MONEY WELL SPENT? AN EMPIRICAL EXAMINATION OF GOVERNMENT EXPENDITURES FOR HIV/AIDS TREATMENT IN SOUTH AFRICA

Dan Friesner, Gonzaga University
Sudesh Mujumdar, University of Southern Indiana
Peter Cashel-Cordo, University of Southern Indiana

ABSTRACT

As governments around the world take the threat of the HIV/AIDS pandemic more seriously, they are allocating more funds to HIV-related prevention and treatment programs. A crucial question is whether these funds are being allocated in an optimal manner. This paper provides an empirical methodology that allows policy makers to determine whether these funds are being allocated in a way that results in the maximum possible reduction in the state of the pandemic. Specifically, we use data envelopment analysis (DEA), a tool commonly used in management science to measure efficiency, as a means of benchmarking the maximum amount of "pandemic reduction" that can accrue from a specific level of funding and a particular economic and epidemiological state. As a case study, we apply our methodology to a panel of 9 South African provinces. The results suggest that most of these countries do not allocate their funds in an efficient manner.

INTRODUCTION

As governments around the world take the threat of the HIV/AIDS pandemic more seriously, they are allocating more funds to HIV-related prevention and treatment programs. For example, in South Africa, where prevalence rates are in excess of twenty percent, the government spent over $28 million US in 2002/2003 and over $45 million US in 2003/2004 to finance HIV-related programs (United Nations 2002; Martin 2003). This trend is not limited to South Africa. The Global Fund, for example, his dispersed nearly $496 million (US) to Sub-Saharan African countries, $159 million to East Asian countries and $119 million to Latin American Countries.[2]

A crucial issue is how to allocate these funds in a manner that causes the highest possible reduction in the spread of the disease as well as the number of deaths from those who have already contracted the disease. There are two commonly used tools that allow policy makers to predict how best to allocate HIV/AIDS funding. The first is the Activity-Based Cost (ABC) Model, which applies concepts used in the accounting discipline as a template for funds allocation (Saxenian and Schwab 2003). The major attribute of this model is that it is a very flexible and requires relatively little data. It is also relatively easy for policy makers to understand and apply. A drawback, however, is that it usually takes as given epidemiological and economic conditions. As such, if those conditions change, the projections produced under the ABC model may become suspect.

A second model used to forecast optimal funding patterns is the Goals Model produced by The Futures Group International (http://www.futuresgroup.com/WhatWeDo.cfm?page=Software&ID=GOALS). Unlike the ABC Model, the GOALS model applies regression and simulation analyses to current and projected economic, demographic, epidemiological and funding data to forecast the spread of the disease, as well as the impact of changes in funding on the spread of the disease. Moreover, policy makers have the ability to adjust the model to determine how

[2] http://www.theglobalfund.org/en/funds_raised/commitments/

sensitive the results are to the changes in these factors. While this model provides very accurate and precise predictions about the optimal allocation of funding, it is extremely complicated and thus is not generally useable for policy makers unless they receive assistance from epidemiologists or bio-statisticians familiar with the model. Additionally, this approach requires either a substantial amount of data or a large number of assumptions to be made about current socio-epidemiological-economic conditions.

An additional drawback to both models is that they only provide information for half of a policy-oriented process. That is, for a policy to be successful (or optimal), the initial decision must not only be based on accurate and precise predictions, but that policy must also be subsequently *evaluated* and revised if need be. With regard to HIV/AIDS funding policies, this implies that not only should current funding decisions be based on what researchers and policy makers predict to be the optimal distribution of funding, but those decisions must be evaluated ex post to determine whether those decisions did, indeed, turn out to be optimal. If they do not turn out to be optimal, it is necessary to determine what factors were responsible for the sub-optimal performance. To our knowledge, no set of quantitative tools has been posited as an effective ex post means of determining whether HIV/AIDS funding policies have been allocated optimally.

