The efficiency of the local health systems: investigating the roles of health administrations and healthcare providers

Laura Anselmi¹, Mylene Lagarde², Kara Hanson³

¹ Manchester Centre For Health Economics, The University of Manchester, Manchester, UK
² Department of Social Policy, London School of Economics, London, UK
³ Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK

Corresponding author: Laura Anselmi
Address: Manchester Centre for health Economics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Room 4.311, Jean McFarlane Building, Oxford Road, Manchester M13 9PL
e-mail: laura.anselmi@manchester.ac.uk

The authors have no conflict of interest to declare.

The authors wish to thank the Ministry of Health, the Ministry of Finance and the National Institute of Statistics of Mozambique for providing all the data used in this analysis. Bina Langa, Amisse Momade, Samuel Cossa, Eduarda Ribeiro, Silvia Bignamini, Cidália Baloi, Célia Gonçalves, Ernesto Mazivila, Quinhas Fernandes, Jorge Perrolas, Costantino Muendane, Mario Tsaqice, Virginia Guibunda, Daniel Simone Nhachengo, Brigida Senate, Lígia Vilanculos, Salomão Lourenço, Lionel Senete provided invaluable support in cleaning the data and interpreting the results. All remaining errors are ours.
Abstract

The analysis of efficiency in healthcare has largely focused either on individual healthcare providers, or on sub-national health systems conceived as a unique decision making unit. However, in hierarchically organised national health services, two separate entities are responsible for turning financial resources into services at the local level: health administrations and healthcare providers. Their separate roles and the one of health administrations in particular have not been explicitly considered in efficiency analysis.

We applied Stochastic Frontier Analysis to district-level panel data from Mozambique to assess districts efficiency in delivering outpatient care. We first assessed the efficiency of the whole district considered as an individual decision making unit, and then we assessed separately the efficiency of health administrations and healthcare providers within the same district. We found that on average only 73% of the outpatient consultations deliverable with the given inputs were realized, with large differences in performance across districts. Individual districts performed differently in administrative or healthcare delivery functions. On average, a reduction of administrative inefficiency by 10 percentage points, for a given budget would increase by 0.2% the volume of services delivered per thousand population per year. Identifying and targeting the specific drivers of administrative inefficiencies can contribute to increase service delivery at local level.

Key words: health system efficiency, health administrations, stochastic frontier analysis, efficiency analysis, Mozambique
1. Introduction

Analysing the productivity and the organizational efficiency of public health systems is critical to address the concerns of policy makers who have to ensure the provision of public services which better satisfy population needs (Hollingsworth 2013). The existence of inefficiencies would imply that public resources could be better used elsewhere, in the healthcare or in other sectors, and that variations in efficiency could lead to uneven quality of service provision and perceptions of unfairness (Smith and Street 2005). Understanding how large inefficiency are and where in the production process improvements could be made, is key for Governments (Hollingsworth and Street 2006, Street and Hakkinen 2009)(Papanicolas and Smith 2014). Add refs

Efficiency in healthcare has been analysed as the difference between the observed and optimal productivity mostly of providers, including hospitals, but also individual practitioners, primary healthcare units, clinics, nursing homes, public health teams and primary healthcare facilities (Hollingsworth and Wildman 2003, Hollingsworth 2008, Hollingsworth and Peacock 2008, Hussey, de Vries et al. 2009, Kirigia, Sambo et al. 2011, Au, Hollingsworth et al. 2014). The efficiency of health systems has also been addressed, mostly at national level through cross-countries comparisons (Gravelle, Jacobs et al. 2003, Hollingsworth and Wildman 2003, Greene 2004, Retzlaff-Roberts, Chang et al. 2004), but also at sub-national level, by comparing states (Kathuria and Sankar 2005, Prachitha and Shanmugam 2012), districts (Kinfu 2013, Kinfu and Sawhney 2015) or lower level health authorities (Giuffrida 1999, Giuffrida, Gravelle et al. 2000, Giuffrida and Gravelle 2001, Puig-Junoy and Ortún 2004, Varela, Martins et al. 2010).
Sub-national health systems have been conceived and analysed as a decision-making units (DMU). However, in many national health systems organized on a hierarchical basis, sub-national health authorities are constituted by two types of entities with distinct functions. Health administrations (HAs) manage the financial resources to purchase health services from local healthcare providers, or to directly provide them with the human and physical inputs they require. Healthcare providers, deliver care according to their capacity and to the needs of their catchment population (WHO 2000, Robinson, Jakubowski et al. 2005). Due to their distinct roles, HAs and HFs can be considered separate DMUs. Measuring their performances separately and understanding how they influence the performance of the health system at sub-national level could provide insights about where improvements could be made (Cacace and Nolte 2011).

Previous studies have assessed cost-efficiency in contracting health services (Puig-Junoy and Ortún 2004) or the relative contribution of administrative costs to inefficiencies of local health authorities (Giuffrida, Gravelle et al. 2000), or have accounted for the hierarchical organization of healthcare delivery in assessing providers’ efficiency through multilevel models (Hauck, Rice et al. 2003, Jacobs, Smith et al. 2006). However, nor the efficiency of two separate DMUs with distinct roles and production processes (HAs and healthcare providers), nor the way in which local HAs’ efficiency influences health providers efficiency have been studies. The role of HAs is even more critical in low and middle-income countries (LMICs), where the scarcity of financial resources is exacerbated (WHO 2000). Further, due to the decentralization processes undergoing in many LMICs, HAs are acquiring growing responsibilities. However, very little is known about how they work, perform and affect the health system at the local level (Wickremasinghe, Hashmi et al. 2016).
We used stochastic frontier analysis (SFA) to assess the efficiency of health districts in delivering outpatient primary care in Mozambique. As in the existing literature, we first assumed that healthcare delivery at local level is the result of an aggregated production process, through which financial resources and healthcare inputs are simultaneously transformed into services. We then analysed separately the efficiency of HAs in managing financial resources to equip and staff health facilities (HFs), and the efficiency of HFs in using these inputs to deliver healthcare. We finally assessed the effect of HAs efficiency on healthcare delivery by including its measure in the district aggregated production function and in the HFs production function.

