



DEPARTMENT OF GLOBAL HEALTH

UNIVERSITY *of* WASHINGTON

PhD in Global Health

Metrics Track

Defining a bottom-up approach to improve HMIS data use in low resource settings

Grégoire Lurton



IHME

Institute for Health Metrics
and Evaluation

Abstract This proposal defines a project for a doctoral research into data processing and analysis methods adapted to Health Information Systems (HIS) in developing countries. I argue that HIS in developing countries have been historically marked by top-down approaches in which methods are defined by remote actors such as foreign or international specialists, and implemented without much local adaptation. This way of creating health information is at odds with a history of local emergence of epidemiological and public health methods, and with the increasing flexibility of data processing and computing techniques.

My research aims at defining data centered methods for local adaptation of health information systems. I will concentrate on three important dimensions of what a bottom-up approach to HIS could be. First, I will define an approach to local indicator definition. Using EMR data from HIV care in different countries, I will show how the most common metric used for patients retention in HIV care is impacted by data quality and local characteristics, and I will explore robust alternatives to this metric. Second, using widely available data from Niger, I will implement a data-hybridization strategy to produce an actionable map of Niger population. Finally, I will use routine data from a Results Based Funding program in Bénin to define an algorithm for data quality screening and facility performance monitoring adapted to local available data.

This project will contribute to research in the Global Health Metrics field in different ways. It will contribute to the research, mainly from the ICTD community, on flexible standards for data collection. It will offer a framework for the analysis and use of real-world data in low-resource settings. Finally, it will contribute to statistical methods for health system data surveillance and monitoring. As a whole, it will provide important contributions to the definition of localized approach to generate information for decision makers in health systems.

Contents

Table of Content	iii
List of Figures	iv
Acronyms	v
1 Introduction	1
1.1 General Overview	1
1.2 Work hypothesis	2
1.3 Aims	3
1.4 Novelty and scientific contribution	4
1.4.1 Local Adaptation and flexible standards	4
1.4.2 Imperfect data Usage	5
1.4.3 Statistical methods for health data monitoring	5
1.5 Timeline	6
2 Flexible Standards	8
2.1 Theoretical Framework	9
2.1.1 HIV care process	9
2.1.2 Data Entry	10
2.1.3 Loss to Follow Up definition	10
2.2 Data	12
2.3 Methods	12
2.4 Timeline	13
3 Data hybridization	15
3.1 Data	16
3.2 Methods	17
3.2.1 Name Matching	17
3.2.2 Locality mapping	17
3.2.3 Population modeling	18
3.3 Output	19
3.4 Timeline	19
4 Data Monitoring	20
4.1 Research questions	20
4.2 Data	21
4.3 Methods	22
4.3.1 Theoretical Framework	22
4.3.2 Data Screening	23
4.3.3 Threshold Definition	24
4.3.4 Risk Classification	25
4.3.5 Validation and Supervision Planning optimization	26
4.4 Timeline	26
References	28

List of Figures

1	HIS Processes and corresponding dissertation aims	3
2	Dissertation Timeline	7
3	Distribution of l_v^i in our data.	9
4	Distribution of f_v^i in our data. We see the visits are scheduled mainly 7, 28, 56 or 84 days after the previous visit	10
5	Distribution of δ_v^i in our data. Most of the data is entered in the first week following the visit, but we see some data can be entered much later	11
6	Individual follow-up and data entry process	11
7	Gantt Chart for Aim 1	14
8	Mapping of Niger population from AFRIPOP	15
9	Comparison of standardized adult age distribution between the 2012 census and the 2016 voting lists	18
10	Gantt Chart for Aim 3	19
11	Distribution of the percentage of correct indicators by report	21
12	Distribution of payment correction percentage for a sample of indicators	22
13	Evolution of payment correction percentage for a sample of indicators	23
14	Gantt Chart for Aim 2	27

Acronyms

ART Antiretroviral Treatment. 8

CEAM Cost Effectiveness Analysis Microsimulation. 12

CENI Comission Electorale Nationale Independante. 16

CUSUM Cumulative Sum. 24

DHIS2 District Health Information Systems. 4, 5, 17

EMR Electronic Medical Records. 4, 5, 8, 10, 12, 13

EWMA Exponentially Weighted Moving Average. 24

ICT Information and Communication Technologies. 4, 5

INS Institut National de la Statistique. 16

LTFU Lost To Follow Up. 3, 4, 8, 10–13

M&E Monitoring and Evaluation. 2

OIF Office International de la Francophonie. 16

OSM OpenStreetMap. 17, 18

PMA Paquet Minimum d'Activités. 21

PRSS Programme de Renforcement du Système de Santé. 20

RBF Results Based Financing. 20, 24

RENACOM Répertoire National des Communes. 16–18

RENALOC Répertoire National des Localités. 16–19

SNIS Système National d'Information Sanitaire. 17

SPC Statistical Process Control. 5, 6, 20

1 Introduction

1.1 General Overview

Evidence based decision making is widely presented as the best way to design public health policies at national and local levels ([Abou-Zahr & Boerma 2005](#), [Shibuya et al. 2005](#), [Bambas Nolen et al. 2005](#), [Mutemwa 2006](#), [Boerma 2013](#)). The ability of statistics to offer rational basis to make decisions ([Desrosières 1993](#), [Porter 1996](#)), allied with progress in data collection and analysis techniques engender enthusiasm and hope in the ability of governments and other actors to design and implement analytical tools to improve the political and administrative processes.

Meanwhile, if this principle seems to be set, what constitutes relevant evidence in an administrative context, and how it should be generated is less clear and depends on local contexts and traditions. Bergeron and Cassel note that knowledge and expertise of public health deciders are "a network involving many different actors, structures and tools, concepts and spatial and institutional arrangements" ([Bergeron & Castel 2014](#)). Producing knowledge for public health decision thus relies on localized and contextual approaches, that are not easily generalized to other situations. In the world of Global Health Metrics, the local and contextual dimension of statistical systems can sometimes be lost in an undue generalization. Meanwhile, the historical conditions of emergence of evidence generation systems for public health are very different in different parts of the world.

In most rich countries, a history of public statistics development resulted in the emergence and stabilization of well defined data sources, of methods and of roles along which the generation of public statistics are produced. In these countries, there seems to be a convergence between the two values noted by Desrosières for statistics : statistics as a measure of reality, and statistics as a norm ([Desrosières 1997, 2014](#)). Meanwhile, this should not hide the important historical debates and struggles that happened around simple objects such as averaging population statistics ([Desrosières 1993, Porter 1996](#)), measuring hospital mortality ([Beyersmann & Schrade 2017](#)) or even collecting hospital data ([Chaperon et al. 1988](#)). As such, strong public health information systems really are the result of local equilibriums and of adaptations to local statistical cultures, but were started by individuals making ad-hoc use of different data sources and inventing and standardizing methods on the run ([Lecuyer 1987](#)).

In sub-Saharan Africa, evidence generation for public health policy makers is in a very different situation. The weakness of statistical systems in sub-Saharan Africa has been well described and documented ([Jerven 2013](#)). Most sub-Saharan countries can't rely on a robust vital statistics system to plan population targeted intervention. National Health Information Systems are often considered weak ([Abou-Zahr et al. 2010](#), [Kiberu et al. 2014](#)), and are affected by multiple uncoordinated demands for data collection and reporting. These demands may come from local authorities or from international donors and partners, which jeopardizes their accuracy ([Sandefur & Glassman 2013](#)).

The development of these top-down statistical systems can be traced, in a long perspective, to the enforcement of a specific mode of administrative control by colonial powers in the XIXth century ([Appadurai 1996](#), [Cordell et al. 2010](#), [Gervais & Mandé 2010](#)). Colonial statisticians, often weakly skilled and trained ([Kateb 1998](#), [Cordell et al. 2010](#)), nonetheless set the nomenclatures and conventions around how land and populations would be described and analyzed ([Rambert 1922](#), [Gervais & Mandé 2010](#)). In the meantime, whereas European statistical systems were developed and structured by social activists in the context of the design of the first modern welfare systems ([Desrosières 1993, 1997](#)), colonial statistical systems were geared towards efficient land administration and exploitation ([Rambert 1922](#), [de Martonne 1931](#)) with little interest on population. In Sénégal in 1926, the first population estimation was made thanks to the availability of taxation records, but with no vital statistics records ([Rousseau 1929](#)).

After the decolonizations, the administrative mindset of evidence generation was completed by the expert mindset of the developmental state ([Bonneuil 2000](#)). Developmentalist expertise transformed African societies into objects of external studies and knowledge jeopardizing the ability of administering political entities to emerge, and prolonging the outwards orientation of colonial statis-

tics. The development and the rollout of neoliberal statistical management in the 90's, and the generalization of [Monitoring and Evaluation \(M&E\)](#) for different projects ([Desrosières 2014](#)) only strengthened and deepened this tendency. The low investment in data collection and excruciating exigency for unrealistic precision, the importation of external methods and definitions, the overarching generalization of categories for administrative simplicity, all these features can be traced back to a long history of producing evidence that are not designed at local level decision making and management.

1.2 Work hypothesis

My main work hypothesis an important reason why health information systems in developing countries are not performing well is that they are and designed and implemented using top-down approaches, not well suited to their purpose.

By top-down I mean systems in which the purported usage of information calls for specific data collection and analytical tools. Most HMIS design recommendations for an ex-ante design of both data analysis and measure framework, from which data collection should be inferred. ([Lippeveld 2000](#), [RHINO 2003](#), [d'Altilia et al. 2005](#), [Health Metrics Network 2008](#)). For example, [M&E](#) guidelines such as those of Global Health agencies like the Global Fund exercise important drives on the organization of health information systems in sub-Saharan African countries on which they are imposing definitions, classifications and reporting methods, to provide sufficient information for the evaluation of funded programs ([The Global Fund 2014](#)). This asymmetric relationship between international and local actors leads to a *reductio ad M&E* of information systems, that are built to answer imposed norms but do not create their own logics, cultures and traditions.

The mindset behind these approaches is influenced both by the colonial administrative tradition, for which standardized statistical tools should provide a normed information, and by the developmentalist expertise tradition, along which an hyper directed data collection system is aimed at providing a unique and narrowly defined piece of knowledge. This mindset is still present in the current tendency of international donors to encourage the development of systems specifically aimed at answering their own particular information needs. National Health Information Systems are then organized around these multiple narrowly defined needs for information, and are directed towards answering questions asked by national or international specialists, with little attention at local level uses.

The critic of this type of statistical systems is often conducted from a postmodernist perspective, questioning the structures and processes that produce statistical knowledge in developing countries. This critic, if it is sometimes useful to the practicing statistician, often poses a radical critic of the quantitative approach to public affairs, offering little place for incremental improvements ([Rottenburg et al. 2016](#)). Building on Latour's call for a critic oriented *towards* its object and "offer an arena in which to gather" ([Latour 2004](#)), my work is aimed at defining what a different approach to building information systems could be. In this dissertation, I explore ways through which a top-down approach to evidence generation in health systems can be reversed, to promote bottom-up data analysis methods that put more emphasis on data already being collected in the developing countries, and from which relevant information is extracted in an ad hoc way.

This bottom-up approach also builds on the latest developments of modern data science that are seldom found in routine health information systems in developing countries.

Data Centrality I take the available data as an object of study *per se*, from which the analysis starts, and which conditions all subsequent developments. This means both adapting analysis aims to available data, and integrating the data collection processes in the models of interest. A corollary of this approach is a more complete expression of the uncertainty of the measure of these quantities, which are increasingly part of the Global Health Metrics practices ([Murray 2007](#)).

Test and validation Strong national statistical systems emerged from both norm and local improvisation, through trial and error ([Lecuyer 1987](#), [Chaperon et al. 1988](#)). Incrementally adapting

current metrics and methods through localized innovation is essential, but the conditions and methods to do this are hard to create. As I develop, test and validate methods I will develop ideas and examples on how this could be made.

Bayesian Perspective Finally, most of the methods used in my dissertation tend to favor a Bayesian perspective on data modeling. Using a Bayesian approach in administrative systems has been noted to be a promising approach (Fienberg 2011, Little 2012). Bryant and Graham also noted how Bayesian methods allow a good combination of a diversity data sources and proper handling of uncertainty (Bryant & Graham 2013). In my different projects, I will work on presenting and displaying the benefits of this Bayesian approach to use and analyse health information systems data.

1.3 Aims

My principal aim will be to define, implement and test *bottom up* approaches to improve Health Information Systems data use in developing countries. This aim will be reached through three specific aims, which address different key processes in health information systems. The Health Metrics Network defined the three main processes of Health Information Systems as the indicators definition and standardization, data sources and data management (database hosting, data cleaning and data processing) (Health Metrics Network 2008).

I will address each of these processes with a specific aim, and will explore how methods based on available data and processes can improve the performance of Health Information Systems.