The purpose of this paper is to posit a third methodology for evaluating the optimality of HIV/AIDS funding policies. We utilize data envelopment analysis (DEA), a tool commonly used in management science to measure efficiency, as a means of benchmarking the maximum amount of "pandemic reduction" that can accrue from a specific level of funding and a particular economic and epidemiological state. Our approach provides several contributions to the literature. First, it strikes a balance between the less data intensive, less precise predictions of the ABC Model and the data-intensive, more complicated GOALS model. Second, not only can it be used to predict how funds should be allocated, but it can also be used to evaluate (ex post) whether those funds (even if allocated according to the recommendations of a model) were, in fact, allocated optimally. That is, it also provides an independent, ex-post means of evaluating the accuracy and precision of the ABC or GOALS models. Lastly, our approach can be used to make inter-country comparisons about the efficiency of HIV/AIDS funding. While the ABC and GOALS models make predictions for a particular country's funding needs, organizations such as UNAIDS or the Global Fund that distribute grants and loans to fight the spread of AIDS need some method for determining the optimal allocation of funds not only within a particular country, but also how money should be allocated between countries.

The remainder of this paper proceeds in three steps. First, we outline our methodology for evaluating the optimal allocation of HIV/AIDS funding. In doing so we also address the assumptions underlying our methodology (including data requirements) as well as the methodology's limitations. Next, we present a case study where we apply the methodology to a panel of South African provinces to determine whether or not the funds currently allocated were (or those funds expected to be allocated are) spent optimally. We conclude our paper by discussing some policy implications from our work and also present some suggestions for future research in this area.

USING DEA TO BENCHMARK THE OPTIMAL ALLOCATION OF HIV/AIDS FUNDING

Our methodology operates under a number of assumptions, the most general of which is that curbing the spread of HIV/AIDS can be characterized as a production process. The output of the production process is the reduction in the scope and impact of the pandemic (i.e., the number of lives and/or

quality of lives that are saved by implementing policies to fight the spread of the disease). The inputs of the production process are those factors describing the current state of the pandemic (inclusive of any historical determinants of the current state), as well as other factors that are hypothesized to increase output. Funding used to fight the spread of HIV/AIDS (whether directly or indirectly allocated to such a purpose), then, is an input in this production process.[1] We make no assumptions about the nature of the production process (i.e., whether there are increasing, constant or decreasing returns to scale) or the number of inputs and outputs employed. The one assumption we do make is that all outputs and inputs are measured in non-negative-valued units, with larger numbers indicating more outputs or inputs. Additionally, if the decision maker is allocating resources in an optimal manner, more inputs utilized should result in at least as much output being produced.

We also make a number of empirical assumptions.[2] First, we assume that each of the inputs and outputs can be measured appropriately with available data, and that those data do not contain significant measurement error. In the case that our data are used to compare the optimal distribution of funding across countries, we also assume that those countries are comparable. That is, each country included in the analysis faces a similar set of technological and resource constraints when attempting to reduce the spread of the disease. Depending on the countries being compared, this may also imply that these nations have comparable political, economic and social structures, as significant differences in these features across countries may imply that a set amount of HIV/AIDS funding will have vastly different impacts in each of those countries, *even if those funds are allocated optimally*.

Given these assumptions, the most parsimonious way to characterize a production process is by depicting it as a production possibilities frontier (PPF). The PPF is a graph that shows the maximum (or efficient) amount of output (or combination of outputs) an entity is capable of producing with a fixed amount of resources and technology. This concept can be explained using a simple example, which is illustrated in Figure 1. Suppose that we have a production process that uses a fixed vector of inputs to produce two outputs: reductions in prevalence and reductions in the number of AIDS-related deaths. The process is said to be efficient if it produces a combination of outputs that are located on the PPF. Alternatively, a combination of outputs relating to a point below the PPF implies that the production process is *inefficient*, or not obtaining the maximum amount of output for the resourced and technology employed. A combination of outputs beyond the PPF is generally impossible, since that would require more than the given amount of resources and/or a higher level of technology than what is currently available to the production process.