This study contributes to the literature on healthcare efficiency in three ways. First, the composite nature of local healthcare systems, resulting from the separate contributions of HAs and HFs to service delivery, is recognised and conceptualized. Second, HAs’ and HFs’ inefficiencies are separately measured and a method to quantify the effect of administrative inefficiency on the performance of HFs and district health systems is proposed. Third, the sources of bias arising from analysing the sub-national health systems as aggregated DMUs are identified.

By separately analysing the efficiency of the decision making units that contribute to healthcare delivery, HAs and HFs, our analysis draws attention upon the overlooked role of HAs. By quantifying the effect of administrative on overall district efficiency, and investigating some of the channels through which it acts, this study highlight how potentially very low cost interventions could target HAs to achieve a more effective use of financial resources at local level.
2. Study setting

The large majority of healthcare services in Mozambique are provided by the public sector. The National Health System (NHS) follows a centralised top-down hierarchical organization, including 10 provinces, 142 districts in the country, and the capital, Maputo City. Districts manage secondary and primary care, provided through district hospitals (DH), health centres (HC) and clinics (MISAU 2012). There is generally one DH or HC per district, often located in the major urban centre. DH and HC serve as district reference facility since along with basic primary care they provide inpatient care and, in DH, minor surgery. Acute and specialised care is provided by 10 provincial and central hospitals.

Like in many sub-Saharan and LMICs, the NHS in Mozambique relies heavily on districts, whose HAs organize service provision in line with the national targets and policies set by the Ministry of Health (MoH) (MISAU 2012, Wickremasinghe, Hashmi et al. 2016). Specifically, district HAs manage financial and non-financial resources to guarantee that HFs have the means to operate and deliver services and are therefore responsible for the staff and equipment input mix in HFs (HST 1998, MISAU 2002, MISAU 2012, Wickremasinghe, Hashmi et al. 2016). The minimum number of staff and equipment are defined by the Ministry of Health (MoH) for each type of HF (clinics, health centres and district hospitals) according to the catchment area of the facility and to whether it provides services such as surgery or maternal care and drive the distribution of healthcare inputs across districts and across HFs within the same district (MISAU 2002, MISAU 2012).

Since the decentralization reform, which began in 2007, the responsibility for hiring human resources (HR) has gradually been devolved to district HAs. District health administrators are responsible for opening staff vacancies, selecting candidates and legalizing the
recruitment according to the national administrative norms. Medical and clinical staff are still recruited by central or provincial administrations and allocated to districts. However, district HAs have first to communicate their needs to get the recruitment process started and then to follow-up with hiring and distributing staff across HFs, paying their salaries and managing their careers and benefits, and ultimately retain them. Housing for key health cadres, for example health technicians and maternal and child nurses, has emerged as the most important factor for HR retention in rural areas (Vio, Buffolano et al. 2013). The MoH recommends that a minimum number of housing should be built next to each HF (MISAU 2007). The management of housing benefit, including building, renting, maintaining and allocating houses, and the daily HR management, depends largely on district health administrators.

While the central or provincial level administrations are responsible for infrastructure building and major maintenance, and for the purchase and distribution of drugs and major equipment, district health administrators are responsible for the direct procurement of small items of equipment and consumables. Therefore, the degree to which the HF’s need for drugs, equipment and consumables is satisfied depends on the efficiency of HAs first in identifying them, and second in channelling and pursuing them with the higher administrative levels, or in directly purchasing what is needed.

District HFs provide healthcare to meet local populations’ demand given their capacity, which is determined by the infrastructure conditions and the availability of staff and equipment. The actions of HF staff operating under the various constraints they face determine service availability and responsiveness to the population needs, for example through HF opening times, attitude towards patients and quality of care provided. In some
cases a pro-active attitude of staff can lead to the involvement of community volunteers to support health personnel in performing basic health-care tasks or HFs maintenance (MISAU 2012). The capacity of HF to deliver healthcare is affected by HAs. The level of human and physical resources available in a HF are correlated and are the results of the HAs capacity to effectively plan, spend resources and allocate inputs. HAs may also influence HFs running through support, monitoring and coordination of various HFs within the district (Sherr, Cuembele et al. 2013)

District recurrent expenditure, including salaries and non-capital expenditure that cover running costs, is mostly funded through provincial and district government grants, as well as non-earmarked donor resources (referred here as provincial, district and donor financial resources for simplicity). District financial resources have progressively increased since the implementation of the decentralization reform in 2008 and represented around 40% of the total executed recurrent expenditure at district level in 2011. Provincial and donor financial resources are allocated to districts according to the number and type of HFs. District own revenues represent a negligible share of their total financial resource and are generated from small activities, such as occasionally renting out a room for local events, and minimally from user fees, which are very low and from which the large majority of the population is exempted. Other donor earmarked funds have supported district level activities, but those resources are often either difficult to track in a systematic way, or not managed by district administrations (MISAU 2012, Quresky, Cossa et al. 2016). All financial resources, at the time covered by this study, were managed by district HAs. No financial resources were managed directly by HFs.
3. A framework for analysis: aggregated versus two-steps district production process

We use two alternative frameworks for the analysis of efficiency at the district level and compare the results obtained.

