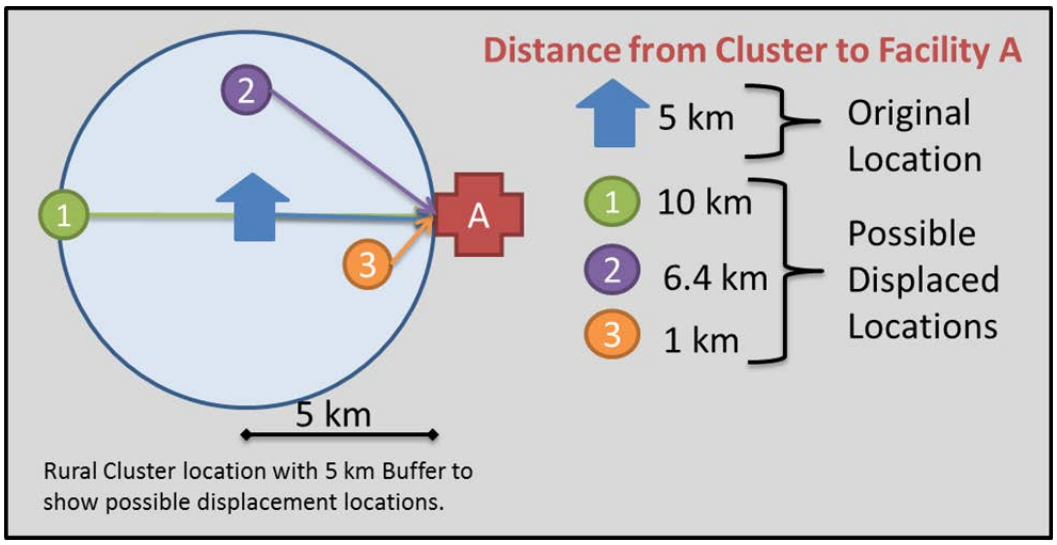


Table 1. Summary of methods appropriate for linking DHS and SPA data

Linking approach	Data inputs	Advantages	Limitations	Notes	Figure 8 panel
Sample domain					
Sample domain	- DHS region data - SPA region data	Can link to sample domain without use of GIS. Does not require a census of facilities, can be done using representative sample.	DHS and SPA domains must be the same or GPS data for one or the other needs to be assigned the domain of the other.	Need - Geographic identifier in dataset – name of administrative region. Do not need GPS data.	--
Service environment					
Administrative boundary	- DHS GPS data - SPA GPS data - Administrative boundary shapefile or geographic Identifiers in DHS SPA datasets	Can link to "service environment" without GIS. Service environment may be representative of that region even if not a census of facilities. (depending on survey sample).	Residents of cluster may not respect admin boundary for care. Administrative areas may be quite large hence; facilities may be far from cluster.	Need - Geographic identifier in dataset – name of admin region. Do not need GPS.	A
Buffer —DHS data	- DHS GPS data - SPA GPS data	Simple GIS tool which is easy to implement.	Without census of facilities links missed. No measurable variation in quality of service within community because all linked to same facilities.	Selection of buffer size should take into account likely local behaver as well as DHS GPS displacement. (min recommended for rural areas is 10 km)	B
Road network	- DHS GPS data - SPA GPS data - Road network data	Allow for linkage with facilities further away that might be used more occasionally e.g. when visiting a market	Requires a high quality road network and information on seasonal passability and if possible local transportation networks (buses etc.)	Road network data; Displacement may move clusters closer/further from a road; Assumes that individuals can travel farther once they have accessed the road due	C

(Continued...)

Table 1. – *Continued*

Linking approach	Data inputs	Advantages	Limitations	Notes	Figure 8 panel
Estimated surface					
Kernel Density Estimation (KDE) —facility data	- DHS GPS data - SPA GPS data - SPA variable to use for density and determinations of kernel size	Can calculate accessibility ratios (pop to facility ratio, pop to staff ratio). KDE disperses effect of facility across space by facility characteristic.	Real reach (kernel) and influence (density) distribution is unknown, this produces a model that is the same for all facilities.	Need-GPS for all facilities; Need-data on facility characteristic that predicts use or reach	D
Spatial interpolation —DHS data	- DHS GPS data - SPA GPS data - DHS indicator summarized at the cluster level to be used for interpolation	Produces indicator estimates of survey data at smaller geographic levels than the survey sample region.	Error measurements and propagation of error across the surface is important for proper use of the data.	Best used with facility catchment information	E
Catchment area					
Buffer or thiessen polygon —facility data	- DHS GPS data - SPA GPS data - Determination of facility catchment area (buffer or thiessen polygons or other)	Focus is on the behavior within the intended catchment area of a specific facility, could be used if SPA was only a sample.	Some facilities will not have any DHS clusters within their catchment area.	Catchment areas for facilities are rarely well defined include GIS boundaries, so some assumptions such as buffer distance need to be used.	F