Suppose that the production process depicted in Figure 1 is operating inefficiently, for example, at point X. The amount of inefficiency can be measured (in relative terms) as the radial distance between the origin and the frontier. Graphically, this distance is given by the ratio 0X/0Y. The numerator of this expression gives the linear distance between the origin and the inefficient point, while the denominator represents the distance between the origin and the projected efficient point (where the process should be producing). By definition, this measure is bounded between zero and one, with values close to zero indicating relatively little efficiency, and values close to one indicating that the process is nearly

[1] In the ABC Model, there is an intermediate step in the production process; namely, that inputs are used to produce "activities", which in turn produce outputs. One could easily adapt our methodology to take these "intermediate" outputs into account. For example, one could postulate a pair of production processes. The first uses inputs to produce "activities". These activities, then form the inputs for a second process which produces reductions in the pandemic.

[2] The effects of violating these assumptions on the results of the analysis will be discussed when addressing the limitations of our approach.

perfectly efficient. Scores between zero and one consequently give the proportion of efficiency obtained by the process' current resource allocation.

We can also define efficiency (as well as potential gains that can accrue from a more efficient allocation of resources) in absolute terms. If the process is producing at point X, it can produce p more units of prevalence reduction and d more units of lives saved by reallocating its resources in a more optimal fashion. Values p and d are generally defined as *potential improvement scores*.

In this paper, we construct the PPF and measure efficiency using data envelopment analysis (DEA).[3] DEA is a linear programming technique that uses data on inputs and outputs over a sample of decision makers and/or time to construct the PPF in a *relative* fashion. That is, DEA identifies the most efficient decision makers (DM) in the dataset and creates the PPF using the outputs (and inputs) of these decision makers. All other firms are benchmarked against this PPF, and the amount of efficiency or inefficiency is determined accordingly.[4] Thus, if a decision maker's efficiency score is 0.65, then that DM is only 65% as efficient as the *most efficient* DM in the data set. Potential improvement scores must also be interpreted in a similar fashion.

Using DEA to create the PPF inherently endows the results with a series of advantages over alternative methods of efficiency measurement. The primary advantage of DEA is that it is nonparametric, and does not make any a priori assumptions about the nature of the production process, or the shape of the PPF. Instead trends in the data itself are used to create the PPF. Additionally, unlike other methods of modeling production processes (such as regression analysis), DEA does not require inputs to be exogenous determinants of outputs. Thus, DEA does not force the researcher to address issues of stationarity or simultaneity bias. Moreover, unlike regression analysis, DEA also allows the researcher to specify multiple outputs. Lastly, if one is using the results of DEA to address policy issues, particularly issues related to the optimal allocation of a fixed amount of resources, then one must compare alternatives in a relative fashion. Since DEA's efficiency scores are constructed relative to the most efficient decision maker(s) in the data set, its findings are consistent with the needs of policy makers.

Using DEA to measure efficiency also has several limitations. Because DEA computes efficiency scores in a relative fashion (and because those scores are bounded between zero and one), the results are not likely to be normally distributed. Consequently, if one is interested in using the efficiency scores to conduct hypothesis tests, then it is necessary to use nonparametric (rank-order) methods in place of the more traditional parametric hypothesis tests. Additionally, as Simar and Wilson (2003) note, one should not use DEA efficiency scores as the dependent variable in a regression analysis since any results obtained are likely to be highly inefficient, and possibly inconsistent.

A second limitation of DEA is that its efficiency scores may change as firms are added to or deleted from the data set. This is especially true if the added or deleted decision makers are producing at a point on the PPF, because adding or deleting these firms will change the shape of the PPF, and thus all efficiency scores computed relative to the frontier. Thus, the researcher must ensure that the collection of decision makers utilized are appropriate for analysis.