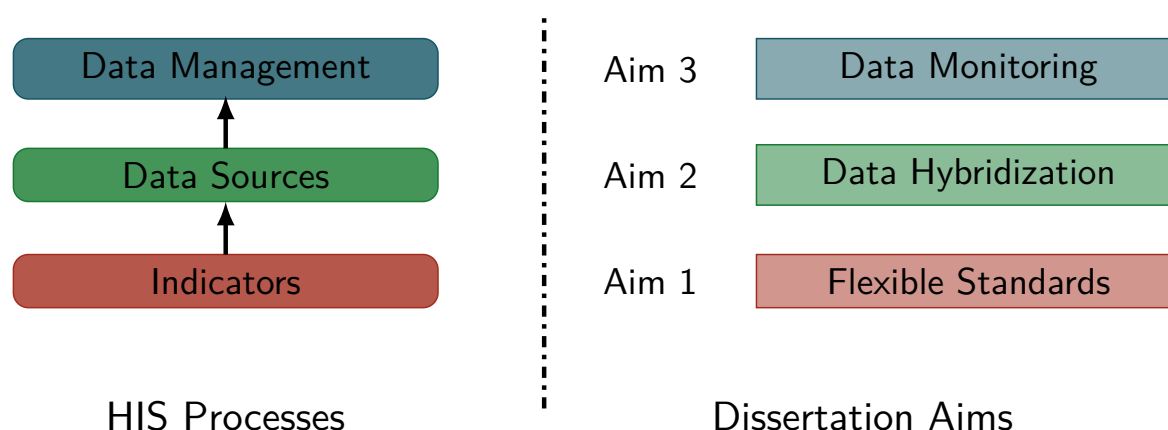


Figure 1: HIS Processes and corresponding dissertation aims

Indicators / Flexible Standards Defining categories and metrics based on which people are going to be counted is an essential piece of the statistical work (Desrosières 1993). It is an essential step in the simplification involved in the activity of measurement. The field of Global Health relies on important taxonomies, like the International Classification of Diseases and on Metrics like the Disability Adjusted Life Years, to unify description and measurement of health across the globe, and allow comparison and benchmarking (Murray 2007, Murray & Frenk 2008). Meanwhile, at local level, the use of globally defined metrics may have its limits, as it does not allow adaptation to local contexts and situations. The measure of retention of HIV patients in care is an example of this. Understanding the outcome of patients after they enter care is essential to evaluating HIV care systems performance, and efforts have been made to track measure to inform strategic planning and programs evaluation (The Global Fund 2014). Meanwhile, the measure used in most settings, namely the proportion of patients that are considered **Lost To Follow Up (LTFU)** after a certain time in care,

is an indicator that was originally designed and used in clinical research settings. The high variation of [LTFU](#) rates between programs and the low specificity of this metric leads researcher to question the way Loss to Follow Up is defined and retention is measured globally ([Chi et al. 2011](#), [Yehia et al. 2012](#), [Grimsrud et al. 2013](#), [Forster et al. 2008](#)). Using [Electronic Medical Records \(EMR\)](#), I will model how the measure of retention is affected by local contexts and data quality, and I will explore more robust ways to measure retention in HIV care.

Data Sources / Data Hybridization Unavailability of population data in a large number of countries is a well known issue ([Mahapatra et al. 2007](#), [Mikkelsen et al. 2015](#)). As a result, spatial distribution of populations, an essential piece of evidence to develop public health policies, is often unavailable at a policy relevant scale. To palliate these shortcomings, approaches for mapping of populations have been developed, involving the use of macro-level rasters of covariates such as land coverage or night lighting imagery ([Linard et al. 2012](#), [Stevens et al. 2015](#)). These approaches give interesting insights into how populations are distributed, and their output can easily be used for other public health work using the GPS coordinates system. Meanwhile, in situations with very low information on populations, the results of these top-down approaches is often little more than an overlay of covariates. Additionally, their results are hard to use for public health policy planning, as they do not link population to commonly used localization conventions such as places names. My second aim is to hybridize multiple data sources on population to produce a population map of Niger linked to the lowest level of population settlement possible.

Data management / Data Monitoring Data quality is an important concern of Health Information Systems professionals ([Shrestha & Bodart 2000](#)) who link it directly to the ability of Information systems to provide good information ([Mphatswe et al. 2012](#)). The definition of data quality is directly linked to an idea of trustworthiness, which may affect the use of certain data sources to generate evidence ([Gimbel et al. 2011](#)), and the quality of health data receives a large amount of attention even from high level actors in Global Health ([Abou-Zahr et al. 2010](#)). The best way to ensure data quality is through routine audits of primary data ([Ronveaux et al. 2005](#)). Meanwhile, this approach is costly and may have some cost-efficiency issues. Additionally, it makes little use of historical data, which makes it vulnerable to issues in primary data. My third aim will be to develop a cost effective approach to data quality screening informed by historical data and including a notion of risk attached to data quality.

1.4 Novelty and scientific contribution

My dissertation is contributing to three main areas of research surrounding Health Information Systems. One is mostly investigated by the [Information and Communication Technologies \(ICT\)](#) and interrogates the importance of local adaptation of information systems and its impact on the definition of standards. The second domain of relevance is more linked to the Global Health field, and on works that interrogate how data collected inside health systems can be used to best inform decision making. The last domain of contribution is mainly of interest for the statistical field, and explores the use of statistical methods for surveillance in applied fields.

1.4.1 Local Adaptation and flexible standards

In the ICT field, defining standards for data collection systems that can be implemented at local levels but respond to national or international norms. Jørn Braa, describing the approach that presided to the design and development of the [District Health Information Systems \(DHIS2\)](#) remarks that "the top-down and all-inclusive approach to standardization [is] common among ministries and central agencies" and pleads for *flexible standards* following the idea that "the individual standards must be crafted in a manner which allows the whole complex system of standards to be adaptive to the local context" ([Braa et al. 2007](#)). The need for local adaptability of Information Systems is seen

as a key issue of ICT in developing countries (Macfarlane 2005, Walsham & Sahay 2006, Walsham et al. 2007, Jacucci et al. 2006) and thus some ICT solutions have been designed and explored in the forms of tools like DHIS2 or standards like the Open Health Information Exchange initiative.

Aims 1 and 3 explore ways in which standards for indicators definition or for the evaluation of data quality can be adapted to specific contexts. In both situations, flexibility is introduced through objective rules : specificity and robustness of the measure of retention in the Aim 1, and cost-effectiveness in the Aim 3. The two rules are different in nature, as one is internal, and based on the characteristics of the data while the other is external, and based on operational characteristics such as the cost of supervision. This approach is innovative, and can be applied to a diversity of questions. An extension of this work could also be the definition of methods for aggregation and comparison of metrics defined and measured using flexible standards.

1.4.2 Imperfect data Usage

Health Information Systems data is often underused, or not used at all by its intended users (Health Metrics Network 2008). This underuse is often blamed on the perceived bad quality of primary data that would make it unfit for statistical analysis (Ronveaux et al. 2005, Makombe et al. 2008, Heunis et al. 2011, Gimbel et al. 2011, WHO 2011, Hahn et al. 2013, Kihuba et al. 2014, Glèlè Ahanhanzo et al. 2015). Approaches to solve this issue have mainly focused on improving primary data quality to improve data use (Braa et al. 2012, Mutale et al. 2013, Ledikwe et al. 2014, Nisingizwe et al. 2014), but some authors have also pointed to the possibility to use even imperfect routine data to answer specific public health questions (Gething et al. 2006, 2007, Wagenaar et al. 2016). These approaches rely on the nature of routine data, which is usually highly dimensional and structured times series. This type of data enables the use of robust modeling methods at an adequate level of aggregation, to control and correct for data errors.

My dissertation contributes to this line of research in two different ways.

Aim 1 and 3 also offer improved insights in routine data quality and its impact on health indicators. Aim 1 explores a micro level modeling of data quality and its impact on the measure of patient retention. It will also define retention indicators less sensitive to data quality. Aim 3 uses expected value imputation and aberrant values detection, which is an essential aspect of understanding data quality. It will also implement solutions to analyze and present the type of data often observed in health systems.

Aim 1 and aim 2 also examine methods to enrich data, using EMR metadata or hybridizing different data sources. Both approaches are an extension of usual approaches to health data, and make full use of the richness of modern data collection tools, and of the wide availability of public data sources.

1.4.3 Statistical methods for health data monitoring

For aim 3, I am interested in three types of literature that address issues of health data monitoring: epidemiological surveillance, health service quality monitoring and industrial Statistical Process Control (SPC).

I will explore a combination of those approaches to provide actionable information on available data in a health system. Indeed, all of these approach rely on a common method, based on the theoretical definition of an expected distribution of some indicator, and then defining a test to classify the observed values as coming from an *in control* or *out of control* process. Using a rare and important data set, I will test and benchmark different algorithm on their ability to detect different types of *out of control* situations.

In the meantime, my implementation of quality monitoring algorithms will differ from existing literature in an essential way. In most quality monitoring applications, the quantity of interest is the success of some type of procedure, measured by an outcome, of which death is usually the least desirable. In the context of Results Based Funding, on the other hand, the measure of success is

much more a measure of access to a wide range of services. The denominator of such a measure is much less clear, as it is not defined at facility level, and there is often little information on target populations. I will thus use methods recently developed in the [SPC](#) field, which aim at monitoring profiles of indicators sets, in order to detect diverging patterns from a standard situation([Woodall & Montgomery 2014](#)). The use of Profile Monitoring for this project is an important innovation in the field of surveillance.

1.5 Timeline

Aims specific timelines are explained in more detail in sections [2.4](#), [3.4](#) and [4.4](#). I will prioritize getting early results on Aim 2 to allow feedback and field testing from Bluesquare and have an opportunity to reorient my methods if needed. This will also leave time me to obtain potential additional data for Aims 1 and 3. All the aims should be finalized and papers written by the end of the first quarter of 2018, which would allow for a finalization of the dissertation by June 2018.

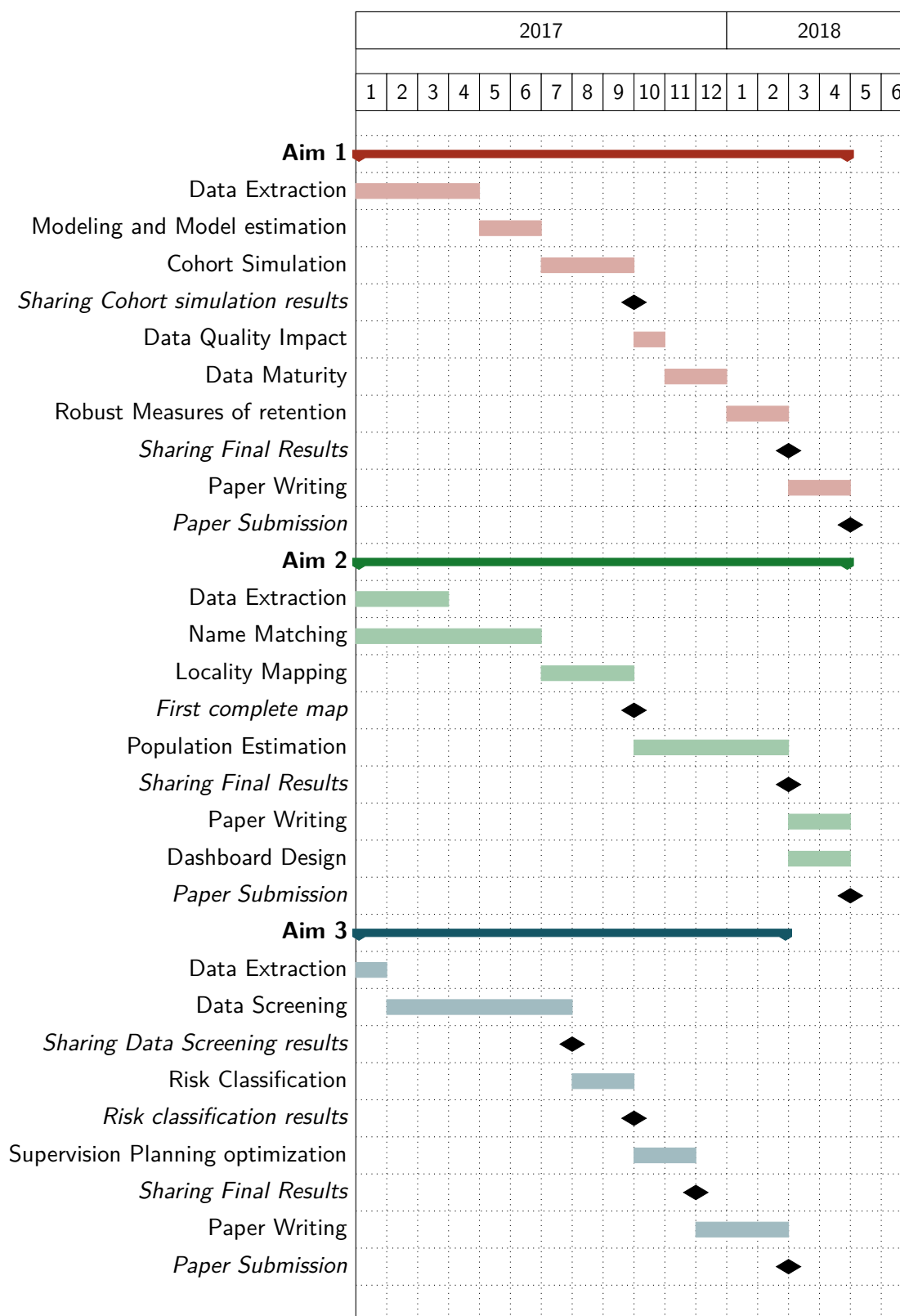


Figure 2: Dissertation Timeline

2 Aim 1 - Flexible Standards for categories definition

The measure of retention of patients under [Antiretroviral Treatment \(ART\)](#) is a key measure of HIV programs performance. HIV being a chronic disease, the measure of success of these treatments is indeed the survival of treated patients, which is directly dependent to their adherence to [ART](#) treatment. As HIV programs have been rolled out in different countries, the importance of [LTFU](#) in [ART](#) cohorts has been recognized, and since the early days of HIV care programs, efforts have been made to better measure and understand the how and why patients remain in care on the long term ([Ioannidis et al. 1999](#), [Lebouché et al. 2006](#), [Moh et al. 2007](#)).

Initiating treatment programs, HIV practitioners used the categories that had been developed and used in clinical trials to categorize patients outcome. This definition is mainly a negative category, grouping patients who failed to attend their medical appointments and for which no definitive had been recorded. [LTFU](#) is thus defined by an absence of data about a patient more than by a positive element of information.