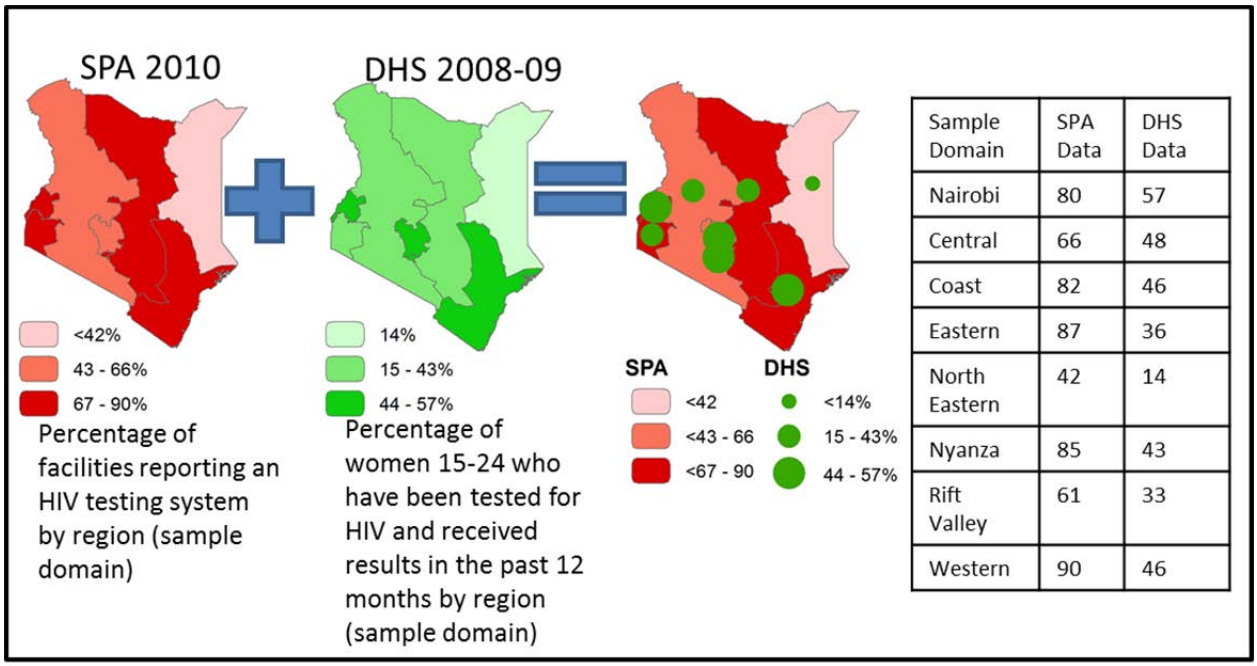
5.1. Recommended Linkage Methods

Many geographic linkage methods could be used for linking the DHS and SPA datasets but not all methods should be used. Table 1 summarizes key information on the linkage methods that can be used while linking the DHS and SPA datasets; while Figure 8 shows examples of these linkages.

Sample domain linkage

The least technical method for linking the DHS and SPA data is to link the surveys at the level of the ***sample domain*** for the surveys. This method can work well when both surveys have the same geographic regions as part of their sampling domains. However, it is important to check that these regions are actually the same by reviewing the sampling information and map in the survey final report and the boundaries comparison on the *DHS Spatial Data Repository* website (spatialdata.dhsprogram.com). Some countries may use the same names for the regions but they may have very different geographic boundaries over time. The two key advantages to using the sample domain method is that the sample is selected to be representative at the sample domain level and the sample domain variables usually already exist in the dataset; this method does not require use of GIS software. The main disadvantage to this method is that the sample domains are typically large geographic regions within a country and thus this high-level linkage hides differences that may exist within sampling domains.

Figure 7. Example sample domain linkage



Sample Domain	SPA Data	DHS Data
Nairobi	80	57
Central	66	48
Coast	82	46
Eastern	87	36
North Eastern	42	14
Nyanza	85	43
Rift Valley	61	33
Western	90	46

Service environment linkage

Several DHS-SPA appropriate linkage methods decrease misclassification bias and other errors by focusing on the service environment around a DHS cluster instead of focusing on a single facility within a certain distance. Figure 8 shows three examples of this method: ***administrative area*** (Panel A), ***buffer*** (Panel B), and ***road network*** (Panel C). This type of "service environment linkage" assigns a summary

(average, minimum, maximum, or count) SPA value to the DHS cluster or an individual or household within a cluster.