First, we followed the approach taken in the existing studies on local health system efficiency implemented in a variety of settings (Giuffrida 1999, Giuffrida, Gravelle et al. 2000, Giuffrida and Gravelle 2001, Puig-Junoy and Ortún 2004, Kathuria and Sankar 2005, Varela, Martins et al. 2010, Prachitha and Shanmugam 2012, Kinfu 2013), and considered healthcare delivery at district level as an aggregated production process. We conceived the district as an individual decision making unit, made of HA and HFs that transform the available inputs (financial resources, staff and equipment) into healthcare, measured here through outpatient consultations (Figure 1).

FIGURE 1

Because of the referral system in place and the inequalities in inpatient and specialized care provision across the country, outpatient consultations are the only output delivered by all HFs and directly comparable across districts. Financial resources include the total recurrent expenditure from multiple sources but managed by the district. Since capital expenditure and infrastructure planning are managed at the central and provincial level, and there were there were minor changes in the number of HFs in the time period considered the number and type of HFs in the district are considered as given at the district level.

Second, we considered that in Mozambique, as in many hierarchically organised NHSs (WHO 2000, Robinson, Jakubowski et al. 2005), the financial management and healthcare delivery functions at local level depend on two separate DMUs. We modelled the healthcare production at the district level as a two-step process. (Figure 2)
In a first-step, HAs use financial resources from three different sources (i.e. government provincial expenditure, government district expenditure, and donor common fund expenditure) to fund HFs input-mix and recurrent expenditure. In a second-step, HFs use the HR and equipment input-mix to deliver healthcare.

HAs provide HFs with the input required (for example medical equipment, staff, drugs, timely payment of energy bills, maintenance commodities for cleaning and hygiene, food in inpatient wards, fuel for transport). HAs identify HFs’ input needs, plan the actions required to satisfy them and identify the options that could minimize costs (for example providing regular maintenance and organizing the purchasing services for all HFs in the district). Differences in HAs abilities (practically planning, management and supervision skills in the HA teams) influence the efficiency through which HAs use the available financial resources to fund HFs recurrent inputs needed to keep them functioning.

4. Methods

4.1. District efficiency assuming an aggregated production process

Following Fried, Lowell et al. (2006), we defined efficiency as the difference between the observed and optimal productivity, measured by ratios of output to input. We used stochastic frontier analysis (SFA) to compare the observed level of output to an optimal production frontier estimated based on the observed inputs and assumptions on the production function. We used a production function, rather than a cost function because information on prices was not available and because, due to central purchasing and national
directives, the prices of the inputs measured and included in the analysis tend not vary across districts.

We used SFA rather than alternative non-parametric techniques because it allows accounting explicitly for measurement error and quantifying the effect of each inputs on the production frontier and of factors outside the producers control (Kinflu 2013). Furthermore, when panel data are available, SFA models outperform non-parametric techniques if the assumed functional form for the production function is close to the underlying production technology (Giuffrida and Gravelle 2001). We assumed a Cobb-Douglas functional form for $f(.)$ to account for inputs complementarities in healthcare delivery, and assuming constant returns to scale.

We formulated the aggregated district SFA model as:

$$Y_{it} = f(X_{it}^D, \beta) E^D(Y_{it}, X_{it}^D)$$

(1)

where $Y_{it}$ is the output (outpatient consultations), delivered by district $i$ at time $t$. $X_{it}^D$ includes healthcare inputs, here defined as total financial resources for outpatient care and HFs staff and equipment. The production frontier $f(X_{it}^D, \beta)$ defines the maximum level of output ($Y_{it}$) attainable by district $i$ at time $t$, for a given combination of inputs ($X_{it}^D$). $\beta$ is the vector of parameters of the production function $f(.)$. $E^D(Y_{it}, X_{it}^D) = \frac{Y_{it}}{f(X_{it}^D, \beta)}$ is the district efficiency calculated as the ratio of observed production ($Y_{it}$) to the maximum number of consultations deliverable with the inputs available and the technology in use, $f(X_{it}^D, \beta)$.

We log transformed (1) to obtain a linear equation and applied the Pitt and Lee (1981) random effects model for panel data with time invariant inefficiency, to allow controlling for district characteristics fixed over time. Important efficiency changes would have been
unlikely since the time period considered covered the same government mandate. We derived the empirical specification of the Pitt and Lee (1981) SFA model following Greene (2008):

$$y_{it} = \alpha + \beta x_{it}^D + \delta z_{it}^D + \varepsilon_{it}^D$$

(2)

where \( \varepsilon_{it}^D = v_{it}^D - u_i^D \) with \( v_{it}^D \sim N(0, \sigma_{v}^2) \), \( u_i^D \sim N^+(0, \sigma_{u}^2) \) and \( \text{cov}(v_{it}^D, u_i^D) = 0 \)

\( y_{it} \) is the natural log of the yearly number of outpatient consultations per-capita.

\( x_{it}^D \) is a vector of inputs that includes: a) the (natural log of) annual recurrent expenditure for outpatient care per capita (see (Anselmi, Lagarde et al. 2015) for details on calculations), a proxy for financial resources, which is the sum of the executed district, provincial and donor expenditure in district \( i \); b) the (natural log of) average HF staff and equipment, measured using an index which averages across HFs the availability of six items: working car, autoclave, motorbike, number of basic, medium and high level trained health cadres. While the index is not comprehensive of all inputs required in the HF, evidence suggests that the availability of HR and major equipment is correlated with that of drugs and other medical supplies (Wagenaar, Gimbel et al. 2014).