As HIV programs were rolled out, it soon became clear that the [LTFU](#) category, in many settings, is a mixed bag of patients with different status ([Kwong-Leung Yu et al. 2007](#), [Dalal et al. 2008](#), [McGuire et al. 2010](#)). This poses problems for HIV programs evaluation, as the true outcome of HIV patient is often unknown, and different authors have proposed ways to correct retention measures by collecting additional data ([Yiannoutsos et al. 2008](#), [Geng et al. 2010](#), [Tassie et al. 2010](#)), or by fitting correcting models ([Brinkhof et al. 2010](#), [Egger et al. 2011](#), [Van Cutsem et al. 2011](#), [Henriques et al. 2012](#), [Verguet et al. 2013](#)). Finally, some authors started paying attention to how the very metric used to measure patients retention could heavily impact the evaluation of HIV programs performance ([Chi et al. 2010, 2011](#), [Fox et al. 2012](#), [Mugavero et al. 2012](#), [Yehia et al. 2012](#), [Grimsrud et al. 2013](#), [Shepherd et al. 2013](#)).

An important underlying hypothesis of the use of the [LTFU](#) metric as a proxy for retention is that, if the patient had been coming to the facility for an appointment, this visit would have been recorded. This hypothesis makes sense considering the notion of [LTFU](#) comes from the clinical trial setting ([Lebouché et al. 2006](#)) but it can be problematic in developing countries, where data quality in care setting is often difficult to guarantee. In these settings, the delay between a visit and the time of data entry will be important, mechanically inflating the number of [LTFU](#) ([Lurton 2012](#)). There is thus an equivalence established between patient's status and data quality in the facility. Meanwhile, assessing data quality of a database and its impact on the measure of retention is seldom made.

With the development and rollout of [EMR](#) in facilities, there is an opportunity to improve on this situation. Most computerized data collection systems collect metadata on data entry, such as the date at which a form was entered, or the role of the person doing the data entry. This metadata could be mobilized to measure the quality of the data, and to understand the impact of data quality on the measure of retention.

This project main aim is to simulate and describe the impact of data quality on the measure of retention of patients in HIV cohorts. This will be achieved through three specific aims :

Impact of data quality on the measure of retention I will model the path through which data quality impacts the measure of retention, and will simulate the results metrics of an HIV cohort with varying data quality.

Data Maturity As the time to data entry is an essential element of data quality and directly impacts the measure of retention, the ability to properly measure retention for a given period improves over time. I will define a metric of data maturity to describe the ability of a database to measure retention for a cohort. This measure will be tested and validated on the simulated cohort.

Robust Retention Metrics Building on the insights gained in the previous aims, I will define, test and validate different robust retention metrics for HIV cohorts. These metrics will be evaluated

on their ability to properly measure the retention of patients in a simulated cohort, and on their robustness to varying data quality.

2.1 Theoretical Framework

2.1.1 HIV care process

HIV patients are visiting hospitals for regular appointments. At each visit, they are given an appointment for their next visit. The date for the v^{th} visit of patient i is noted V_v^i , and the appointment time for this visit is noted A_v^i . Thus, if $V_v^i > A_v^i$, the patient came late to his appointment, if $V_v^i < A_v^i$, the patient came early. We note $l_v^i = A_v^i - V_v^i$, the lag between scheduled appointment and actual visit. Figure 3 shows the distribution of l_v^i in our data. Unrecorded appointments are handled by dividing the time between scheduled appointment and actual visit by 28, and considering the remainder. The spikes around 7, 14 and 21 days show either patients who were given an unrecorded appointment of less than a month, or is proof to the fact that patients have favorite days of visit.

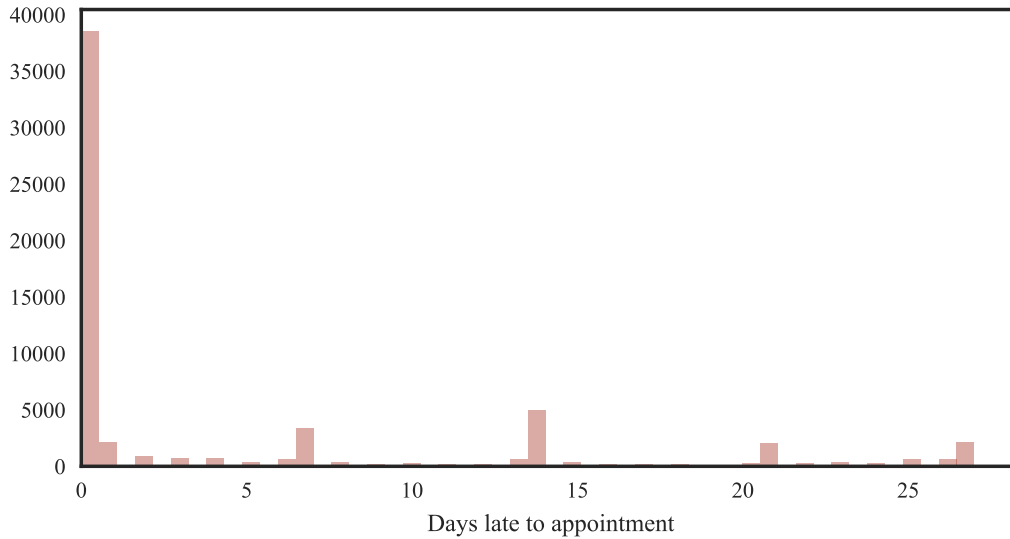


Figure 3: Distribution of l_v^i in our data.

The time elapsed between a visit and the next scheduled appointment is set by the national norms regarding the frequency at which HIV patients should be evaluated, depending on their condition and medical history. Patients recently enrolled in care will be seen more frequently than patients with a longer follow-up and no complications. We note (f_v^i) the visit frequency regimen of patient i at visit v . This unit is usually around a multiple of 28 days, as patients are likely to have a favorite week day for visit (cf. *supra*). Figure 4 shows the distribution of f_v^i in our data.

We can finally express the time between two visits as :

$$V_{v+1}^i - V_v^i = f_v^i + l_v^i$$

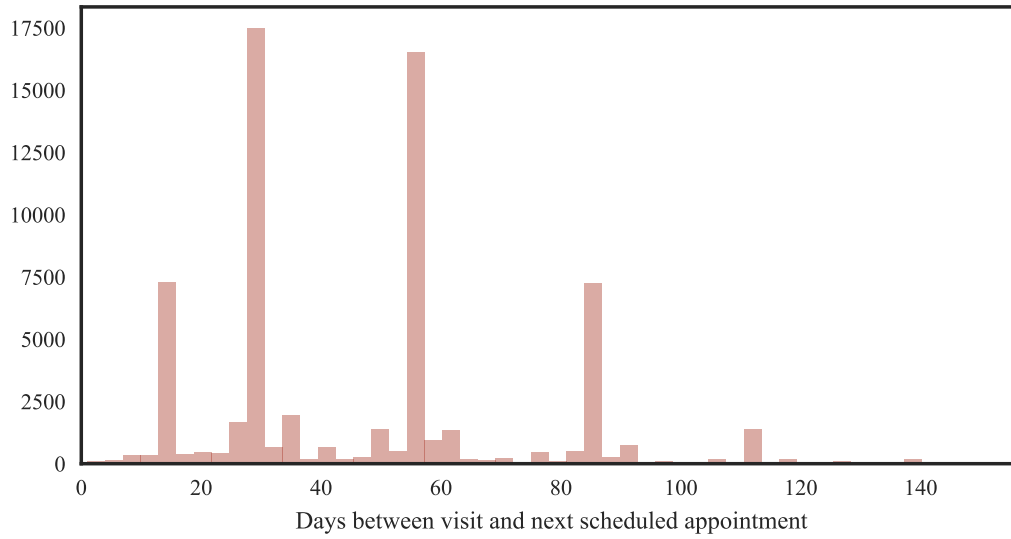


Figure 4: Distribution of f_v^i in our data. We see the visits are scheduled mainly 7, 28, 56 or 84 days after the previous visit

2.1.2 Data Entry

The date at which a visit is recorded in an EMR database is $R(V_v^i)$. By definition, $R(V_v^i) \geq V_v^i$, and the delay in data entry is noted :

$$R(V_v^i) - V_v^i = \delta_v^i \geq 0$$

δ may vary in a facility, depending on the workload, staffing or other factors. In some cases, the visit has not and will never be recorded. I will note this situation as $\delta \rightarrow \infty$. Figure 5 shows the distribution of δ in our data.

Finally, data entry is interrupted at date T_{close} before the data is used for analysis. The time elapsed between patient i 's last visit and the closing date is noted as $G_i = T_{close} - \max_v(A_v^i)$. For simplicity, we will equate the date of database closure with the date of analysis in a first step, and will relax this assumption when we will be measuring data maturity.

Figure 6 shows how these different parameters can play out for a given patient. This imaginary patient had a first visit on January 1st, and had an appointment scheduled on February 1st, to which he came 6 days late. After three months of being seen monthly, he switched to a quarterly follow up. He was early to his July appointment, but came to every appointment until the end of the year. The data was entered very quick at the beginning of the year, but v_2^i was never recorded. v_4^i was entered before v_3^i , and v_6^i could not be entered before the database was frozen for analysis on January 15 of the following year. This example gives a rough demonstration of the different situations and problems that can be encountered when analyzing the follow-up of the patient.

2.1.3 Loss to Follow Up definition

A central piece of the LTFU definition is the *grace period* during which a patient, even if he did not return to a facility, is considered actively followed. This *grace period* is denoted G_0 .

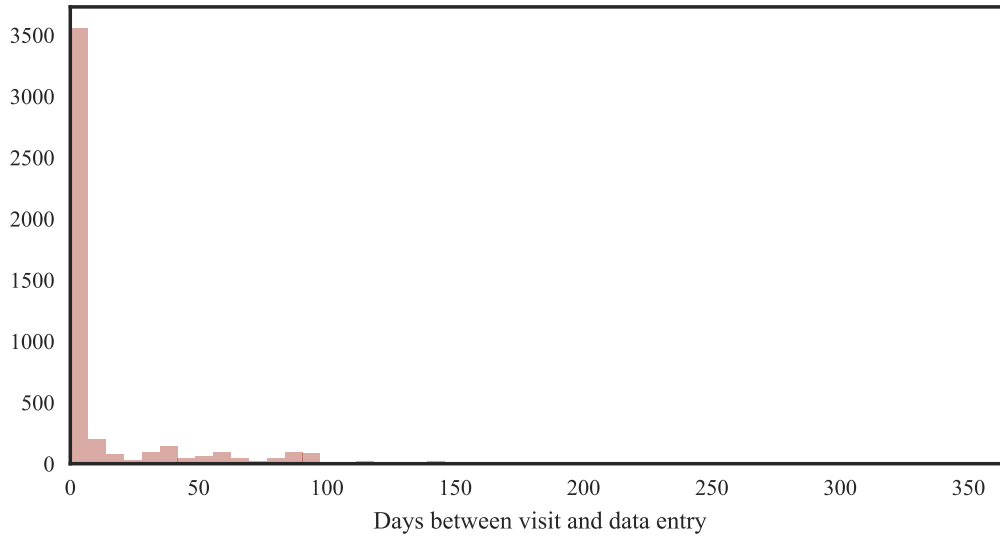


Figure 5: Distribution of δ_v^i in our data. Most of the data is entered in the first week following the visit, but we see some data can be entered much later

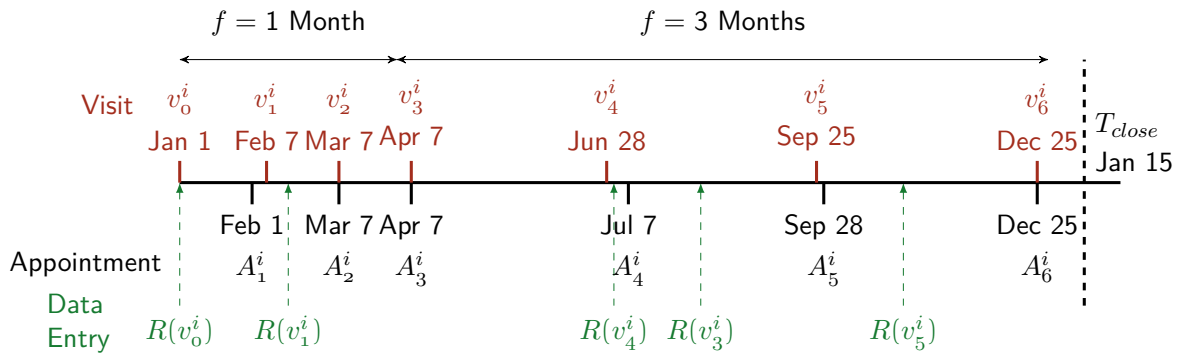


Figure 6: Individual follow-up and data entry process

A patient i is considered **LTFU** if he is late to his latest scheduled appointment for more than G_0 days.

$$l_{v^*}^i > G_0$$

Looking closer at this definition, we can see it regroups three different situations :

1. $l_{v^*}^i \rightarrow \infty$: The patient is **LTFU** and will never come back to the facility.
2. $\infty > l_{v^*,i}^i > G_0$: The patient is late to his appointment but will come back in the future.
3. $\delta_{v^*+1}^i > G_0$: The patient came for his visit $v^* + 1$ but data entry took longer than the grace period and the visit was not entered at the time of database closure.

Using this definition, we can express the probability that a patient is identified as **LTFU** based on the data at hand. Let's $X = 1$ be the event that a patient is actively in care, and $X = 0$ the event that the patient is **LTFU**. We can get $p(X = 0 | l_{v^*}^i \leq G_0)$ as the combination of elements we can measure :

$$p(X = 0 | l_{v^*}^i > G_0) = 1 - p(\infty > l_{v^*}^i > G_0) - p(\delta_{v^*+1}^i > G_0)$$

We can understand $\infty > l_{v*}^i \leq G_0$ as an intrinsic myopia of the health system, who can not predict the future, and $\delta_{v*+1} > G_0$ as a data quality measure. Differentiating between these two terms is important in order to understand uncertainty in the LTFU rate and better measure retention in the cohort.