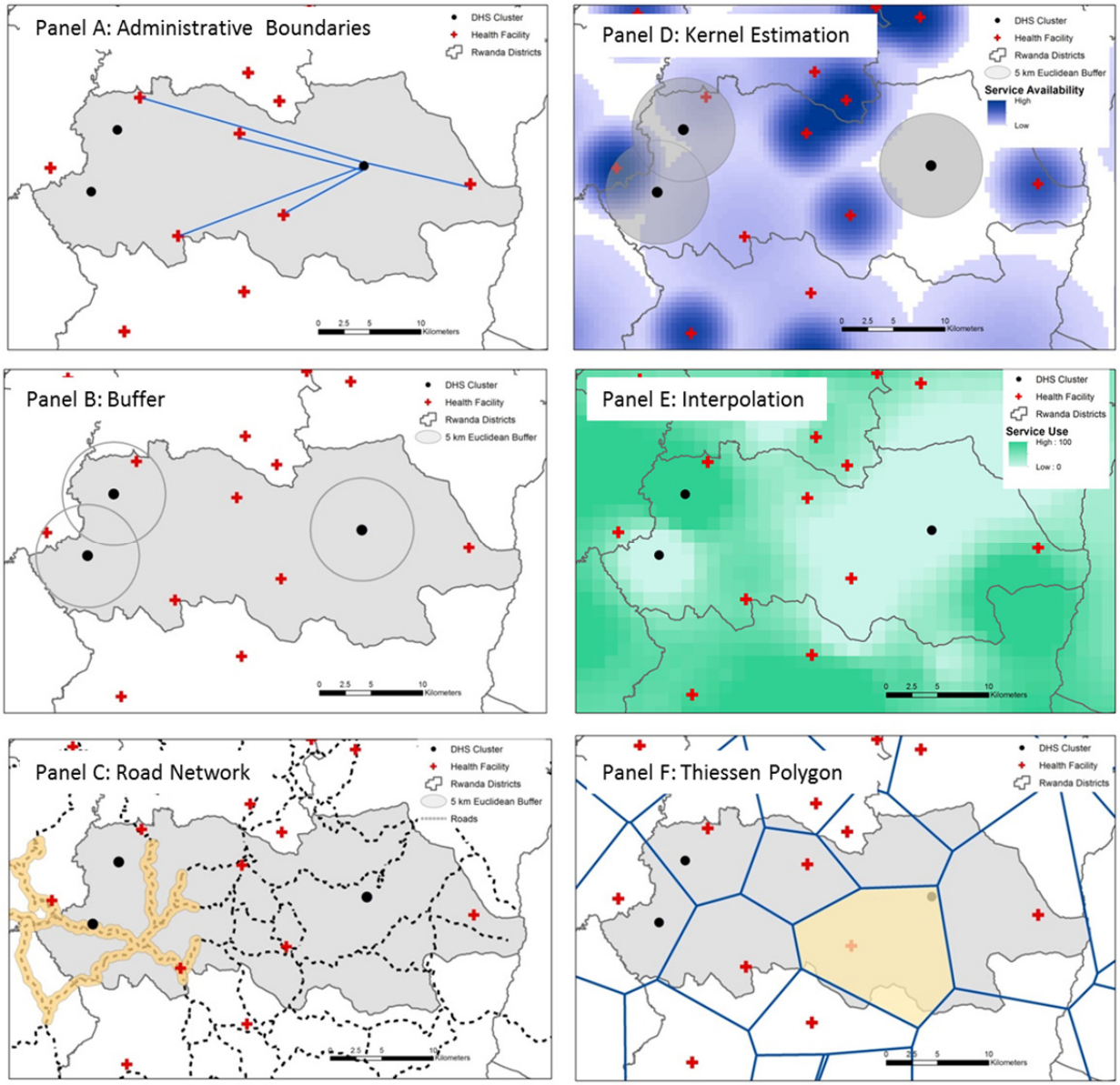
Estimated surface linkage

The most technically difficult linkage method to use when linking the DHS and SPA data involves geo-spatial estimation techniques. ***Kernel Density Estimation*** (KDE) creates a projected map surface that visually describes the estimated influence of a given facility over space. The further away a facility is located from a cluster, the less influence it has and the lower the density of the facility score at the cluster location. Panel D in Figure 8 shows an example of this type of surface created using a service availability score. Once the KDE is created, a buffer extraction can be used to determine the average value around a DHS cluster. While KDE in this case uses facility-level data, DHS data can also be used to create estimated surfaces. These surfaces are created using ***spatial interpolation*** techniques include Kriging and Geo-Bayesian statistics. These techniques draw on information across space to estimate the values at locations that the DHS survey did not take place (Figure 8 Panel E). Guidelines around creating interpolated surfaces with DHS data are presented elsewhere (DHS Spatial Interpolation Working Group 2014). This surface can then be used in a catchment area linkage method (see next bullet).

Catchment area linkage

Typically, questions that are asked in linking the DHS and the SPA start with the individual/household focus and use facility characteristics to draw conclusions about why a certain outcome exists for the individual. Another way to phrase the questions focuses on the facility and uses the DHS information to understand the population that the facility is meant to serve. This linkage method requires known catchment areas or creation of estimated catchment areas around a facility. Then DHS clusters that fall in that catchment area are assigned to that facility. Panel F in Figure 8 shows an example of this method using estimated catchment areas created through ***thiessen polygons***. Buffers of particular sizes depending on the estimated draw of a facility type can also be used; for example, a hospital might draw patients from a further distance than a dispensary (See Colston and Burgert 2014 for more on this method). Instead of assigning specific DHS clusters to facilities, an interpolated surface (see previous bullet) can be created and the value of the surface contained in the catchment area assigned to the facility.