For each HF we calculated the average ratio of available resources to the minimum standard set by norms for each item (MISAU 2002). Details on the calculation and further discussion of the implications are provided in Anselmi, Lagarde et al. (2015). We used existing norms as a benchmark to standardize availability of items across different types of HFs (MISAU 2012). We averaged across items to account for the complementarity between staff and equipment availability required to deliver effective healthcare and across HF within the same district to get a measure of availability in HFs rather than in the district.
$z_{it}^D$ is a vector of district characteristics that capture heterogeneity in the production technology and accommodates shifts of the production frontier. $z_{it}^D$ includes: a) the (natural log of) number of HFs per type per 100,000 population; b) the percentages of HFs with access to electricity and to running water (time variant); c) the percentages of population that are economically active and illiterate (time invariant); and d) provinces fixed effects to control for the influence of provincial management and other provinces characteristics. District characteristics and provincial dummies control for differences in the epidemiological profile, for which specific data were not available, and in differences in demand for healthcare across geographic areas (Anselmi, Lagarde et al. 2015). Provincial dummies also control for differences across provincial administrations which in a hierarchical system influence both HAs and HFs activities. The observed error term ($\varepsilon_{it}^D$) is a combination of the normally distributed stochastic error term ($v_{it}^P$) and the non-negative half-normally distributed term ($u_{it}^D$). $u_{it}^D$ defines how far the $i^{th}$ district operates below the stochastic production frontier and measures the time invariant inefficiency in production (Kinfu 2013).

We finally obtained estimates of the district relative efficiency score ($\widehat{E}_{it}^D$) from the Pitt and Lee (1981) as follows (Greene 2008):

$$\widehat{E}_{it}^D = e^{-u_{it}^D} \approx 1 - \widehat{u}_{it}^D$$

where $\widehat{u}_{it}^D$ is the input-oriented inefficiency scores obtained with the JLMS estimator (Jondrow, Materov et al. 1982). $u_{it}^D = \frac{-\lambda \varepsilon_{it}}{\sigma}$ where $\varepsilon$ is half-normally distributed and can be estimated using the values provided by the SFA estimates.
4.2. District efficiency assuming a two-step production process

First step: health administration efficiency

In analogy with the SFA model used for the district integrated production process, we derived the empirical specification of the district HA's SFA model as:

\[ y_{it}^A = \alpha + \beta x_{it}^A + \delta z_{it}^A + \epsilon_{it}^A \quad (4) \]

where \( \epsilon_{it}^A = v_{it}^A - u_i^A \) with \( v_{it}^A \sim N(0, \sigma_{v_{it}^A}^2) \), \( u_i^A \sim N^+(0, \sigma_{u_i^A}^2) \) and \( \text{cov}(v_{it}^A, u_i^A) = 0 \).

\( y_{it}^A \) is the natural log of the average HFs staffing and equipment index described in section 4.1.

\( x_{it}^A \) includes the (natural log) of the recurrent expenditure per HF, for each source of funding. To standardize the availability of financial resource per HF across districts accounting for differences in number and type of existing HFs, we weighted each HF type based on the estimated cost of the minimum staff and equipment they should have according to norms (MISAU 2012).

\( z_{it}^A \) includes the (natural log) of the total number of HFs, the percentages of HF with access to water and electricity, the average ratio of houses for personnel available and in good conditions compared to the minimum number established by norms (MISAU 2012) as a proxy for district capacity for HR retention, and provinces fixed effects.

As before in equation (3), we estimated \( \hat{E}_{t}^A \) that quantifies the efficiency of HA \( i \).

Second step: health facility efficiency
In the second step we derived the empirical specification of the district HFs production process as:

\[ y_{it} = \alpha + \beta y_{it}^A + \beta x_{it}^F + \delta z_{it}^F + \varepsilon_{it}^F \]  

(5)

where \( \varepsilon_{it}^F = v_{it}^F - u_{it}^F \) with \( v_{it}^F \sim N(0, \sigma_{vF}^2) \), \( u_{it}^F \sim N^+(0, \sigma_{uF}^2) \) and \( \text{cov}(v_{it}^F, u_{it}^F) = 0 \)

As in section 4.1, \( y_{it} \) is the natural log of the district total outpatient consultations per capita. \( y_{it}^A \) is the (natural log of) average availability of staffing and equipment with respect to norms, as in section 4.2.1. \( x_{it}^F \) includes the (natural log of the) number of DH, HC and clinics per 100,000 population. \( z_{it}^F \) includes the percentages of the district population that are economically active and illiterate (time invariant) to control for catchment population characteristics, the percentages of HFs in the district with electricity and with access to running water (time variant), to control for HF conditions, and provinces fixed effects.

As before, we estimated \( \widehat{E}_i^F \), which quantifies the average efficiency of HFs in district \( i \).

Having estimated the district, HA and HFs efficiency scores (\( \widehat{E}_i^D \), \( \widehat{E}_i^A \) and \( \widehat{E}_i^F \)), we plotted them in a graph and calculated their correlation and rank correlation to compare the measured district performance under the assumption of an aggregated or two-step production process.

4.3. Testing for the effect of health administration efficiency on district and health facility efficiency

We tested for the effect of district administrative efficiency on district and HFs production (and efficiency) by including \( \widehat{E}_i^A \) into the stochastic frontier defined in equations (2) and (5).
and looking at the significance of the coefficient associated with it, as done by Greene and Segal (2004) to test the effect of efficiency on profitability.

We evaluated the goodness of fit of the district and HF models without and with inclusion of HA efficiency looking at the log-likelihood and the Akaike information criterion (AIC), testing the hypothesis of a null coefficient associated with $\hat{E}_{iA}$ and performing a Likelihood-ratio (LR) test. Higher log-likelihood, lower AIC and rejection of the LR test null hypothesis indicated that the SFA model including administrative efficiency better fitted the data.