Third, one must ensure that the inputs and outputs employed in the analysis are reliable and representative of the production process being analyzed. If the data contain significant measurement error, if important

[3] We utilize the Data Envelopment Analysis Program (DEAP) version 2.1 to run the analysis. This package is written by Tim Coelli, and can be downloaded free of charge (along with a detailed instruction manual) on the world wide web at http://www.uq.edu.au/economics/cepa.

[4] A more technical discussion of DEA can be found in the Appendix of this paper.

inputs and/or outputs are missing, or if irrelevant inputs and/or outputs are included in the analysis, the results calculated by DEA may be incorrect. As a result, any results must be carefully interpreted in the context (or conditional upon) the data used to generate those findings.[5]

AN EMPIRICAL APPLICATION

As an empirical illustration, we apply our methodology to a panel of 9 South African provinces. South Africa provides an interesting case analysis for several reasons. First, South Africa's prevalence rates are among the highest in the world. Thus, it is imperative that what fixed resources are available to combat the pandemic are allocated as efficiently as possible. Second, since the late 1990's the national government has acknowledged the severity of its pandemic, and has allocated a significant portion of its annual budget towards HIV/AIDS programs.[8] Third, sufficient data resources exist (and are freely available on the world wide web) to run a reasonably complete analysis.

It is also important to note that, while our data represent a timely and relevant application of our methodology, *we do not intend our case study to be a complete discussion and/or critique of the efficiency of South African HIV/AIDS funding policies.* Instead, our goal is twofold. First, we intend the analysis to be an illustrative example of our methodology. Perhaps more importantly, we interpret our findings only as a first step in this investigation. Our goal is merely to call attention to a potential problem, and to encourage future work in this area (using more detailed and exhaustive data) to perform a more complete analysis of the efficiency of South African HIV funding.

As mentioned earlier, our methodology can be used as both a method of predicting the optimal amount of pandemic reduction with a fixed amount of funding as well as an ex post method of evaluating the efficiency of past funding levels. In this application, we take the former perspective. Our particular objective is to predict, based on current epidemiological and economic estimates, the optimal amount of pandemic reduction that should occur based on an efficient use of *current budget projections on HIV/AIDS provincial allocations.* The time frame of our analysis is a moving four-to-five-year window. That is, we examine how allocated funds (or commitments for fund allocations) for the 9 South African provinces from 2002/2003-2005/2006 impact pandemic estimates 4/5 years later (2007-2010). Consequently, when conducting our analysis we implicitly assume that all estimates and projections are accurate and precise. Additionally, we assume that all funds allocated to the provinces are spent within that year. Failure to meet these assumptions may limit the generality of our findings.

Our data for the analysis come from several sources. Projections on the extent of the pandemic come from the Actuarial Society of South Africa's AIDS Model (http://www.assa.org.za/aidsmodel.asp). In particular, we define our baseline measures of the pandemic as the number of HIV-infected individuals and the number of AIDS-related deaths in each province and year, each expressed as a proportion of the total provincial population. That is, we utilize a general prevalence rate and a general death rate for each province and year. The ASSA's Model provides estimates and projections for these variables (as well as supporting information such as each province's total population) from

[5] This last limitation is also true of most other alternatives to DEA, including regression analysis.

[8] In fact, so much money was allocated during the first few years of this campaign that the provincial governments had difficulty spending all of the allocated funds. However, as local health departments have increased their ability to react to the pandemic, this problem was reduced significantly (Hickey wt al 2003). The starting date for analysis represents a point in time where i) the vast majority of funds allocated in a specific year are spent during that time frame and ii) previous funding for HIV/ AIDS has been relatively sparse, so that we can reasonably abstract from the influence of past funding on current and future epidemiological conditions.