Figure 8. Visualizations of linkage methods

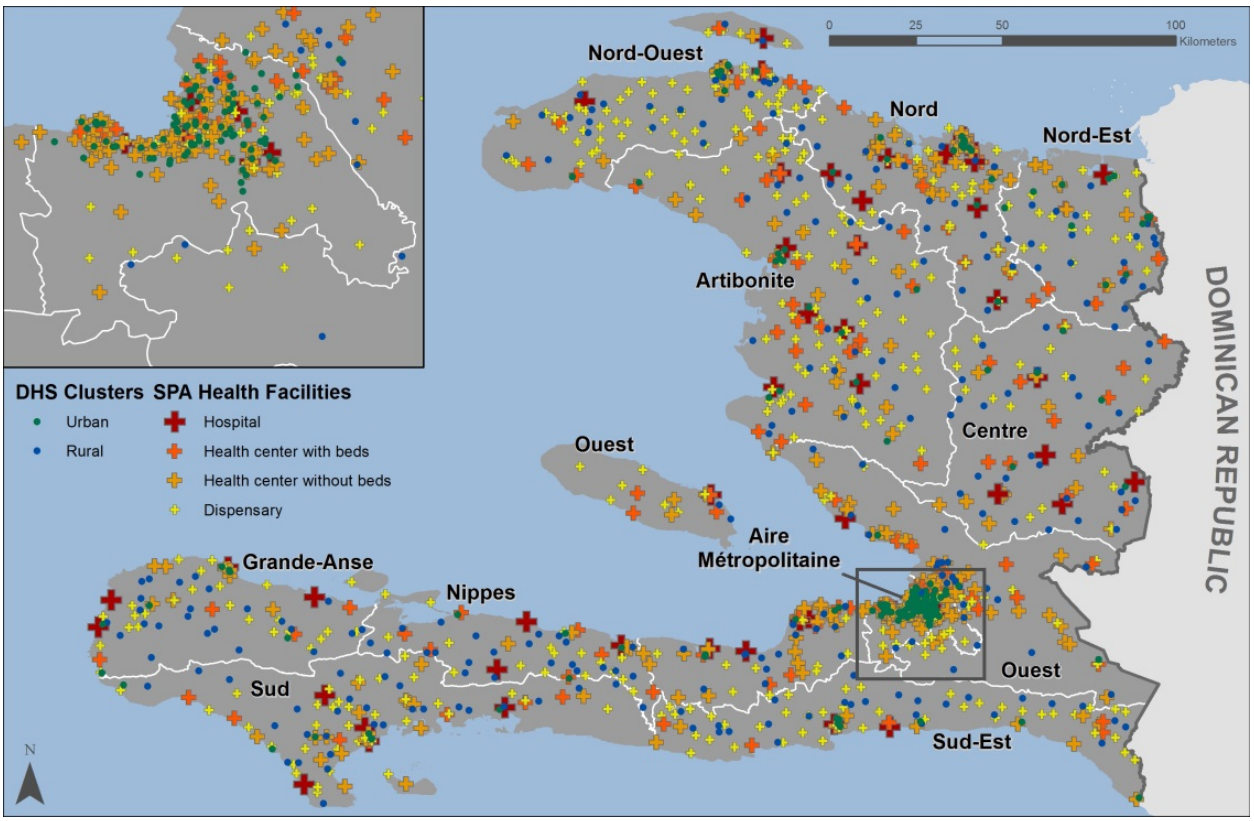


6. Case Study: Haiti Birth Delivery Services

The process of framing an appropriate question to be answered using the DHS and SPA datasets, and selecting a linkage method, was laid out in sections 4 and 5. This section demonstrates a possible question and linkage method applied to the Haiti 2012 DHS data and the Haiti 2013 SPA data. The SPA was a census of all facilities. The two surveys took place within a year of each other with the DHS fieldwork January to June 2012 and the SPA fieldwork March to July 2013. The linkage between the surveys was done using 10-kilometer buffers to estimate the service environment around the DHS cluster locations.

The program question of interest in this case study was: "Are women more likely to deliver in a health facility if they live in a service environment where there is at least one facility providing high quality obstetric care?"

Figure 9. Haiti DHS and SPA survey locations



6.1. Methods

The DHS survey included 400 clusters, plus an additional 45 clusters from camps composed of internally displaced persons (Institut Haïtien de l'Enfance and MEASURE DHS 2012). A total of 5,515 women between the age of 15 and 49 who had given birth in the five years preceding the survey were interviewed. The SPA survey was a census of health facilities that collected data on 907 public and private health facilities from hospitals down to the dispensary level (Institut Haïtien de l'Enfance and

MEASURE DHS 2013). Figure 9 shows the distribution of DHS cluster locations (displaced for confidentiality) and SPA facilities from the surveys.