We further compared the aggregated district SFA in equation (2) to a specification including $\hat{E}_{iA}$ but not HF staff and equipment. If the inclusion of staff and equipment levels, along with HA efficiency, did not improve the fit for the data, we could infer that the availability of staff and equipment depends on financial resources available as well as the HAs efficiency in managing them. A two-step model accounting for administrative efficiency would then be preferable.

We tested if accounting for heterogeneity in administrative efficiency affected the evaluation of district and HFs performance, by calculating the correlation and rank-correlation of the efficiency scores estimated from (2) and (5) and from the relative counterparts including $\hat{E}_{iA}$. High correlation would indicate a small effect of HA on district performance.

**4.4. Data**

Data on HF type, staff and equipment for 2008 – 2011 were derived from the National Health Information System (NHIS), as provided by the MoH in June 2012 (MISAU 2012).
After cross-checking the information with provincial and district annual reports and with local health authorities, we retained only information on equipment items that was found reliable and for 1.5% of the HFs we replaced staff and equipment figures in a specific year with the 2008-2011 average, when the discrepancy between the two was higher than fifty percent.

Figures for provincial and district recurrent expenditure for 2008-2011 were obtained from the Ministry of Finance (MF 2012) and MoH budget execution reports (MISAU 2012). To get disaggregated district figures for all sources of expenditure, we assumed that donor and provincial expenditure were allocated to each district according to the number and type of HFs, as it used to be historically (MISAU 2012), and that donor expenditure benefited exclusively primary and secondary care, entirely managed at district level.

Population figures were based on annual projections from the 2007 Census (INE 2010). District socio-economic indicators were estimated by the National Institute of Statistics from 2007 Census data (INE 2008, INE 2010), and were therefore time invariant over the period considered in the analysis.

We merged data at the district level to obtain a single four-year panel database (2008-2011). We excluded Maputo City from the analysis, due to the presence of specialised and private healthcare providers, and eight districts, due to incomplete information. All models are estimated using Nlogit 5.

The descriptive statistics of the variables included in the analysis (Table 1), revealed considerable heterogeneity across districts in terms of service delivered, access to healthcare, expenditure and population characteristics, driven by differences in population density and urbanization. Heterogeneity in expenditure by source reflected the undergoing
decentralization process involving the gradual devolution of financial resources to districts, which explains the low but increasing district expenditure per-capita.

TABLE 1

5. Results

5.1. District efficiency assuming an aggregated or a two-step production process

The average district efficiency differed if evaluated at the HA, HF or aggregated level. Table 2 presents the coefficients of the stochastic frontier and the efficiency scores for district, HA and HFs.

On average districts delivered only 73% of the potentially deliverable outpatient consultations. The presence of inefficiency could be attributed to inefficiencies in healthcare delivery, since for given healthcare inputs HFs delivered only 74% of the attainable outpatient consultations, but also to administrative inefficiency. Indeed, on average HAs reached only 66% of the HFs staff and equipment levels that they could potentially achieve for the given financial resources. The significance of the variance parameters confirmed the presence of inefficiency in each of the production processes considered.

TABLE 2

When healthcare delivery was analyzed as a unique district process, as expected, we found that staff and equipment, financial resources per capita, the presence of a district hospital and the proportion of HF with access to water and electricity were positively associated with the highest attainable delivery of outpatient consultations per capita per year (Table 2, Column 1). On the contrary, the proportions of illiterate and economically active population
were negatively correlated with it, reflecting higher barriers to health care use in rural areas where education is lower and where a higher proportion of the population recorded as economically active (due to higher employment in agriculture rather than informal services and businesses).

When healthcare delivery was analyzed as a two-step production process at district level, we found more precise indications of how specific inputs contribute to district efficiency. Donor and district expenditures, the proportions of HFs with access to water and electricity, and the availability of staff housing, were positively correlated with the highest availability of staff and equipment that HAs could have managed to place in HFs (Table 2, Column 2). On the contrary, the number of HF in a district was negatively correlated with it for given HAs’ inputs. Interestingly, provincial expenditure has no significant effect, probably due to its reduction face of an increase in district expenditure arising from decentralization policies.

As expected, we found that higher availability of staff and equipment in HFs had a positive effect on the number of outpatients consultations per-capita per year deliverable by HFs in a given district (Table 2, Column 3). The coefficient associated with HF staff and equipment, which was hypothesized to depend on financial resources and HA efficiency, was larger in the HF production frontier, where expenditure was not included (Table 2, Column 3). It was smaller in the district production function where expenditure was included (Table 2, Column 1). The numbers of HCs and clinics in a district were positively correlated with the HFs attainable output. This suggests that the non-correlation between the total number of HFs and the district attainable output would therefore reflect the opposite effect of the number of HF on the potential service deliverable and on HA management performance. Higher proportions of illiterate and economically active population in the district were negatively
correlated with the delivery of outpatient consultations, likely reflecting utilization constraints in more rural areas.

The comparison of the efficiency scores and relative rankings of district, HA and HFs through scatter plots (Figure 3) and correlation coefficients, indicated that the district aggregated health care delivery process is closer to the HFs than to the HAs process in the two-step model. HA efficiency and ranking are only mildly correlated with district efficiency and rankings (Perason’s correlation: 0.11, p-value: 0.00 and Kendall’s rank-correlation: 0.07, p-value: 0.09), as illustrated by the dispersion of the relative plots in Figure 3.a. Similarly, HA and HF efficiency scores and their ranks are different (Figure 3.b) and not significantly correlated (Perason’s correlation: 0.01, p-value: 0.87 and Kendall’s rank-correlation: -0.05, p-value: 0.26). Aggregated district and HF efficiency scores and their ranks were highly correlated (Perason’s correlation: 0.86, p-value: 0.00 and Kendall’s rank-correlation: 0.88, p-value: 0.00) and very similar, as illustrated by their plots close to the 45 degree line (Figure 3.c).