2.2 Data

The data used for this aim are EMR database obtained from the HIV programs. I currently work with an EMR from Kenya, and part of IHME's ABCE study. In this facility, 4833 patients have been registered for HIV care, from 2005 to June 2012, totaling 69591 recorded visits. Data entry time is easily available for at least 4853 of these visits.

I hope to obtain two additional EMR databases. The ISanté EMR, used in Haïti, and multiple implementations of the FUCHIA system, developed by Médecin Sans Frontières and used by the national HIV program in Niger. This would allow me to better measure variability in appointments lags and data entry issues. All the data will be analyzed anonymously and in aggregate form. For each patient, I only use visit dates and scheduled appointment dates. If scheduled appointment dates are missing, they will be imputed using observed f_v^i in the data. I will also use the metadata collected in the EMR, especially the dates of data saving, to estimate data quality by measuring δ distribution.

I will not report precise data on HIV programs performance, and all the data used will be used to inform the simulation model.

2.3 Methods

This work will be done in three main steps. First, I will estimate the relevant quantities from the data at hand. In a second step, I will simulate a cohort and its monitoring, using estimated quantities as parameters. Finally, I will use this simulation, varying different parameters, to answer my main questions of interest.

Modeling - The different parameters described earlier will be modeled and estimated from the cohort data I will have at hand, using a Bayesian approach. The two most important parameters for this work are δ , the time before visit has been recorded in a database, and l , the time between an appointment and the actual visit.

A first approach to modeling these parameters is to express δ as a Gamma distribution $G(\alpha, \beta)$, with mean $\frac{\alpha}{\beta}$ representing the mean time to data entry of a visit form in the EMR. A Gamma distribution allows for a very long tail on the right, which will allow us to include data loss ($\delta \rightarrow \infty$).

l could be modeled using a mixture model, to take into account the multimodal nature of the l . I will also have to consider the hypothesis that the multimodal l distribution results from unrecorded shorter term appointments, and if this is the case we may need to use a long tailed distribution as for δ .

Simulation - Using the parameters estimated in the previous step, I will simulate an HIV patients cohort. This simulation will be made using the Cost Effectiveness Analysis Microsimulation (CEAM) framework developed at IHME to simulate epidemiological cohorts, in which I will feed draws from the posterior distributions of the previously estimated models.

In a second step, I will simulate the data entry process, for any given month, by drawing a δ for each visit that will have been generated (initial or returning), and then computing the date of data entry for the information related to this visit.

As a result of this simulation, I will have all the information needed to estimate retention and measure of retention for the cohort. Finally, varying the relevant parameters, I will be able to measure the quantities of interest for my study aims.

Quantities of interest - This simulated data will then be used to estimate our elements of interest :

1. **Measuring data quality impact** : From the cohort simulations, I will measure the LTFU rate using different distributions of δ . Different scenarios will be considered for data quality, varying both the mean and variance of δ . Perfect data quality will be compared to situations with long delays of data entry, and situations with important data loss (high variance of δ). The resulting observed variation in the LTFU rate will be described as the impact of data quality on the measure of retention.
2. **Data maturity** As data is being entered in the EMR, or as missed visits are finally being made, the data for a given period will get completed, and patients actively on care are more and more considered so. As data maturity grows in the EMR, the data quality induced error is lowered. Varying T_{close} can thus have an impact on the measure of retention of a patient on a given date. I will carry out the measure of retention using different closing dates for the database, and only using the data recorded before the closing date. These measures will allow me to define and test a Data Maturity metric, based on a combination of f , l and δ that will allow us to identify the optimal minimum date of analysis to estimate retention rates in a program, and the optimal grace period G_0 to use for different levels of maturity.
3. **Robust measures of retention** Finally, we will consider more robust metrics that can be considered good proxies for retention. These metrics will include :
 - The ratio of the corrected average number of registered visits on the expected number of followed patients
 - The ratio of new to returning patients in the facility
 - The probability that the rate of LTFU is higher than a given threshold

For each of these metrics, I will evaluate their capacity to measure retention in the cohort, by comparing with the reference measure of LTFU measured with perfect data. I will also evaluate the sensibility of these metrics to data quality and data maturity.

2.4 Timeline

I am still hoping to obtain additional data for this aim. I anticipate starting to work on the estimation of the model parameters from this data in April 2017. Based on these parameters, I will be able to implement the cohort and data entry simulation and will have first results to share on this simulation in September 2017. Based on this simulation, I will estimate data quality impact, data maturity impact and I will test robust measures of retention during the last quarter of 2017 and the first quarter of 2018, and I plan on having final results by February 2018. Figure 7 summaries this timeline.

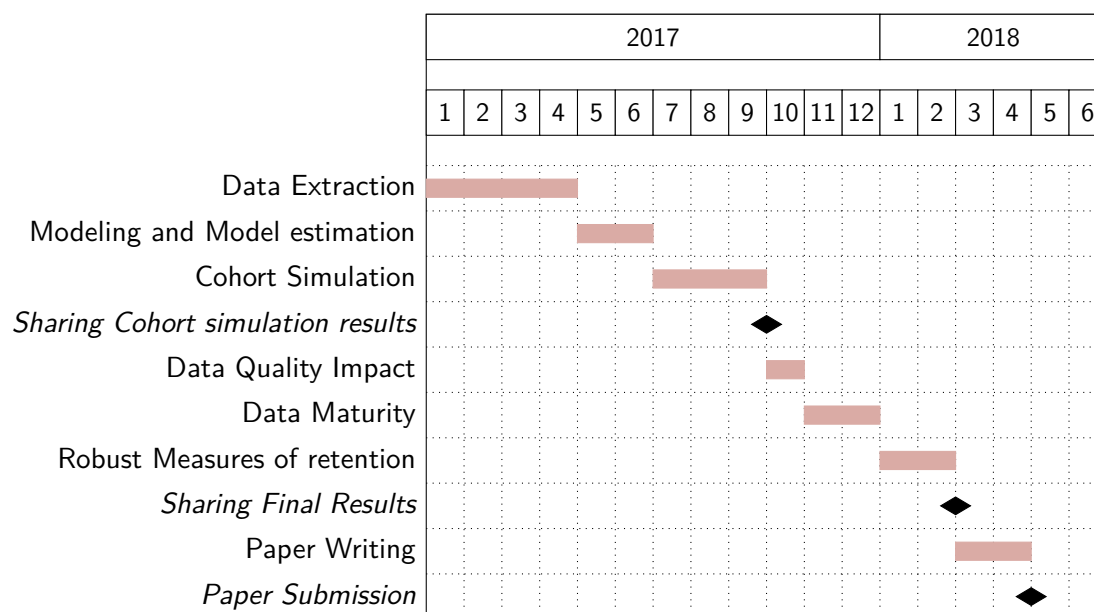


Figure 7: Gantt Chart for Aim 1



Figure 8: Mapping of Niger population from AFRIPOP

3 Aim 2 - Data hybridization for population mapping in Niger

In 1935, British geographer Charles Fawcett defined three phenomenons that could be described on a population map ([Fawcett 1935](#)):

1. The actual number of the people within given areas
2. The density of the population in these areas
3. The grouping, or arrangement, of the population in space.

Each of these elements is important to design public health programs, to plan infrastructure development or to implement quick emergency response ([Bambas Nolen et al. 2005](#), [Thieren 2005](#)). Meanwhile, the lack of good quality data on populations size and locations in low resource and developing countries is well known ([Mikkelsen et al. 2015](#)). As a consequence, most current approaches to population mapping rely on models aimed at mapping density surfaces ([Linard et al. 2012](#)). This approach makes best use of the availability of large datasets for land use and other usable covariates, and of the computational ability to interpolate these different data sources for population distribution ([Stevens et al. 2015](#)). As is evident for Niger in Figure 8, in countries with little urbanization, and poor population data, these methods end up displaying an overlay of covariate layers more than they present a credible distribution of populations.

Mapping the density of the population is useful for descriptive purposes, and the production of density rasters can be essential to the spatial modeling of diseases and other population phenomenons. Meanwhile, these maps do not display actual numbers of people, nor a comprehensive display of populations groupings. Moreover, being able to query places by their name is an essential feature for managers. The information needed to produce a precise and actionable map of population in Niger can be found scattered in different unrelated data sources. This project is exploring an innovative approach to provide a population map in Niger, through the hybridization of multiple data sources and the use of Bayesian models for administrative data.

3.1 Data

aim2:data

Voters list as a demographic data source A data source that is, to my knowledge, seldom used to inform population mapping for public health purposes, is voters registration lists. This low usage can be linked to concerns regarding the completeness of this data in settings where only subsets of the population are eligible to vote, to issues regarding the regular update of this lists, or to the political purported use of these data, that could favor non-random missingness in data in some situations.

There is meanwhile a case to be made for the use of voters' registration data to estimate size and the spatial distribution of populations. By definition, voters' registration should aim at being as complete as possible a register of adults in the nation. Moreover, in most democracies, some form of national elections are held at least every five years, leading to an update at least partial of voters' registrations. In sub-Saharan Africa, between the years 2015 and 2016, 27 countries were supposed to hold national elections, leading to a theoretical registration of more than half of the adult population of the continent. Finally, for transparency and accountability reasons, electors registries are supposed to be accessible.

Due to the sensitive and political use of these data, the quality of voters registries are often described as not being trustworthy. On the other hand, for the same sensitivity reasons, voters registries are receiving a high level of scrutiny from different actors, and are audited sometimes multiple times before validation. This level of scrutiny before validation is much higher than the attention given to a lot of studies or other often used data sources.

The Niger 2016 elections voters registry In Niger, presidential and parliamentary elections were held in February 2016. Voters lists were updated during the second half of the year 2015, under the supervision and control of a mission of the [Office International de la Francophonie \(OIF\)](#). The operations for registration of voters were conducted during the third quarter of 2015¹. A first version of the voters list was published on December 21, 2015, tallying 7,569,172 voters, out of 8,569,309 that were expected based on the 2012 census².

Final lists were validated in early January 2016 after being corrected for some incoherencies noted by the supervisory body³. A final report on these lists was published in may 2016⁴. The [Comission Electorale Nationale Independante \(CENI\)](#) later made these lists fully available on its website, from which I extracted, anonymized and formatted the lists.

RENALOC and RENACOM The [Répertoire National des Localités \(RENALOC\)](#) is a geolocalized repertory of all localities in Niger. The 2012 version was downloaded as a pdf file from the [Institut National de la Statistique \(INS\)](#) website. The tables were extracted in bulk from this file using the Tabula Package, and then processed in Python to recompose the geographic structure of the document. The final data consists in 34507 localities, for which the INS provides the number of inhabitants, by gender, as well as the number of households, and the number of agricultural households. For most of the localities, a GPS coordinate is recorded, as well as the type of locality (neighborhood, village, camp, water well, hamlet).

The 2001 version of this database, named [Répertoire National des Communes \(RENACOM\)](#), contains similar information. Meanwhile, the number of places identified varies, and for places identified in [RENALOC](#) and [RENACOM](#), some names spelling vary. I retrieved the [RENACOM](#) in Excel tabular format directly from the [INS](#) website.

¹<http://www.ceni-niger.org/article-region/#more-24>

²http://www.iinanews.org/page/public/news_details.aspx?id=98929&NL=True

³<http://www.nigerinter.com/2016/01/le-fichier-electoral-du-niger-valable-sous-reserves/>

⁴<http://www.nigerinter.com/2016/05/remise-officielle-du-rapport-du-fichier-electoral-au-ministre-detat-a-linterieur-par-le-cfeb/>

OpenStreetMap OpenStreetMap (OSM) is "a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license"⁵. Its API allows an easy query of its content, from which we can retrieve places names and community generated GPS coordinates. This data can provide additional precisions on where some localities are, but is much less complete than both RENALOC and RENACOM.

DHI2 The Niger *Système National d'Information Sanitaire (SNIS)* is currently implementing DHIS2. The portal to DHIS2 already makes available some limited geolocalization data, regarding the health districts divisions, and some health facilities coordinates⁶.

3.2 Methods

To make the best use of this data, I will implement an approach based on a minimal modelling of primary population data, and geared towards the anchoring of population in callable named localization. To achieve this, my project has three main components.

3.2.1 Name Matching

Due to the history of the creation and administration of the Nigerien territory, multiple different spellings are in use for most localities names in Niger. There are no obvious reasons to prioritize one spelling over another for this project. To the contrary, I want users to be able to use whichever spelling of a name they prefer to query their results.

In collaboration with Fahad Pervaiz, a PhD student in the department of Computer Science at UW, I am designing a matching algorithm for different spellings of the same locality names in Niger. Our approach relies on the use of a mixture of standard string matching algorithms. We use these algorithms for each pair of data sources and define a heuristic to combine them and select best matches. We also enrich these heuristics by defining patterns and features that allow a first classification and simplification of names to improve matching performance. These patterns may be data source specific to reflect explicit or non-explicit conventions used in each data source.

After this first round of unsupervised matching, we will manually confirm some of the matches with the help of members of the OSM community in Niger. Using this validated training set, we will fit supervised algorithms to improve our previous matching approach.

As a result of this step, I will have a consolidated list of localities in Niger, with different possible spellings of names for each of them.

3.2.2 Locality mapping

The three data sources that include GPS coordinates (RENALOC, RENACOM, OSM) have GPS coordinates for different subsets of localities in Niger. It appears that RENALOC GPS coordinates are of very low quality, and that OSM coordinates are sometimes rough estimates of exact locations with a rounding factor. I will design an algorithm to attach, for each identified locality, the most probable GPS localization.