Variable operationalization

For this case study, SPA data on quality obstetric delivery services were operationalized through identification of health facilities that provided the six signal functions for basic emergency obstetric care (BEmOC) (WHO 2013). These six functions are:

- Administration of antibiotics
- Administration of uterotonic drugs/oxytoxics
- Administration of anticonvulsants
- Manual removal of placenta
- Assisted vaginal delivery
- Removal of retained products

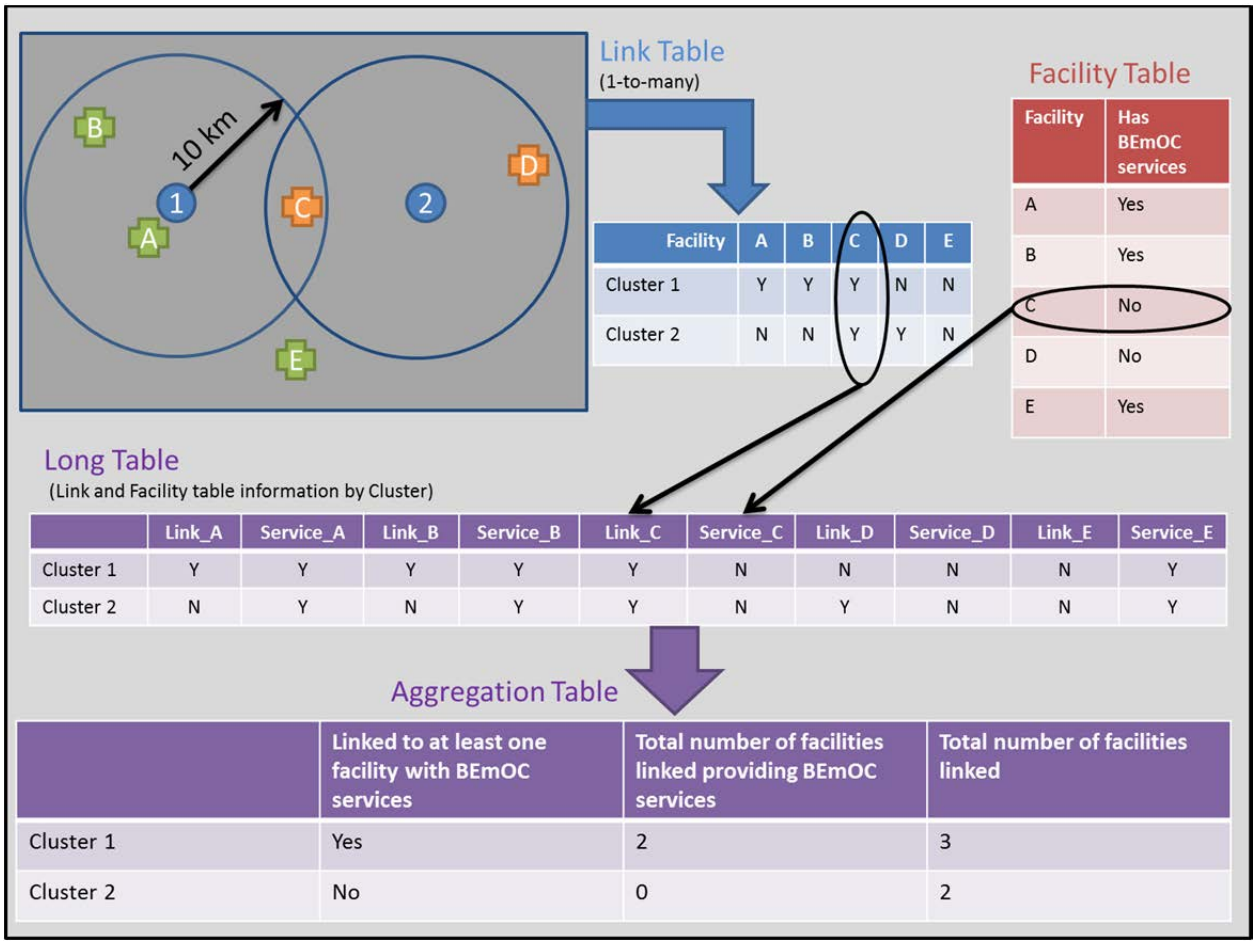
Facilities that provided all six functions were classified as providing BEmOC; facilities that provided zero to five of these functions were classified as *not providing BEmOC*.

DHS data was operationalized to identify women who had given birth in the last five years in a health facility. Facility classification was not harmonized between the DHS and SPA survey instruments so it was not possible to link a woman who reported using a health facility type in the DHS survey to the same type of facility in the SPA survey.

Linkage operationalization

The service environment linkage between the DHS and SPA GPS data was operationalized using a 10 km buffer. The steps to create the final aggregated table used for analysis are illustrated in Figure 10. First, within a GIS software all facilities within 10 km of the cluster are listed as linked ("Y" for yes in the "Link Table"). Second, the facility-level data on BEmOC availability was linked to each health facility in the "Long Table." Third, the data were summarized to indicate whether at least one facility linked with the cluster provided BEmOC services ("Facility Table"). Fourth, an "Aggregation Table" was created summarizing the data from the "Long Table." Finally, the "Aggregation Table" is then linked to the individual women's information on health facility delivery for analysis (not shown in figure).