FIGURE 3

5.2. The effect of health administration efficiency on district and health facility efficiency

Table 3 shows SFA estimates including HA efficiency for aggregated district, with and without HF staff and equipment (Columns 1 and 2), and for HFs.

TABLE 3
When included in the district production function, the HA efficiency score had a positive and significant effect on the deliverable outpatient consultations, as illustrated by the coefficient in the aggregated district production frontier (Table 3, Column 1). When comparing the aggregated district model without and with HA efficiency (Table 2, Column 1 and Table 3, Column 1) the higher log-likelihood, the lower AIC criterion and LR test suggested that accounting for administrative efficiency improved the model fit. The LR test (LR=6.61>3.84) rejected the hypothesis of the coefficient associated with the administrative efficiency being null.

After the inclusion of the HA efficiency score, the coefficient associated with district expenditure was larger and the coefficient associated with the HF staff and equipment was no longer significant. This indicate that the effect of HF staff and equipment in an aggregated district production function is fully explained by the availability of financial resources and the efficiency of HAs in using them. The LR test (LR=0.414<3.84) comparing the model specification in Table 3, Column1 and Table 3, Column2 confirmed that the inclusion of HF staff and equipment did not improve the model fit.

Once the availability of inputs for health service delivery is accounted for, HA efficiency has no significant effect on the estimated HFs’ production frontier and efficiency scores (Table 3, Column 3). This suggests that HA efficiency influences healthcare delivery only through the level of staffing and equipment in HFs. The smaller log-likelihood, the higher AIC criterion and the LR test (LR=2.11<3.84) comparing the model in Table 2, column 2 with the model in Table 3, Column 3 confirm that the inclusion of administrative efficiency does not improve the model fit to the data.
While the inclusion of HA efficiency in the district and HFs healthcare production models affected the responsiveness (elasticity) of healthcare delivery to each input, but not the measured district or HFs efficiencies. Indeed the coefficients associated with the inputs changed when the HA efficiency scores were included in the analysis, but not much the efficiency scores. Both the district and HFs efficiency scores resulting from the SFA models with and without HAs’ efficiency were very highly correlated (Pearson’s correlation: 0.99, Kendall’s rank-correlation: 0.99), indicating that the measured performance was not affected by the inclusion of HA efficiency.

5.3. Robustness checks

SFA estimates of efficiency may be sensitive to the assumptions made about the distribution of the inefficiency term and the definition of the model used (Street 2003, Kumbhakar, Lien et al. 2014). We therefore calculated the confidence intervals of the efficiency scores and performed a number of robustness checks. We carried out the same analysis first assuming an exponential distribution for the inefficiency terms $u^D_i$, $u^A_i$, and $u^F_i$, rather than half-normal, and second measuring healthcare output in service units. The latter is a composite weighted measure used in Mozambique’s NHS planning which includes inpatient days, institutional deliveries, vaccinations doses, outpatient consultations and maternal and child health consultations (MISAU 2012). Under the assumption of exponential distribution for the inefficiency term, the stochastic frontier coefficients associated with input factors and the relative efficiency scores were very similar to those in the original model. When output was measured in service units, differences in the magnitude of coefficients, and for some district characteristics in significance, were found, reflecting the different nature of the
production process. However, the significant and positive effect of administrative efficiency on the district but not on the HFs production frontier, was confirmed by both robustness checks. The correlation and rank correlation between the efficiency scores obtained from the original model and the one with exponential distribution (between 0.98 and 0.99) and the one with service units output (0.52 and 0.60) indicate sensitivity of results to the output definition rather than distributional assumptions.

Although less frequently done in the literature, we estimated alternative SFA models assuming a translog production function and using a true random effects model. The non-convergence of the maximum likelihood function, due to collinearity in the data, indicated that these models would not better fit the data.

6. Discussion

In this study we set out to assess the efficiency of health districts in delivering outpatient primary care in Mozambique and specifically of health administration and healthcare providers in performing their roles. We estimated efficiency both assuming an aggregated healthcare delivery process and a two-step process, where HAs and HFs hold the separate responsibilities of managing financial resources to purchase healthcare inputs and of using them to deliver services. We found evidence of inefficiency at the aggregated district level (average efficiency score 73%) that could be attributed to both HAs (average efficiency score 66%) and HFs (average efficiency score 74%). We found variation in performance across districts, and in the performance of the same district when evaluated at the HA, HF or aggregated district level. Performances at the HFs and aggregated district level were similar, since they were both evaluated with respect to the volume of services delivered. We found
evidence that administrative efficiency, affected HFs and district healthcare delivery through the availability of staff and equipment.

The results obtained are in line with those of the literature on sub-national health systems in LMICs settings (Kathuria and Sankar 2005, Varela, Martins et al. 2010, Kinfu 2013). Indeed, we found a great variability in input availability and in environmental factors, which affected district performance, emphasizing the importance of accounting for heterogeneity in efficiency analysis (Greene 2004). However, the comparison of the aggregated versus the two-step district healthcare production revealed that in settings where at local level financial administration and healthcare delivery functions are attributed to separate organizations, a model accounting for administrative inefficiencies would perform better. This finding would apply to hierarchically organised public health systems, such as Mozambique, but possibly also to the analysis of efficiency in other contexts (for example the allocation of inputs across specialties in a hospital) or in other public sectors where the financial management and service delivery functions are split.