1. Get RENACOM GPS coordinates for localities where they are available.
2. Fit some models to correct GPS coordinates in glsrenaloc using localities with both RENALOC and RENACOM GPS coordinates. Use the best performing approach, as evaluated with cross-validation, to correct RENALOC coordinates

⁵http://wiki.openstreetmap.org/wiki/About_OpenStreetMap

⁶<http://www.snisniger.ne/dhis-web-commons/security/login.action>

3. Fit some models to correct GPS coordinates in **OSM** using localities with both **OSM** and **RENACOM** GPS coordinates. Use the best performing approach, as evaluated with cross-validation, to correct **OSM** coordinates

For steps 2 and 3, different linear models will be tested, as well as non linear Machine Learning approaches allowing for different local corrections. For localities with **RENALOC** and **OSM** coordinates but no **RENACOM**, I will evaluate if the results 2 or 3 or a combination of both performs best, using localities with GPS from the three data sources as training set.

As a result of this step, I will have the most complete and accurate grid possible of named localities in Niger.

3.2.3 Population modeling

Finally, I will model Niger population using its voters list by voting precinct as a main data source. I could not source an example of using voters list as a source for demographic estimation. Meanwhile, in a country like Niger where elections are held much more regularly than censuses, using voting lists, a quasi complete enumeration of the population, to estimate population size and structure, does not seem unreasonable.

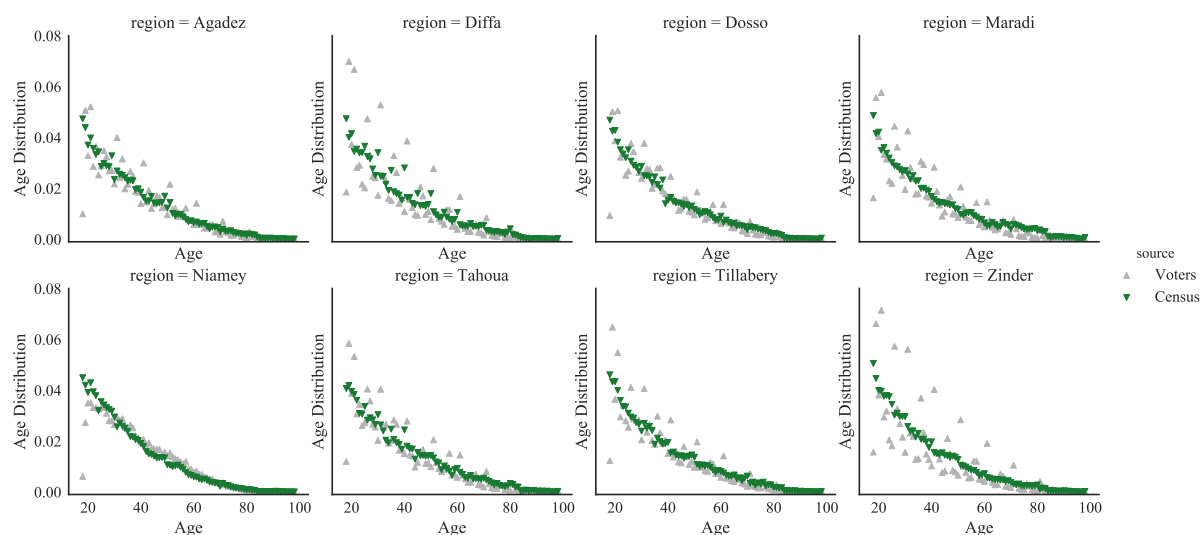


Figure 9: Comparison of standardized adult age distribution between the 2012 census and the 2016 voting lists

A specificity of the voting list is that it does not include children under 18, as they are not allowed to vote. Additionally, I listed in section ?? some issues that can be raised regarding the quality and completeness of voters lists, and I should assess their quality and correct their count to match the census results. Finally, as voting lists are very local, I will need to determine the most appropriate level of aggregation to get a meaningful estimation of the population age and gender distribution. Figure 9 compares the standardized age distribution of adults in the 2012 census and in the 2016 voters list at regional level. We can see there is more variability in the voters lists age structure than in the census. Concordance between the two age distributions seems to vary between regions.

I will model population size and age and gender distribution at regional and health zone level, using the electoral lists as input data, and the 2012 census regional distributions of population by age and gender as validation data. A simple approach will be used to keep this part of the project tractable, and easily reproducible locally. Age distributions from the voters lists will be aggregated at the regional or health zone level, and a number of children will be computed using widely available life tables. I will use multiple life tables to get an estimation of uncertainty on this estimation. The

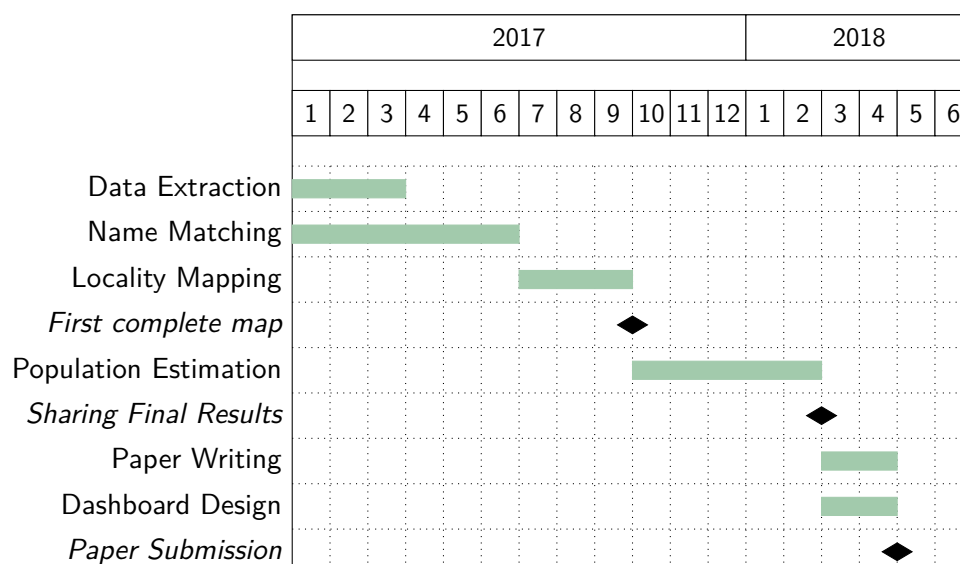


Figure 10: Gantt Chart for Aim 3

total population numbers will then be modeled in a linear regression framework, using the 2012 census numbers as validated results.

3.3 Output

The output of the project will be an interactive map, allowing the query of our results for local practitioners. This dashboard will have the following feature :

1. An interactive map of Niger localities, selectable by clicking, or panning for multiple selection
2. An estimation of the population in the zone selected on the map
3. A histogram representing the age structure of the population in the localities selected on the map
4. A search box through which the user will be able to search for a given locality with different spellings. Every name linked to a mapped locality will be searchable and will return the different matching localities in a hierarchized way.

3.4 Timeline

I am currently trying to obtain more detailed census data from Niger Census. Meanwhile, the name matching and locality mapping work is already well advanced as we have a first set of matched names from the unsupervised approach, and I have already explored approaches to GPS correction for [RENALOC](#). I anticipate three more months on the name matching and one month to confirm the locality mapping and the overlay with other *ad hoc* layers such as health services and health administration map, and should have completed mapping data by September 2017. I plan 5 months of work for the population estimation, and should have my final results by February 2018. Figure 10 summaries this timeline.

4 Aim 3 - Health System Data Monitoring and Screening

The third aim of my dissertation is the development of methods for screening and monitoring of data collected in health systems to inform and orient supervisions. Specifically, I will consider data collected in [Results Based Financing \(RBF\)](#) programs in Bénin. RBF is a mode of financing of health systems based on the ex-post payment of health facilities by a national financing body or program. This mode of financing alleviates the burden of planning for coordinating bodies, but transfers it to increased reporting needs for facilities. In order to receive these payments, facilities have to report monthly on a set of indicators on which these payments will be based ([Musgrove 2011](#)).

In a RBF system, understanding and monitoring results is thus important on two main levels. First, Program managers want to have an accurate evaluation of the activity they should be paying for, as the accurate evaluation of this amount is key to the success of the program. Underfinancing facilities endangers their ability to operate in good conditions, but overpaying some facilities may be detrimental to the overall sustainability of the project. Second, being able to measure and compare facilities performance is essential to identify weakly performing facilities, and to start implementing correcting measures.

In Bénin, the [Programme de Renforcement du Système de Santé \(PRSS\)](#) has been launched in 2011 by a consortium formed by the Beninese government, the World Bank, the GAVI Alliance and the Global Fund. As part of this program, a comprehensive RBF program has been implemented in all 34 health zones of the country. To allow the management of RBF reporting data, the software OpenRBF, developed by the Belgian startup Bluesquare has been implemented.

In systems using OpenRBF, data is collected in facilities for indicators contracted in the RBF program, and are reported at district level on a monthly basis. The District administrators are in charge of entering the data in the OpenRBF database. This monthly data is then aggregated, and checked on a quarterly basis for quality. Data quality check is made through a field visit made by project managers in facilities, who will then check the quality of primary data collection in facilities (reports, charts) and the quality of reporting, by comparing collected primary data and reported numbers.

This system allows to improve the confidence and exactitude of reported numbers, on which payments to facilities depend. Meanwhile, it is costly and does require regular field visits by program managers ([Antony et al. 2017](#)). This system has proved its worth in allowing RBF program managers to access credible data on facilities performance. Meanwhile, in a program managing close to 800 facilities, it is increasingly unceasingly difficult to validate data, and to monitor service quality. I will develop an approach to screen indicators reported by facilities, to help program managers making decisions. The decision framework in which we operate is quite simple, and has two main outcomes : validate data or not, raise a service quality issue or not. These two outcomes are not independent, and all depend on an evaluation of the normality of reported data when compared to previously observed data.

4.1 Research questions

This project will be aimed at developing and validating a generic framework for screening and validating data reported by facilities in the Bénin RBF program. I will develop this framework in three specific objectives.

Data screening In a first step, I will implement and compare different approaches to data monitoring in order to detect and report anomalies in the observed series. Combining methods used for syndromic surveillance, health system performance monitoring, and industrial SPC, I will define an algorithm to screen reported data, and spot abnormal values.

Classification In a second step, I will test a classification method to differentiate spotted abnormalities between data quality issues and other issues.

Program Management Finally, I will simulate different strategies to prospectively collect data in facilities, so as to allow an optimal performance of imputation models at a minimum cost.

4.2 Data

OpenRBF has been implemented in Bénin since March 2012, and has been rolled out in every départements of the country. Data is collected and validated monthly for the [Paquet Minimum d'Activités \(PMA\)](#), which is a minimum package of services that includes most standard primary care activities (immunization, antenatal care, malaria care). As of January 2017, the data consists in 16131 facility/months of [PMA](#) data in 671 different facilities. In each report, a median of 75% of indicators are correctly reported, as shown in figure 11.

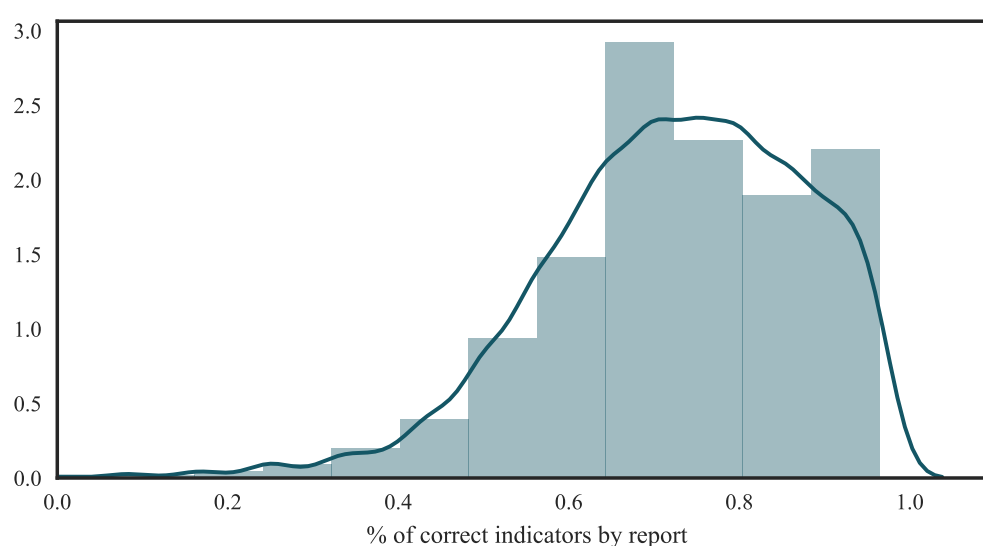


Figure 11: Distribution of the percentage of correct indicators by report

Figure 12 shows the distribution of the amounts by which payments have been corrected after data validation. We see that the distributions are skewed on the left, showing that most corrections lead to diminutions of the amounts. In the meantime, the distributions are very peaked around a mean close to 0, which leads me to hypothesize that most of the corrections can be attributed to situations of reported service provision that could not be fully documented, and a can thus be understood as data quality issues more than deliberate data manipulation.

Figure 13 shows the evolution of monthly correction amounts percentage for the same facilities. The pattern of these evolutions is similar for each facility, and suggests an *in control* state with low variance, and some episodes of aberrant values. Some of these *out of control* situations appear to have some persistence as the early 2014 in the CSA Kpasse. My work will be to detect these episodes based on reported data.