Figure 10. Illustration of linkage steps



Facility	A	B	C	D	E
Cluster 1	Y	Y	Y	N	N
Cluster 2	N	N	Y	Y	N

Facility	Has BEmOC services
A	Yes
B	Yes
C	No
D	No
E	Yes

	Link_A	Service_A	Link_B	Service_B	Link_C	Service_C	Link_D	Service_D	Link_E	Service_E
Cluster 1	Y	Y	Y	Y	Y	N	N	N	N	Y
Cluster 2	N	Y	N	Y	Y	N	Y	N	N	Y

	Linked to at least one facility with BEmOC services	Total number of facilities linked providing BEmOC services	Total number of facilities linked
Cluster 1	Yes	2	3
Cluster 2	No	0	2

6.2. Results

The percentage of DHS clusters linked to at least one health facility from the SPA that meet the BEmOC requirements is shown in Figure 11. Over 91 percent of urban clusters have access to a BEmOC facility within the 10 km buffer radius. Rural clusters, on the other hand, have more limited access: only 55 percent are within the 10 km buffer.

The percentage of women delivering at a health facility in the last five years who live within 10 km of at least one facility providing BEmOC is 47 percent (Figure 12). Women in urban or rural areas are somewhat more likely to deliver in a health facility if they are linked to a health facility providing BEmOC services. Although, in general, women in rural areas have a smaller proportion of health facility deliveries.

While Figure 11 shows access to a BEmOC facility, Figure 12 shows utilization of health facility by individual women according to residence.

Figure 11. Percentage of DHS clusters having a BEmOC facility within 10 km, by residence

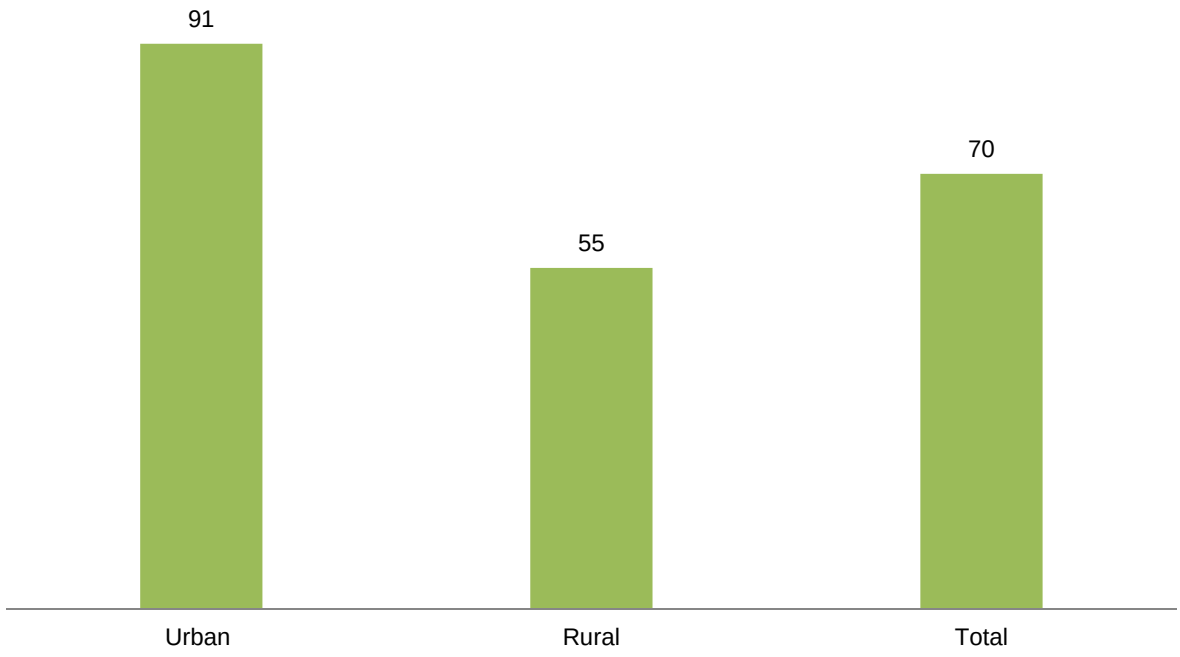
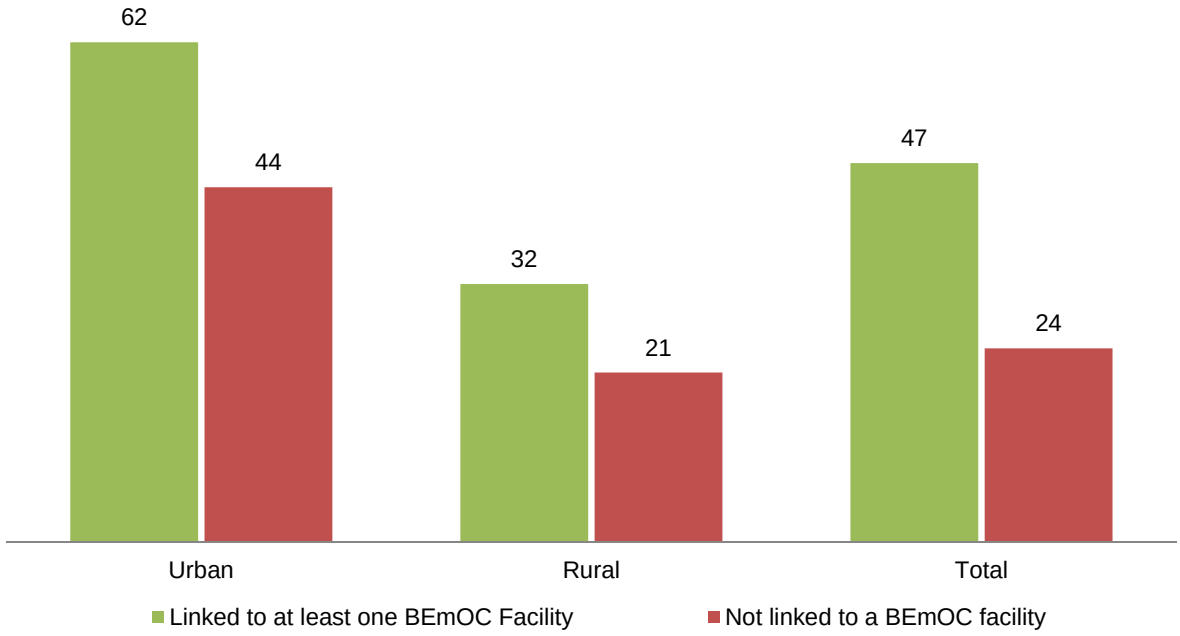