Results suggest that the existing studies, which assume an aggregated production process at the subnational level, tend to capture HFs rather than HA performance, and may generate misleading conclusions about the factors that influence district performance. First of all they ignore the role of HAs, which determine the availability of inputs for health care delivery in HFs. Secondly, it may lead to biased estimates of healthcare delivery responsiveness to specific inputs. For example, where an input such as the number of facilities has opposing effects on HA and HF production frontier, these may cancel each other out leading to an insignificant coefficient on district healthcare delivery. Estimating the effect of specific
inputs correctly and precisely could avoid the formulation of ineffective policy recommendations and provide indications for more specifically targeted intervention.

Although the final estimated district efficiency in healthcare delivery is similar in the aggregated and two-steps models, the analysis of the production process and performance of distinct organizations separately provides more precise insights on where inefficiencies are. For example, we found that in the period of study only 66% of the attainable levels of staff and equipment for given financial resources were available in HFs. Although the focus of the policy debate is on the improvement of health care conditions and delivery processes at the HF level, increasing HA efficiency is crucial to promote the effective use of available financial resources at district level. Increasing HA efficiency may translate into up to an average 34% increase in HFs staff and equipment without additional recurrent expenditure. Tackling administrative efficiencies will result into increased healthcare delivery. On the contrary, increasing financial resources without addressing HA inefficiencies may not produce the expected outcome, even when HFs are efficient. On average, a reduction of administrative inefficiency by 10 percentage points, for a given budget would increase by 0.2% the volume of services delivered per thousand population per year.

The study presents limitations common to SFA models or related to data availability, particularly in LMICs, and common to existing studies of sub-national health systems efficiency. The definition of input and outputs is limited by data availability and may reduce the potential to evaluate the performance of sophisticated production processes, with multiple objectives, such as healthcare including not only outpatient but also inpatient more complex care. Furthermore, no distributional and quality concerns were explicitly included. However, the staff and equipment index first includes essential healthcare inputs, HR
availability in particular, which are likely to be correlated with other inputs (Wagenaar, Gimbel et al. 2014), and second, it incorporates distributional concerns since it accounts for resource distribution across HFs according to the minimum requirement set by norms.

In developing both the aggregated and the two-step models we follow previous studies focusing on similar settings (Kinfu 2013) and use a Cobb-Douglas production function which assume complementarities between inputs. Although the Cobb-Douglas functional form assume constant returns to scale, and the same returns to scale across districts, it is possible that returns to scale vary at different levels of inputs and across districts. In particular provincial administrations may influence returns to scale and economies of scale and in spite of national directives prices vary across districts (Arndt, Jones et al. 2015). Differences in the incentives for staff, such as the implementation of pay for performance schemes, which began to be applied in some provinces only in 2012, may also affect productivity. Such differences should be considered by further studies.

Since measures of disease prevalence were not available at district level, the coefficients associated with district characteristics should be interpreted with caution since they are potentially biased by this omission.

Although potential limitations related to the data available exist, in this study the use of routine data has allowed to advance in comparison with previous studies focusing on similar settings.

Our findings lad to a number of recommendation for further research investigating efficiency of sub-national health system with the aim of informing policies improving effectiveness in the use of public health resources. First, modelling separately the functions of HAs and HFs within districts can allow assessing the efficiency of each institutions and
understand which districts are underperforming in which functions. Second, the use of different models for HAs and HFs production process allows to identify the effect of each input within the whole district healthcare delivery process. Third, while the inclusion of HA efficiency in the district and HFs production function does not affect the performance measurement, it improves the precision of estimates of the effect of each input. Finally, research on health system performance should explicitly account for the organizational architecture of the health system and the role of health administrations. The use of contextual knowledge about the healthcare production processes to inform efficiency analysis are recommended to enhance the analysis of healthcare production processes and the understanding of which specific policy interventions can improve their efficiency.

7. Conclusion

Studying the effect of administrative efficiency on district and HFs performance has shown that analyzing each production process separately is more informative than assuming an aggregated production process at the health system sub-national level. While district performance measured under the assumption of an aggregated production process reflects HFs performance, it is driven by both the performances of HA in using financial resources to staff and equip HFs and the performance of HF in delivering healthcare.

Differences in performance across districts and HA, HFs or aggregated level within districts exist. The significant effect of administrative efficiency on healthcare delivery through the availability of staff and equipment in HFs, indicates that improving local capacity for resource administration may alone increase the volume of healthcare delivered at no
additional cost. Further research on health system efficiency should separately assess HA and HF performances and investigate their specific drivers.
References


Figure 1. District aggregated healthcare production process

Figure 2. Two-step district healthcare delivery production process in Mozambique
Figure 3. Comparison of district, health administration and health facilities efficiency scores and rankings, Mozambique (2008-2011)

a) Health Administration and District

b) Health Administration and Health Facilities

c) Health Facilities and District
### Table 1. District descriptive statistics, Mozambique 2008-2011