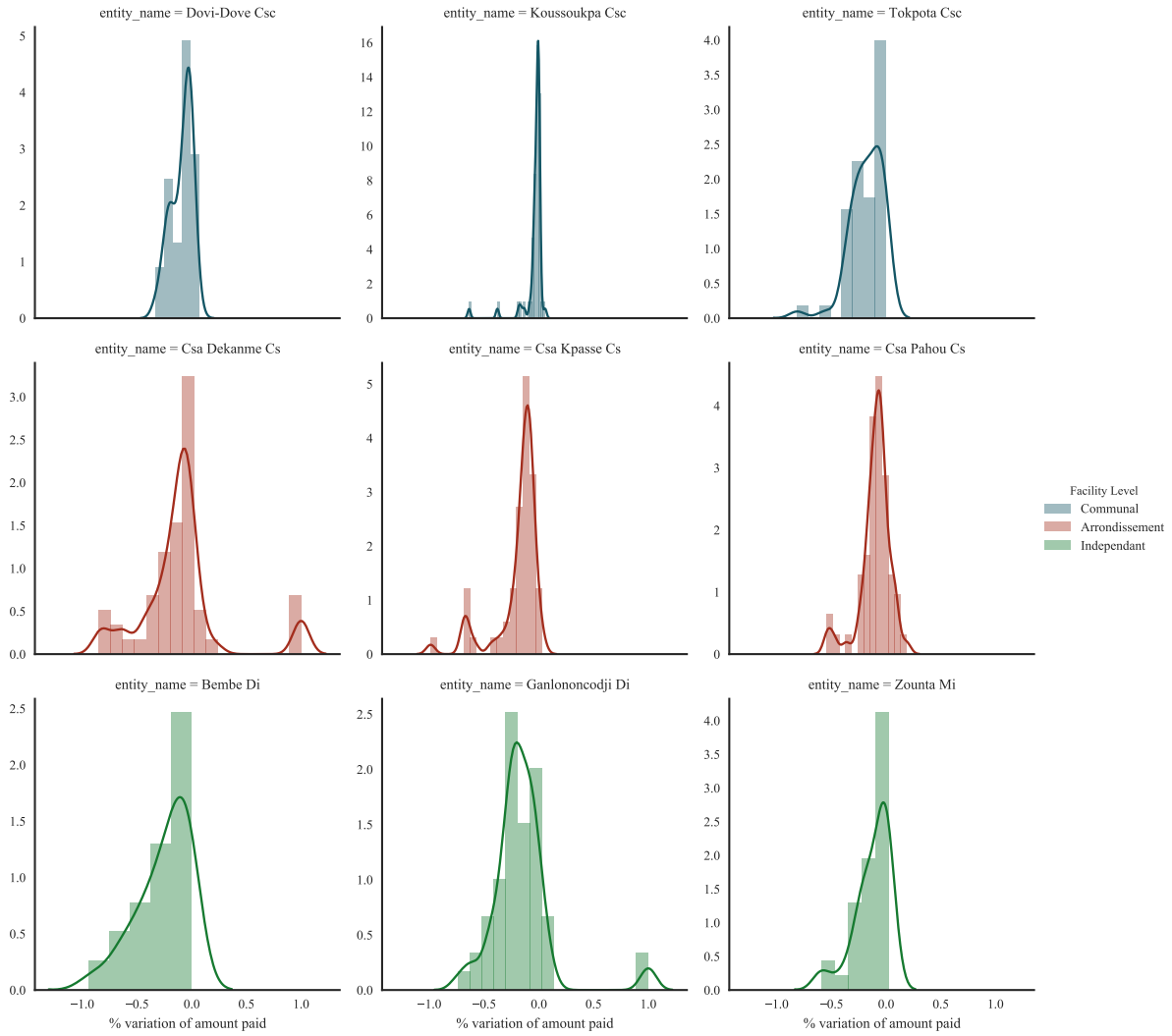


Figure 12: Distribution of payment correction percentage for a sample of indicators

4.3 Methods

4.3.1 Theoretical Framework

Each month, each facility sends a report, composed of indicators. Each indicator I has a unitary cost C . Thus, on month t , facility f will receive a payment of the amount :

$$P_{t,f} = \sum_i I_{i,t,f} C_{i,t,f}$$

When taking into consideration a monthly report from a facility, we want to measure the marginal and joint probabilities of two events Q and D , where :

1. $Q = 1$ if the quality of care in the facility appears in accordance with the quality in other facilities.
2. $D = 1$ if the data appears of sufficient quality to base payment of the facility on it.

We are thus most interested in three distinct joint events :

- $D = 0$ the data appears to be of insufficient quality to draw conclusions

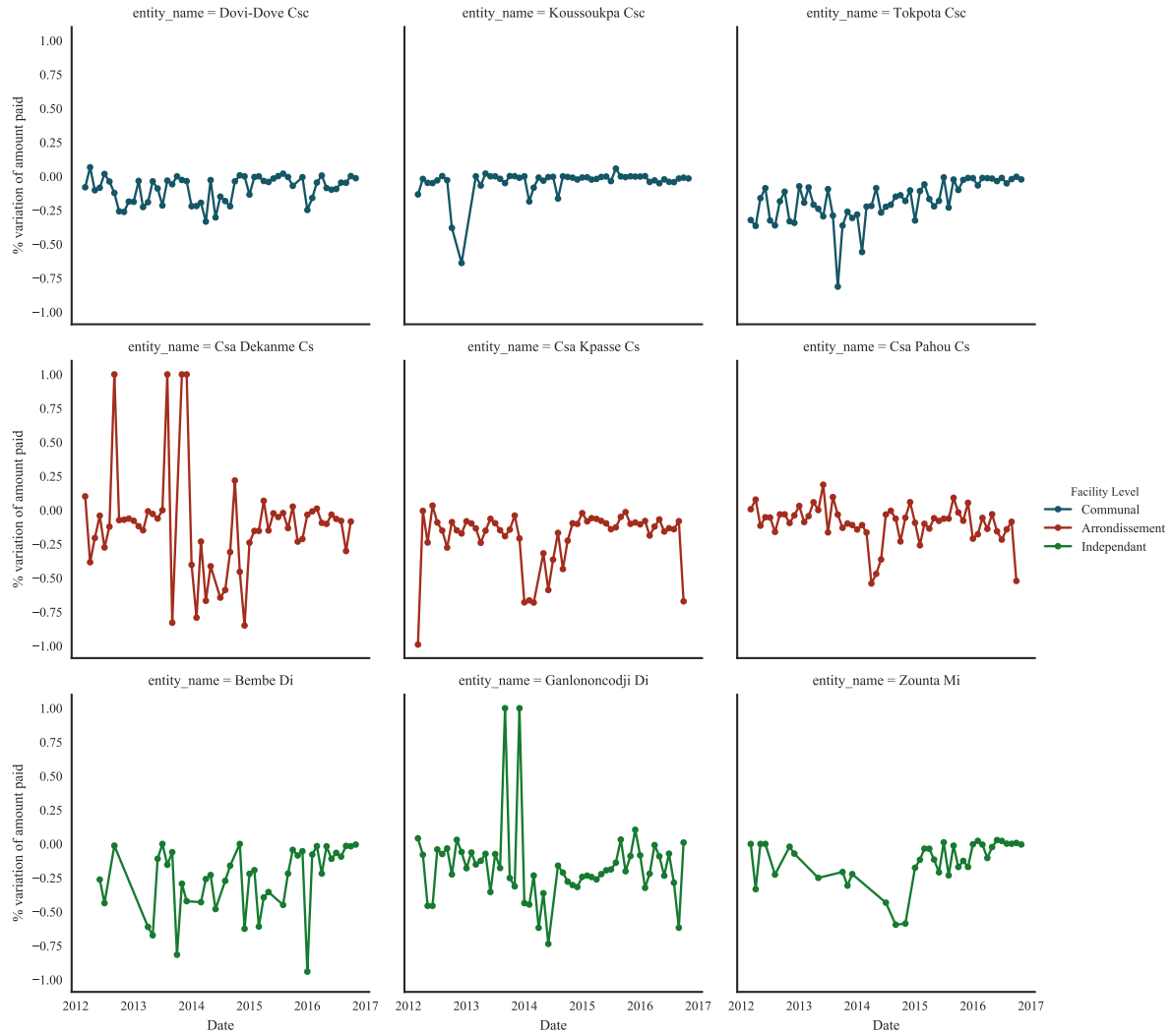


Figure 13: Evolution of payment correction percentage for a sample of indicators

- $D = 1 \cap Q = 0$ the data appears sufficient but quality of service seems poor
- $D = 1 \cap Q = 1$ the data appears sufficient and quality of service seems good

The definition of D results from a budgetary arbitrage. A report is considered of insufficient quality if the expected monetary value of the payment made to facility based on this report is too severely affected by data quality. We are then interested in the difference between the payment based on reported data (\bar{P}) and the expected payment based on all validated information available (P^*).

I will approach this problem in two steps. I will first test multiple approaches for process monitoring, to detect and spot unlikely data patterns. I will then define a approach to classify detected abnormalities as either $D = 0$ or $D = 1 \cap Q = 0$.

4.3.2 Data Screening

Methods for continuous data inspection and screening come from the industrial statistics domain and are increasingly applied for healthcare monitoring (Woodall 2006, Woodall & Montgomery 2014). Spiegelhalter et al. provided a nice overview of how these tools can be used for health systems, for three main functions of interest : rating, screening and surveillance (Spiegelhalter et al. 2012). These methods are based on the standardization of data distributions, and the analysis of observed data to

detect values diverging from their expected distributions. This second stage, geared toward hypothesis testing, can be made using visual tools based on adaptation of classical charts like Shewart charts, [Cumulative Sum \(CUSUM\)](#) control charts or [Exponentially Weighted Moving Average \(EWMA\)](#) charts. I will compare and combine three approaches to this data screening question.

Syndromic Surveillance A first approach is to define the expected distribution of a given series based on its past values. This approach is widely used in syndromic surveillance, and builds on a model-based [CUSUM](#) approach, where the expected count of cases for a given period is modeled based on historical data, and the [CUSUM](#) is updated at each period using the forecasting errors ([Fricker et al. 2008](#)). Fricker writes a general version of model-based [CUSUM](#) as monitoring $S(t)$ where :

$$S(t) = \max(0, S(t-1) + x(t) - k)$$

where $x(t) = \frac{Y(t) - \hat{Y}(t)}{\sigma_\epsilon}$ is a standardized prediction error for $Y(t)$ on σ_ϵ the variation of the model error, and k is a tuning parameter that will have to be optimized to obtain detection the desired characteristics for our model, including the expected monetary cost of error.

A reference implementation of this approach was developed by Farrington et al. ([Farrington et al. 1996](#)) and updated by Noufaily et al. ([Noufaily et al. 2013](#)), and is implemented in *R* ([Salmon 2016](#)). I will implement this approach to test its ability to detect and differentiate between data quality issues and regime changes for certain indicators series. I may also use the Z-Score aggregation approach described in ([Bardsley et al. 2009](#)) to aggregate multiple indicators result for a given facility month.

Hierarchical Modeling A second approach defines expected values from a facility based on a hierarchical model including all facilities. Using Bayesian hypothesis testing strategies based on multiple sampling from the posterior distribution of facility level random effects from this hierarchical model, this approach allows me to spot unlikely numbers for a given facility as well as facility level unusual performances for a given indicator ([Ohlssen et al. 2007](#)). For time t^* , $\tilde{I}_{i,t^*,f}$ will be simulated by first simulating parameters α_f and β_i and second drawing from the corresponding Poisson distribution using the different parameters.

Profile Monitoring Finally, we want to be able to detect pattern variations in reports. Indeed, the methods we described until now are targeted at monitoring univariate series for which desirable variations are clearly identified. In *in control* situations, the number of cases of monitored diseases are expected to stay low, and mortality rates are supposed to stay low. In the monitoring of [RBF](#) data, meanwhile, there is no clear denominator to construct rates, and there is little sense of what an *in control* situation should look like for individual indicators. Profile monitoring is an increasingly popular research area ([Woodall & Montgomery 2014](#)), and has many trends and applications, but I couldn't find any application in the healthcare sector. Its main idea is to monitor how indicators in the monitored process are related by a functional relationship, instead of monitoring individual indicators. This relationship can be specified in parametric ([Mahmoud 2011](#)) or non parametric ways([Chicken 2011](#)), and I will explore the options offered by the latter approach for the type of data we have.

4.3.3 Threshold Definition

For each of these approaches, I will need to define thresholds to define unlikely data. This threshold will be a combination of analysis of the deviation of the observed indicators from the expected distributions, and will also incorporate an element of financial risk for. We noted the constraints of the [RBF](#), which should aim at spending their budget efficiently, but ensure facilities have sufficient budget to operate their basic activities. I thus define the threshold of data credibility from a financial perspective as :

- For the upper limit, we want to identify the risk that the cost of bad data for the system exceeds a certain threshold. This can be defined as

$$\bar{P} - P^* > \eta_1 \Rightarrow D = 0$$

with $\eta_1 > 0$.

- For the lower limit, we consider the risk of underfunding the facilities, if the payment claimed is significantly smaller than the expected payment. This can be defined as

$$\frac{\bar{P}}{P^*} > \eta_2 \Rightarrow D = 0$$

with $0 < \eta_2 < 1$.

For any given pair of thresholds (η_1, η_2) , for each report computed in month t for facility f , I am thus interested in evaluating :

$$p(D = 0) = p(\bar{P} - P^* > \eta_1 \cup \frac{P^*}{\bar{P}} > \eta_2)$$

$$p(D = 0) = 1 - p(\bar{P} - P^* < \eta_1 \cap \frac{P^*}{\bar{P}} < \eta_2)$$

For any given month t^* in the data, using validated data from previous periods $t < t^*$, I will sequentially evaluate each approach in this section. I will be able to evaluate the ability of each approach to identify low quality data, or data that is too different from expected its values.