Figure 12. Percentage of women delivering at health facility in the past five years who live within 10 km of at least one facility providing BEmOC, by residence



6.3. Discussion

This case study linked DHS and SPA datasets from Haiti to examine women's decision to give birth in a health facility which provides BEmOC. Many factors likely influence a women's decision to give birth in a health facility; this study looked at only one possible factor. For women who went to a facility to deliver, it is not possible to know if they went because of the availability of the BEmOC services or for other reasons. Additionally, it is not possible to know if they went to a specific facility used in the linking, another facility within the service environment, or a facility outside of the service environment definition used in this case study. Other factors such as education level, household wealth status, and empowerment (decisionmaking power), along with the BEmOC standards and an operationalized score, are examined more thoroughly in a separate report (Wang et al. 2014).

This Haiti case study highlights the importance of stratifying results by urban-rural residence. The results show that access to and utilization of health facilities varies substantially depending on the women's residence. Looking at just access to a facility that provides BEmOC services, the DHS survey findings showed that only 55 percent of the respondents living in rural areas were within 10 km of a facility providing BEmOC services, compared with 91 percent of respondents living in urban areas.

7. Drawing Conclusions from DHS and SPA Linkage

This report has outlined the benefits that can be achieved by linking DHS and SPA survey data. Such linkage has the potential of generating answers to key programmatic and policy questions. However, many considerations have to be taken into account when linking these data, particularly geographic limitations. Framing the question of interest and appropriately operationalizing DHS and SPA variables are key steps in creating an appropriate linkage between the two survey datasets.

The Haiti case study presented in section 6 describes one approach taken to exploring the availability of delivery services in the service environment within which pregnant women live, and assessing women's care-seeking or utilization of health facility delivery services. This is one piece in a larger analysis designed to assist decisionmakers in understanding the factors associated with women's utilization health facilities for delivery. Other, non-facility-related factors likely influence a woman's decision to deliver in a health facility including level of education and household wealth status. The appropriate linkage of the DHS and SPA data is one component in an analysis that may require other advanced statistical methods such as multivariate regression models. While clearly there are limitations to creating an appropriate link between the DHS and SPA datasets, the opportunity for linkage exists and allows for better understanding of the data.

Looking beyond DHS and SPA surveys, program managers and policymakers will probably want to link data from other program-specific surveys, facility and household or population-based surveys, or data from routine health information systems. The MEASURE Evaluation document *"Using Geospatial analysis to inform decision making in targeting health facility-based programs: A Guidance Document"* (Colston and Burgert 2014) describes ways to leverage various types of data to improve understanding of the dynamics between populations and facilities within a country. Alternatively, household and facility data collection methods specific to the researcher's question of interest may be possible (time and funding permitting). If you are preparing a household or facility survey and anticipate linking them together, several alternative sampling methods may be considered that allow for expansion of the types of linkages possible; however, they may also limit the representativeness of the population survey, the facility survey, or both. Some descriptions of these methods and their pros and cons are presented in the conclusion of the Skiles et al. paper (2013).

This document has described the considerations to be made before linking the DHS and SPA data and methods for creating linkages between DHS and SPA datasets. It is important when interpreting the results of such analysis to be cognizant of limitations to the questions posed and the data used. These data and linked analyses may be a starting point for further investigation of care-seeking behaviors, access to health services, and issues of quality of services that require other and more advanced quantitative and qualitative studies.

References

Bailey, P.E., E.B. Keyes, C. Parker, M. Abdullah, H. Kebede, and L. Freedman. 2011. "Using a GIS to Model Interventions to Strengthen the Emergency Referral System for Maternal and Newborn Health in Ethiopia." *International Journal of Gynecology & Obstetrics* 115(3): 300-9.