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient consultations per capita per year</td>
<td>1.08</td>
<td>0.46</td>
<td>0.19</td>
<td>2.50</td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF staff and equipment index</td>
<td>45.80</td>
<td>14.67</td>
<td>17.35</td>
<td>92.49</td>
</tr>
<tr>
<td>Total expenditure per capita (MZM)</td>
<td>138.26</td>
<td>101.67</td>
<td>6.84</td>
<td>646.12</td>
</tr>
<tr>
<td>Government district expenditure per capita (MZM)</td>
<td>41.74</td>
<td>54.90</td>
<td>0.00</td>
<td>339.98</td>
</tr>
<tr>
<td>Government provincial expenditure per capita (MZM)</td>
<td>20.55</td>
<td>15.67</td>
<td>2.44</td>
<td>127.83</td>
</tr>
<tr>
<td>Donor provincial expenditure per capita (MZM)</td>
<td>12.63</td>
<td>8.24</td>
<td>2.08</td>
<td>77.93</td>
</tr>
<tr>
<td>Government district expenditure per HF (1,000 MZM)</td>
<td>5.28</td>
<td>8.44</td>
<td>0.00</td>
<td>72.15</td>
</tr>
<tr>
<td>Government provincial expenditure per HF (1,000 MZM)</td>
<td>134.29</td>
<td>47.77</td>
<td>45.06</td>
<td>276.03</td>
</tr>
<tr>
<td>Donor provincial expenditure per HF (1,000 MZM)</td>
<td>85.31</td>
<td>24.54</td>
<td>44.02</td>
<td>146.32</td>
</tr>
<tr>
<td><strong>District characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District hospitals (per 100,000 inhabitants)</td>
<td>0.16</td>
<td>0.30</td>
<td>0.00</td>
<td>1.68</td>
</tr>
<tr>
<td>Health centres (per 100,000 inhabitants)</td>
<td>1.43</td>
<td>1.68</td>
<td>0.00</td>
<td>12.39</td>
</tr>
<tr>
<td>Clinics (per 100,000 inhabitants)</td>
<td>7.39</td>
<td>5.40</td>
<td>1.09</td>
<td>41.12</td>
</tr>
<tr>
<td>Total number of HF per district</td>
<td>8.70</td>
<td>3.72</td>
<td>3.00</td>
<td>20.00</td>
</tr>
<tr>
<td>HF with water (percentage)</td>
<td>42.27</td>
<td>28.78</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>HF with electricity (percentage)</td>
<td>29.09</td>
<td>27.24</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>HF housing availability (ratio actual to norms)</td>
<td>0.98</td>
<td>0.66</td>
<td>0.00</td>
<td>3.63</td>
</tr>
<tr>
<td>Economically active population (percentage)</td>
<td>72.77</td>
<td>8.59</td>
<td>38.50</td>
<td>87.80</td>
</tr>
<tr>
<td>Illiterate population (percentage)</td>
<td>55.85</td>
<td>15.26</td>
<td>14.40</td>
<td>79.80</td>
</tr>
</tbody>
</table>

N=532 (133 districts over 4 years)
<table>
<thead>
<tr>
<th>Production Process (Decision Making Unit)</th>
<th>Aggregated (District)</th>
<th>Two-step (Health Administration)</th>
<th>Two-step (Health Facilities)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF staff and equipment index</td>
<td>0.132**</td>
<td>0.201***</td>
<td></td>
</tr>
<tr>
<td>Total expenditure per capita (MZN)</td>
<td>0.271***</td>
<td>0.037***</td>
<td></td>
</tr>
<tr>
<td>Government district expenditure per HF (MZN)</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government provincial expenditure per HF (MZN)</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donor provincial expenditure per capita (MZN)</td>
<td>0.160***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Inputs**

- **District characteristics**
  - District/Rural hospitals (per 100,000 inhabitants)
  - Health centres (per 100,000 inhabitants)
  - Clinics (per 100,000 inhabitants)
  - Total number of HF
  - HF with electricity (percentage)
  - HF with water (percentage)
  - HF housing availability (ratio actual/minimum standard)
  - Illiterate population (percentage)
  - Economically active population (percentage)
  - Constant
  - Provincial dummies

**Variance Parameters**

- Lambda
- Sigma(u)

**Log likelihood**

- Mean
- Std. Dev.
- Minimum
- Max

***, **, * indicate significance at 1%, 5%, 10% level.
N=532 (133 districts over 4 years)
Standard Errors in parenthesis
All Stochastic Frontier Inputs in natural log
### Table 3. District and health facility stochastic frontiers and efficiency scores, accounting for health administration efficiency, Mozambique (2008-2011)

<table>
<thead>
<tr>
<th>Production Process (Decision Making Unit)</th>
<th>Aggregated (District)</th>
<th>Two-step (Health Facilities)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td><strong>Stochastic Frontier</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF staff and equipment index</td>
<td>0.037</td>
<td>0.251***</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>Total expenditure per capita (MZM)</td>
<td>0.308***</td>
<td>0.317***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.025)</td>
</tr>
<tr>
<td><strong>District characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District/Rural hospitals (per 100,000 inhabitants)</td>
<td>0.247***</td>
<td>0.253***</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Health centres (per 100,000 inhabitants)</td>
<td>0.013</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Clinics (per 100,000 inhabitants)</td>
<td>0.050</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>HF with electricity (percentage)</td>
<td>0.002**</td>
<td>0.003**</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>HF with water (percentage)</td>
<td>0.001**</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Illiterate population (percentage)</td>
<td>-0.016***</td>
<td>-0.016***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Economically active population (percentage)</td>
<td>-0.010**</td>
<td>-0.010**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Health Administration efficiency score</td>
<td>0.290**</td>
<td>0.336***</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.506</td>
<td>-0.441*</td>
</tr>
<tr>
<td></td>
<td>(0.309)</td>
<td>(0.265)</td>
</tr>
<tr>
<td>Provincial dummies</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Variance Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda</td>
<td>1.799***</td>
<td>1.801***</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.267)</td>
</tr>
<tr>
<td>Sigma(u)</td>
<td>0.333***</td>
<td>0.333***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.036)</td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>34.068</td>
<td>33.861</td>
</tr>
<tr>
<td>AIC</td>
<td>-24.100</td>
<td>-25.700</td>
</tr>
<tr>
<td><strong>Efficiency scores</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.731</td>
<td>0.730</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.179</td>
<td>0.179</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.234</td>
<td>0.238</td>
</tr>
<tr>
<td>Max</td>
<td>0.967</td>
<td>0.967</td>
</tr>
</tbody>
</table>

***, **, * indicates significance at 1%, 5%, 10% level.
N=532 (133 districts over 4 years)
Standard Errors in parenthesis
All Stochastic Frontier Inputs in natural log