Using standard diagnostic evaluation methods, I will test and evaluate the ability of each approach to properly identify aberrant data. I will also test the benefit of combining these different approaches to improve the diagnostic performances. Once this will be done, I will try to give a better understanding of whether the spotted abnormalities are due to a data quality issue, or to problems in health services provision.

4.3.4 Risk Classification

As described in section 4.3.1, I am intersted in differentiating as much as possible between three situations :

- $D = 0$ the data appears to be of insufficient quality to draw conclusions
- $D = 1 \cap Q = 0$ the data appears sufficient but quality of service seems poor
- $D = 1 \cap Q = 1$ the data appears sufficient and quality of service seems good

The first step of the work, by screening out situations that appear to be *in control*, will allow me to rule out situations that appear to be $D = 1 \cap Q = 1$. Meanwhile, I still want to differentiate between $D = 0$ and $D = 1 \cap Q = 0$ for situations that have raised an alarm during data screening. For any indicator or report spotted as behaving in an unexpected way, I will want to understand if the problem detected can be linked to data quality issues or to a modification of the conditions of care in the facility.

I will try to solve this question by handling it as a classification problem. Using the confirmation data, I will be able to classify data as $D = 1$ or $D = 0$. Using a combination of the outputs of each data screening method as features I will implement and test simple supervised classification methods, to differentiate between $D = 0$ and $D = 1$ in abnormal data. This second layer of classification will thus allow me to differentiate between the $D = 0$ and $D = 1 \cap Q = 0$ situations of abnormal observed data.

4.3.5 Validation and Supervision Planning optimization

Performing algorithmic verification comes with a risk of creating reinforcing errors. Since we fit the expected distribution of I_i using only the data that was previously validated on the field, the periodic creation of such validation data points for every facilities is essential to ensure these expected distributions are still in line with reality. Meanwhile, in facilities that demonstrate good measured data quality, data will be verified less frequently than previously. This holds potential risks for our approach to be able to detect variations in performance. The question we know ask is thus, what is the minimum of validated data that has to be available in a facility to allow algorithmic validation to perform well ?

The question at stake here is to find a best performing arbitrage between data verification costs and benefits of verification. To identify a best performing data validation strategy, we will compare four strategies :

1. **Reactive validation only** : On site data validation is only performed in facilities were data quality issues have been suspected algorithmically.
2. **Reactive and yearly validation** : On site data validation is performed in facilities in which data issues have been suspected, but every facility has to be visited at least once per year.
3. **Reactive and bi-annual validation** : Same as the previous scenario, but every facility has to be visited at least once per semester.
4. **Reactive and quarterly validation** : Same as the previous scenario, but every facility has to be visited at least once per quarter.

The comparison will be made by simulating progressively the implementation of each strategy on the available data. We will be describing the performances of each approach in the same way we described the performance of the full model in section 4.3.2.

4.4 Timeline

The data for this aim is already available and has been partially cleaned. I envision needing four months to implement the different data monitoring approaches, and four months for risk classification and supervision planning optimization.

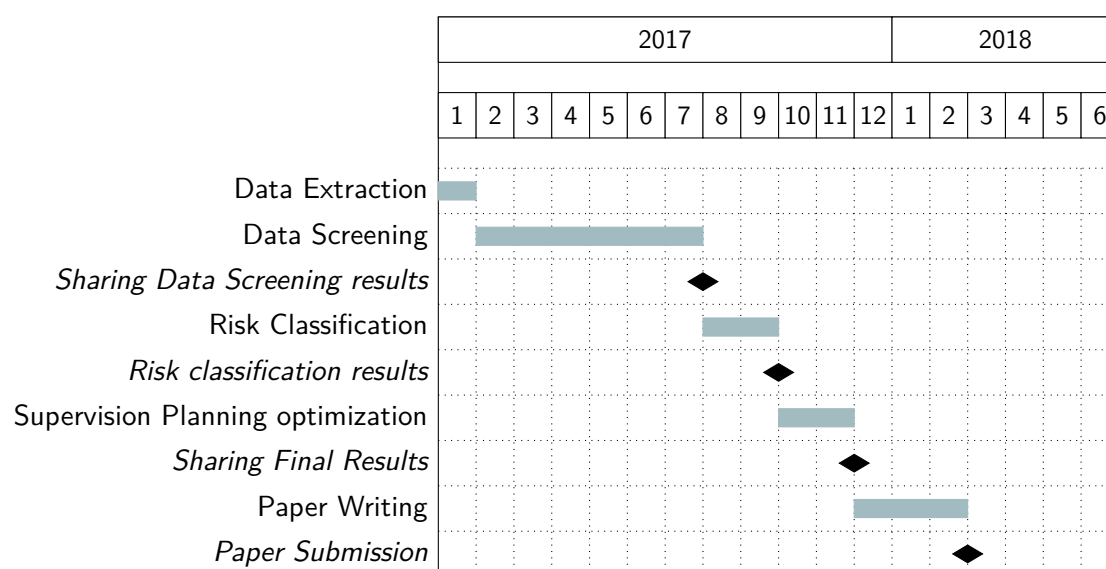


Figure 14: Gantt Chart for Aim 2

References

- Abou-Zahr, C. & Boerma, T. (2005), 'Health information systems: the foundations of public health', *Bulletin of the World Health Organization* **83**, 578–583.
- Abou-Zahr, C., Gollogly, L. & Stevens, G. (2010), 'Better data needed: everyone agrees, but no one wants to pay', *The Lancet* **375**(9715), 619–621.
- Antony, M., Bertone, M. P. & Barthes, O. (2017), 'Exploring implementation practices in results-based financing: the case of the verification in Benin', *BMC Health Services Research* **17**(1).
- Appadurai, A. (1996), Number in the Colonial Imagination, in 'Modernity at Large', University of Minnesota Press, pp. 114–135.
- Bambas Nolen, L., Braveman, P., Dachs, N., Delgado, I., Gakidou, E., Moser, K., Rolfe, L., Vega, J. & Zarowsky, C. (2005), 'Strengthening health information systems to address health equity challenges', *Bulletin of the World Health Organization* **83**(8), 597–603.
- Bardsley, M., Spiegelhalter, D. J., Blunt, I., Chitnis, X., Roberts, A. & Bharania, S. (2009), 'Using routine intelligence to target inspection of healthcare providers in England', *Quality and Safety in Health Care* **18**(3), 189–194.
- Bergeron, H. & Castel, P. (2014), Savoirs de politique, politique des savoirs, in 'Sociologie politique de la santé', Presses Universitaires de France.
- Beyersmann, J. & Schrade, C. (2017), 'Florence Nightingale, William Farr and competing risks', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **180**(1), 285–293.
- Boerma, T. (2013), 'Public health information needs in districts.', *BMC Health Services Research* **13**(Suppl 2), S12–S12.
- Bonneuil, C. (2000), 'Development as Experiment: Science and State Building in Late Colonial and Postcolonial Africa, 1930-1970', *Osiris* **15**, 258–281.
- Braa, J., Hanseth, O., Heywood, A., Woinshet, M. & Shaw, V. (2007), 'Developing Health Information Systems in Developing Countries : the flexible Standards Strategy', *Management Information Systems Quarterly* **31**(Suppl 1), 381–402.
- Braa, J., Heywood, A. & Sahay, S. (2012), 'Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania.', *Bulletin of the World Health Organization* **90**(5), 379–84.
- Brinkhof, M., Spycher, B., Yiannoutsos, C., Weigel, R., Wood, R., Messou, E., Boulle, A., Egger, M., Sterne, J. & leDEA (2010), 'Adjusting Mortality for Loss to Follow-Up: Analysis of Five ART Programmes in Sub-Saharan Africa', *PLoS ONE* **5**(11), e14149.
- Bryant, J. & Graham, P. (2013), 'Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources', *Bayesian Analysis* **8**(3), 591–622.
- Chaperon, J., Villard, J. M. & Riou, F. (1988), 'L'information médicale enjeu de la gestion hospitalière', *Politiques et Management Public* **6**(2), 35–46.
- Chi, B., Cantrell, R., Mwango, A., Westfall, A., Mutale, W., Limbada, M., Mulenga, L., Vermund, S. & Stringer, J. (2010), 'An Empirical Approach to Defining Loss to Follow-up Among Patients Enrolled in Antiretroviral Treatment Programs', *American Journal of Epidemiology* **171**(8), 924–931.

- Chi, B., Yiannoutsos, C., Westfall, A., Newman, J., Zhou, J., Cesar, C., Brinkhof, M., Mwango, A., Balestre, E., Carriquiry, G., Sirisanthana, T., Mukumbi, H., Martin, J., Grimsrud, A., Bacon, M., Thiebaut, R. & leDEA (2011), 'Universal Definition of Loss to Follow-Up in HIV Treatment Programs: A Statistical Analysis of 111 Facilities in Africa, Asia, and Latin America', *PLoS Medicine* **9**, e1001111.
- Chicken, E. (2011), Nonparametric Nonlinear Profiles, in 'Statistical Analysis of Profile Monitoring', Wiley series in probability and statistics, Wiley, pp. 157–188.
- Cordell, D., Ittmann, K. & Maddox, G. (2010), Counting Subjects - Demography and Empire, in K. Ittmann, D. Cordell & G. Maddox, eds, 'The demographics of empire: the colonial order and the creation of knowledge', Ohio University Press.
- Dalal, R., MacPhail, C., Mqhayi, M., Wing, J., Feldman, C., Chersich, M. & Venter, W. (2008), 'Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa', *Journal of Acquired Immune Deficiency Syndromes* **47**(1), 101.
- d'Altilia, J.-P., de Lamalle, J.-P., de Caluwe, P., Greindl, I., Lecharlier, F. & Wodon, A. (2005), *Système d'Information Sanitaire*, l'Harmattan.
- de Martonne, E. (1931), La cartographie à l'exposition coloniale de Vincennes, in 'Annales de Géographie', Vol. 40, JSTOR, pp. 449–478.
- Desrosières, A. (1993), *La Politique des grands nombres*, La Découverte.
- Desrosières, A. (1997), 'The administrator and the scientist: How the statistical profession has changed', *Statistical Journal of the UN Economic Commission for Europe* **14**(1), 31–50.
- Desrosières, A. (2014), *Prouver et Gouverner*, La Découverte, Paris.
- Egger, M., Spycher, B., Sidle, J., Weigel, R., Geng, E., Fox, M., MacPhail, P., van Cutsem, G., Messo, E., Wood, R., Nash, D., Pascoe, M., Dickinson, D., Etard, J.-F., McIntyre, J., Brinkhof, M. & leDEA (2011), 'Correcting Mortality for Loss to Follow-Up: A Nomogram Applied to Antiretroviral Treatment Programmes in Sub-Saharan Africa', *PLoS Medicine* **8**(1), e1000390.
- Farrington, C. P., Andrews, N. J., Beale, A. D. & Catchpole, M. A. (1996), 'A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease', *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **159**(3), 547.
- Fawcett, C. B. (1935), 'Population Maps', *The Geographical Journal* **85**(2), 142.
- Fienberg, S. E. (2011), 'Bayesian Models and Methods in Public Policy and Government Settings', *Statistical Science* **26**(2), 212–226.
- Forster, M., Bailey, C., Brinkhof, M., Graber, C., Boule, A., Spohr, M., Balestre, E., May, M., Keiser, O., Jahn, A. & Egger, M. (2008), 'Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings', *Bulletin of the World Health Organization* **86**, 939–947.
- Fox, M. P., Larson, B. & Rosen, S. (2012), 'Defining retention and attrition in pre-antiretroviral HIV care: proposals based on experience in Africa: Editorial', *Tropical Medicine & International Health* **17**(10), 1235–1244.
- Fricker, R. D., Hegler, B. L. & Dunfee, D. A. (2008), 'Comparing syndromic surveillance detection methods: EARS versus a CUSUMbased methodology', *Statistics in Medicine* **27**(17), 3407–3429.

- Geng, E., Glidden, D., Emenyonu, N., Musinguzi, N., Bwana, M., Neilands, T., Muyindike, W., Yiannoutsos, C., Deeks, S., Bangsberg, D. et al. (2010), 'Tracking a sample of patients lost to follow-up has a major impact on understanding determinants of survival in HIV-infected patients on antiretroviral therapy in Africa', *Tropical Medicine & International Health* **15**, 63–69.
- Gervais, R. & Mandé, I. (2010), How to Count the Subjects of Empire ? Steps towards an Imperial Demography in French West Africa before 1946, in K. Ittmann, D. Cordell & G. Maddox, eds, 'The demographics of empire: the colonial order and the creation of knowledge', Ohio University Press.
- Gething, P., Noor, A., Gikandi, P., Ogara, E., Hay, S., Nixon, M., Snow, R. & Atkinson, P. (2006), 'Improving imperfect data from health management information systems in Africa using space-time geostatistics.', *PLoS medicine* **3**(6), e271–e271.
- Gething, P., Noor, A., Goodman, C., Gikandi, P., Hay, S., Sharif, S., Atkinson, P. & Snow, R. (2007), 'Information for decision making from imperfect national data: tracking major changes in health care use in Kenya using geostatistics', *BMC Medicine* **5**(1), 37.
- Gimbel, S., Micek, M., Lambdin, B., Lara, J., Karagianis, M., Cuembelo, F., Gloyd, S., Pfeiffer, J. & Sherr, K. (2011), 'An assessment of routine primary care health information system data quality in Sofala Province , Mozambique', *Population Health Metrics* **9**(1), 12–12.
- Glèlè Ahanhanzo, Y., Ouendo, E.-M., Kpozèhouen, A., Levêque, A., Makoutodé, M. & Dramaix-Wilmet, M. (2015), 'Data quality assessment in the routine health information system: an application of the Lot Quality Assurance Sampling in Benin', *Health Policy and Planning* **30**(7), 837–843.
- Grimrud, A. T., Cornell, M., Egger, M., Boule, A. & Myer, L. (2013), 'Impact of definitions of loss to follow-up (LTFU) in antiretroviral therapy program evaluation: variation in the definition can have an appreciable impact on estimated proportions of LTFU.', *Journal of clinical epidemiology* **66**(9), 1006–13.
- Hahn, D., Wanjala, P. & Marx, M. (2013), 'Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics', *Global Health Action* **6**(1), 1–10.
- Health Metrics Network (2008), Framework and standards for country health information system, Technical report.
- Henriques, J., Pujades-Rodriguez, M., McGuire, M., Szumilin, E., Iwaz, J., Etard, J.-F. & Ecochard, R. (2012), 'Comparison of Methods to Correct Survival Estimates and Survival Regression Analysis on a Large HIV African Cohort', *PLoS ONE* **7**(2), e31706.
- Heunis, C., Wouters, E., Kigozi, G., Engelbrecht, M., Tsibolane, Y., van der Merwe, S. & Mothlanke, S. (2011), 'Accuracy of Tuberculosis Routine Data and Nurses' views of the TB-HIV Information System in the Free State, South Africa', *Journal of the Association of Nurses in AIDS Care* **22**(1), 67–73.
- Ioannidis, J., Taha, T., Kumwenda, N., Broadhead, R., Mtimavalye, L., Miotti, P., Yellin, F., Contopoulos-Ioannidis, D. & Biggar, R. (1999), 'Predictors and impact of losses to follow up in an HIV1 perinatal transmission cohort in Malawi.pdf', *International Journal of Epidemiology* **28**, 769–775.
- Jacucci, E., Shaw, V. & Braa, J. (2006), 'Standardization of health information systems in South Africa: The challenge of local sustainability', *Information Technology for Development* **12**(3), 225–239.