Baker, J.B., and L. Liu. 2006. "The Determinants of Primary Health Care Utilization: A Comparison of Three Rural Clinics in Southern Honduras." *GeoJournal* 66(4): 295-310.

Bourne, P.A., D. Eldemire-Shearer, T.J. Paul, J. Lagrenade, and C.A. Charles. 2010. "Public and Private Health Care Utilization Differences between Socioeconomic Strata in Jamaica." *Patient Relat Outcome Meas* 1: 81-91.

Brondi, L., H. Nair, and H. Campbell. 2014. *Linking Houhold Surveys and Assessments of Service Provision in MNCH: A Review*. Eduburgh, Scotland: Centre for Population Healht Sciences the University of Edinburgh.

Burgert, C.R., J. Colston, T. Roy, and B. Zachary. 2013. *Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys*. DHS Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International.

Colston, J., and C.R. Burgert. 2014. *Using Geospatial Analysis to Inform Decision Making in Targeting Health Facility-Based Programs: A Guidance Document*. Chapel Hill, North Carolina, USA: MEASURE Evalatuion.

DHS Spatial Interpolation Working Group. 2014. *Spatial Interpolation with Demographic and Health Survey Data: Key Consideration*. DHS Spatial Analysis Reports No. 9. Rockville, Maryland, USA: ICF International.

Hong, R., L. Montana, and V. Mishra. 2006. "Family Planning Services Quality as a Determinant of Use of IUD in Egypt." *BMC Health Serv Res* 6: 79.

ICF International. 2013. *Service Provision Assement Survey Interviewer's Manual*. DHS Basic Documentation No. 2. Calverton, Maryland, USA: ICF International.

Institut Haïtien de l'Enfance, and MEASURE DHS. 2012. *Enquete Mortalite, Morbidite et Utlization des Services 2012*. Calverton, Maryland, USA: ICF International.

Institut Haïtien de l'Enfance, and MEASURE DHS. 2013. *Evaluation de la Prestation des Services de Soins de Sante 2013*. Rockville, Maryland, USA: ICF International.

Institute for International Programs, J.H.U. 2014. *Linking Household Surveys and Health Service Assessments: Technical Consultation on Study Design*. Paper read at Improving Coverage Measurement for MNCH (ICM) at Baltimore, Maryland, USA.

Kyei, N.N.A., O.M.R. Campbell, and S. Gabrysch. 2012. "The Influence of Distance and Level of Service Provision on Antenatal Care Use in Rural Zambia." *PLoS ONE* 7(10): e46475.

Mensch, B., M. Arends-Kuenning, and A. Jain. 1996. "The Impact of the Quality of Family Planning Services on Contraceptive Use in Peru." *Stud Fam Plann* 27(2): 59-75.

Montana, L., R. Bessinger, and S. Curtis. 2001. *Linking Health Facility and Population Level Data in Kenya Using Geogrpahic Information Systems*. Calverton, Maryland, USA: ORC Macro.

Mwaikambo, L., B. Wilkes, and N. Jayarajan. 2014. *Mapping: A Skill for the Future*. Available online at http://measureevaluation.wordpress.com/2014/07/30/mapping-a-skill-for-the-future/

Ozawa, S., and D.G. Walker. 2011. "Comparison of Trust in Public vs Private Health Care Providers in Rural Cambodia." *Health Policy Plan* 26(Suppl 1): i20-9.

Perez-Heydrich, C., J.L. Warren, C.R. Burgert, and M. Emch. 2013. *Guidelines on the Use of DHS GPS Data*. Spatial Analysis Reports No. 8. Calverton, Maryland, USA: ICF International.

Skiles, M., C.R. Burgert, S. Curtis, and J. Spencer. 2013. "Geographically Linking Population and Facility Surveys: Methodological Considerations." *Population Health Metrics* 11(1): 14.

Wang, W., S. Wang, T. Pullum, and P. Ametepi. 2012. *How Family Planning Supply and the Service Environment Affect Contraceptive Use: Findings from Four East African Countries*. DHS Analytical Studies No. 26. Calverton, Maryland, USA: ICF International.

Wang, W., M. Winner, C.R. Burgert, and J. Colsotn. 2014. *Influence of Service Readiness on Use of Facility Delivery Care: A Study Linking Health Facility Data and Population Data in Haiti*. DHS Working Papers No. 114. Rockville, Maryland, USA: ICF International.

World Health Organization (WHO). 2013. *Service Availability and Readiness Assesment (SARA): An Annual Monitoring System for Service Delivery, Reference Manual*. Geneva, Switzerland: WHO.

World Health Organization (WHO). 2014. *Service Availability and Readiness Assessment (SARA)*. Available online at http://www.who.int/healthinfo/systems/sara_introduction/en/

World Health Organization (WHO), and US Agency for International Development (USAID). 2012. *Measuring Service Availability and Readiness*. Geneva, Switzerland: WHO.

Yanagisawa, S., V. Mey, and S. Wakai. 2004. "Comparison of Health-Seeking Behaviour between Poor and Better-off People after Health Sector Reform in Cambodia." *Public Health* 118(1): 21-30.