- Jerven, M. (2013), *Poor Numbers: how we are misled by African development statistics and what to do about it*, Cornell University Press.
- Kateb, K. (1998), 'La gestion statistique des populations dans l'empire colonial français', *Histoire & Mesure* **13**(1-2), 77–111.
- Kiberu, V. M., Matovu, J. K. B., Makumbi, F., Kyozi, C., Mukooyo, E. & Wanyenze, R. K. (2014), 'Strengthening district-based health reporting through the district health management information software system: the Ugandan experience.', *BMC medical informatics and decision making* **14**(1), 40–40.
- Kihuba, E., Gathara, D., Mwinga, S., Mulaku, M., Kosgei, R., Mogoa, W., Nyamai, R. & English, M. (2014), 'Assessing the ability of health information systems in hospitals to support evidence-informed decisions in Kenya', *Global health action* **7**, 24859–24859.
- Kwong-Leung Yu, J., Chen, S., Wang, K.-Y., Chang, C.-S., Makombe, S., Schouten, E. & Harries, A. (2007), 'True outcomes for patients on antiretroviral therapy who are lost to follow-up in Malawi', *Bulletin of the World Health Organization* **85**(7), 550–554.
- Latour, B. (2004), 'Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern', *Critical Inquiry* **30**(2), 225–248.
- Lebouché, B., Yazdanpanah, Y., Gerard, Y., Sissoko, D., Ajana, F., Alcaraz, I., Boitte, P., Cadoré, B. & Mouton, Y. (2006), 'Incidence rate and risk factors for loss to follow-up in a French clinical cohort of HIV-infected patients from January 1985 to January 1998', *HIV medicine* **7**(3), 140–145.
- Lecuyer, B.-P. (1987), Médecins et observateurs sociaux : les Annales d'Hygiène publique et de médecine légale (1820-1850), in 'Pour une Histoire de la Statistique', Vol. 1, INSEE, Paris, pp. 445–476.
- Ledikwe, J., Grignon, J., Lebelonyane, R., Ludick, S., Matshediso, E., Sento, B. W., Sharma, A. & Semo, B.-w. (2014), 'Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana.', *Health Research Policy and Systems* **12**, 7–7.
- Linard, C., Gilbert, M., Snow, R. W., Noor, A. M. & Tatem, A. J. (2012), 'Population Distribution, Settlement Patterns and Accessibility across Africa in 2010', *PLoS ONE* **7**(2), e31743.
- Lippeveld, T. (2000), Routine Data Collection Methods, in 'Design and Implementation of Health Information Systems', pp. 88–113.
- Little, R. (2012), 'Calibrated Bayes, an alternative inferential paradigm for official statistics', *Journal of Official Statistics* **28**(3), 309–372.
- Lurton, G. (2012), Report on the assessment of Health Information System for HIV in Sierra Leone, Consultancy Report, Freetown.
- Macfarlane, S. (2005), 'Harmonizing health information systems with information systems in other social and economic sectors', *Bulletin of the World Health Organization* **83**, 590–596.
- Mahapatra, P., Shibuya, K., Lopez, A., Coullare, F., Notzon, F., Rao, C. & Szreter, S. (2007), 'Civil registration systems and vital statistics: successes and missed opportunities', *The Lancet* **370**(9599), 1653–1663.
- Mahmoud, M. (2011), Simple Linear Profiles, in 'Statistical Analysis of Profile Monitoring', Wiley series in probability and statistics, Wiley, pp. 21–92.

- Makombe, S., Hochgesang, M., Jahn, A., Tweya, H., Hedt, B., Chuka, S., Kwong-Leung Yu, J., Aberle-Grasse, J., Pasulani, O., Bailey, C., Kamoto, K., Schouten, E. & Harries, A. (2008), 'Assessing the quality of data aggregated by antiretroviral treatment clinics in Malawi', *Bulletin of the World Health Organization* **86**(4), 310–314.
- McGuire, M., Munyenembe, T., Szumilin, E., Heinzelmann, A., Le Paih, M., Bouithy, N. & Pujades-Rodríguez, M. (2010), 'Vital status of pre-ART and ART patients defaulting from care in rural Malawi', *Tropical Medicine & International Health* **15**, 55–62.
- Mikkelsen, L., Phillips, D., Abou-zahr, C., Setel, P., De Savigny, D., Lozano, R. & Lopez, A. (2015), 'A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress', *The Lancet* **386**(10001), 1395–1406.
- Moh, R., Danel, C., Messou, E., Ouassa, T., Gabillard, D., Anzian, A., Abo, Y., Salamon, R., Bissagnene, E., Seyler, C., Eholie, S. & Anglaret, X. (2007), 'Incidence and determinants of mortality and morbidity following early antiretroviral therapy initiation in HIV-infected adults in West Africa', *AIDS* **21**(18), 2483.
- Mphatswe, W., Mate, K. S., Bennett, B., Ngidi, H., Reddy, J., Barker, P. M. & Rollins, N. (2012), 'Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa', *Bulletin of the World Health Organization* **90**(3), 176–82.
- Mugavero, M., Westfall, A., Zinski, A., Davila, J., Drainoni, M.-L., Gardner, L. I., Keruly, J., Malitz, F., Marks, G., Metsch, L., Wilson, T. & Giordano, T. (2012), 'Measuring retention in HIV care: the elusive gold standard', *Journal of Acquired Immune Deficiency Syndromes* **61**(5), 574–80.
- Murray, C. (2007), 'Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators', *The Lancet* **369**(9564), 862–873.
- Murray, C. & Frenk, J. (2008), 'Health metrics and evaluation: strengthening the science', *The Lancet* **371**, 1191–1199.
- Musgrove, P. (2011), Financial and other rewards for good performance or results: a guided tour of concepts and terms and a short glossary, Background Brief, The World Bank.
- Mutale, W., Chintu, N., Amoroso, C., Awoonor-Williams, K., Phillips, J., Baynes, C., Michel, C., Taylor, A. & Sherr, K. (2013), 'Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative', *BMC health services research* **13 Suppl 2**(Suppl 2), S9–S9.
- Mutemwa, R. I. (2006), 'HMIS and decision-making in Zambia: Re-thinking information solutions for district health management in decentralized health systems', *Health Policy and Planning* **21**(1), 40–52.
- Nisingizwe, M. P., Iyer, H. S., Gashayija, M., Hirschhorn, L., Amoroso, C., Wilson, R., Rubyutsa, E., Gaju, E., Basinga, P., Muhire, A., Binagwaho, A. & Hedt-Gauthier, B. (2014), 'Toward utilization of data for program management and evaluation: Quality assessment of five years of health management information system data in Rwanda', *Global Health Action* **7**(1).
- Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N. & Charlett, A. (2013), 'An improved algorithm for outbreak detection in multiple surveillance systems', *Statistics in Medicine* **32**(7), 1206–1222.
- Ohlssen, D., Sharples, L. & Spiegelhalter, D. (2007), 'A hierarchical modelling framework for identifying unusual performance in health care providers', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **170**(4), 865–890.

- Porter, T. (1996), *Trust in numbers: The pursuit of objectivity in science and public life*, Princeton University Press.
- Rambert, G. (1922), La cartographie à l'exposition coloniale de Marseille, in 'Annales de Géographie', Vol. 31, JSTOR, pp. 433–448.
- RHINO (2003), introducing an analytical framework for understanding performance of routine health information systems.pdf, Technical report.
- Ronveaux, O., Rickert, D., Hadler, S., Groom, H., Lloyd, J., Bchir, A. & Birmingham, M. (2005), 'The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems', *Bulletin of the World Health Organization* **83**(7), 503–510.
- Rottenburg, R., Merry, S., Park, S.-J. & Mugler, J., eds (2016), *The World of Indicators - The Making of Governmental Knowledge through Quantification*, Cambridge Studies in Law and Society, Cambridge University Press, Cambridge.
- Rousseau, R. (1929), La population du Sénégal en 1926, in 'Annales de Géographie', Vol. 38, JSTOR, pp. 399–403.
- Salmon, M. (2016), Advances in count time series monitoring for public health surveillance, PhD thesis, lmu.
- Sandefur, J. & Glassman, A. (2013), 'The Political Economy of Bad Data: Evidence from African Survey & Administrative Statistics', *The Journal of Development Studies* **51**(2), 116–132.
- Shepherd, B. E., Blevins, M., Vaz, L. M. E., Moon, T. D., Kipp, A. M., José, E., Ferreira, F. G. & Vermund, S. H. (2013), 'Impact of definitions of loss to follow-up on estimates of retention, disease progression, and mortality: application to an HIV program in Mozambique.', *American journal of epidemiology* **178**(5), 819–28.
- Shibuya, K., Scheele, S. & Boerma, T. (2005), 'Health statistics : time to get serious', *Bulletin of the World Health Organization* **83**(10), 722–723.
- Shrestha, L. & Bodart, C. (2000), Data Transmission, data processing and data quality, in 'Design and Implementation of Health Information Systems', pp. 128–145.
- Spiegelhalter, D., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C. & Grigg, O. (2012), 'Statistical methods for healthcare regulation: rating, screening and surveillance', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **175**(1), 1–47.
- Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. (2015), 'Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data', *PLOS ONE* **10**(2), e0107042.
- Tassie, J.-M., Malateste, K., Pujades-Rodríguez, M., Poulet, E., Bennett, D., Harries, A., Mahy, M., Schechter, M., Souteyrand, Y., Dabis, F. & leDEA (2010), 'Evaluation of Three Sampling Methods to Monitor Outcomes of Antiretroviral Treatment Programmes in Low- and Middle-Income Countries', *PLoS ONE* **5**, e13899.
- The Global Fund (2014), 'The Global Fund Measurement Guidance'.
- Thieren, M. (2005), 'Health information systems in humanitarian emergencies', *Bulletin of the World Health Organization* **83**, 584–589.
- Van Cutsem, G., Ford, N., Hildebrand, K., Goemaere, E., Mathee, S., Abrahams, M., Coetzee, D. & Boulle, A. (2011), 'Correcting for Mortality Among Patients Lost to Follow Up on Antiretroviral Therapy in South Africa: A Cohort Analysis', *PLoS ONE* **6**, e14684.

- Verguet, S., Lim, S. S., Murray, C. J. L., Gakidou, E. & Salomon, J. A. (2013), 'Incorporating loss to follow-up in estimates of survival among HIV-infected individuals in sub-saharan Africa enrolled in antiretroviral therapy programs', *Journal of Infectious Diseases* **207**(1).
- Wagenaar, B. H., Sherr, K., Fernandes, Q. & Wagenaar, A. C. (2016), 'Using routine health information systems for well-designed health evaluations in low- and middle-income countries.', *Health policy and planning* **31**(1), 129–35.
- Walsham, G., Robey, D. & Sahay, S. (2007), 'Foreword: Special issue on information systems in developing countries', *Management Information Systems Quarterly* **31**(2), 317–326.
- Walsham, G. & Sahay, S. (2006), 'Research on information systems in developing countries: Current landscape and future prospects', *Information Technology for Development* **12**(1), 7–24.
- WHO (2011), Assessment of health facility data quality in Uganda, Technical report.
- Woodall, W. (2006), 'The use of control charts in health-care and public-health surveillance', *Journal of Quality Technology* **38**(2), 89–104.
- Woodall, W. & Montgomery, D. (2014), 'Some current directions in the theory and application of statistical process monitoring', *Journal of Quality Technology* **46**(1), 78.
- Yehia, B. R., Fleishman, J. A., Metlay, J. P., Korthuis, P. T., Agwu, A. L., Berry, S. A., Moore, R. D. & Gebo, K. A. (2012), 'Comparing different measures of retention in outpatient HIV care', *AIDS* **26**(9), 1131–1139.
- Yiannoutsos, C. T., An, M.-W., Frangakis, C. E., Musick, B. S., Braitstein, P., Wools-Kaloustian, K., Ochieng, D., Martin, J. N., Bacon, M. C., Ochieng, V. & Kimaiyo, S. (2008), 'Sampling-Based Approaches to Improve Estimation of Mortality among Patient Dropouts: Experience from a Large PEPFAR-Funded Program in Western Kenya', *PLoS ONE* **3**(12), e3843.