1986 to 2010. Budget-related information comes from Hickey et al (2003). Lastly, as a control for current socio-economic conditions, we employ Human Development Index (HDI) projections created by the United Nations Development Programme in South Africa (http://www.undp.org.za/sahdr2000/hdr.appendices.pdf).[9]

Consistent with our methodology, we define pandemic reduction in South Africa as a production process. Of particular interest is how we define the inputs and outputs of that process, given the fact that we have a small panel at our disposal (36 observations). We specify four inputs and two outputs. The first input is the projected HDI for each province and time period. Our second input is the (projected) percent of each province's budget that is spent on health-related programs. The third input is the ratio of (projected) provincial HIV spending per capita to the minimum value for each of the 9 provinces in that year. Lastly, we compute the ratio of projected, "discretionary" (or non-conditional grant) HIV spending to conditional grant HIV spending for each province and year. The first input is intended to capture the current state of welfare within each province and year. The second measure attempts to control for the indirect expenses of HIV, such as longer hospital stays, that occur in each province and year due to increases in the pandemic. The final two measures control for direct HIV/AIDS expenditures. The third is a measure of the relative monetary commitment of each province to fighting the pandemic, while the latter measures the flexibility each province has to allocate funds. Specifically, each province receives conditional grants from the national government, which are intended for particular purposes and programs. However, provinces also obtain other funds (which may come from the national government or through each province's tax revenues), a significant portion of which can be allocated as the provincial governments see fit. Our rationale is that a higher degree of flexibility allows each province to tailor spending in a manner that best fits their needs, and thus, should be a more efficient allocation of those funds.

Defining the outputs is slightly more problematic. As stated earlier, our baseline measures of the pandemic are the projected prevalence rate (or proportion of the population that is HIV-infected) and the projected AIDS death rate (or the proportion of the population that dies from AIDS-related illnesses in a given time period). One potential problem is that higher prevalence and death rates imply a worsening pandemic; however, to be consistent with a production process, higher positive values should indicate a declining pandemic. To address this issue, we transform our baseline measures by examining the *inverse* of the projected prevalence rate and the *inverse* of the projected death rate.

Another issue related to defining outputs is that of path dependency. Path dependency implies that provinces that have already taken significant steps to reduce the pandemic receive two potential advantages over provinces that have not taken such steps. First, reductions in the pandemic today increase the likelihood that the pandemic will decline (or at least remain constant) in the future. Additionally, lower states of pandemic may increase the effectiveness of funding allocated to reduce the state of the disease in a more than proportionate fashion. In either case, the production process should account for current and past epidemiological conditions. One possible approach is to include this information

[9] This is a commonly used measure in the field of international and development economics. The index essentially takes data on a number of variables commonly believed to influence a country's welfare and standard of living, including GDP, education levels, women's rights, etc, and aggregates this information into a single index. We employ this overall index as opposed its disaggregated components because we have a relatively small panel (9 provinces over 4 years) and using so many variables as inputs may simply be asking too much of the data.

as inputs in the production process.[10] Another possibility (which we use) is to measure our outputs relative to current and past conditions. That is, we re-define our outputs to measure the projected state of the pandemic relative to its recent history. While there is no universally accepted measure that takes all of these factors (time dependency and the nature of outputs required for the analysis) into account, we propose the following measure, which is parsimonious, yet familiar to most practicing public health researchers. It is essentially a variant of the chi-square statistic:

$$output_{jit} = \frac{\left(p_{jit} - e_{jit}\right)^2}{e_{jit}}$$

where:
j indexes each output;
i indexes each province;
t indexes time;
p describes the projected state of the output 5 years from time t;
e describes the expected value for p based on the history of those projections.

In our case, we measure e as a 5 year average of past values. That is, if p denotes the inverse prevalence rate for a province in 2009, then e denotes the average inverse prevalence rate for that same province from 2004-2008. In South Africa, prevalence and death rates are expected to increase, but at a declining rate over the next decade. That is, the impact of the disease is expected to level off. As a result, our inverse rates should be start off as very high values, but decline at a declining rate. What our output measure captures are significant deviations from that trend. Ideally, if funding is effective, we should see greater deviations from the trend (in a positive fashion if the prevalence or death rate drops) in our new output measure. However, if funding is ineffective, we should see only minor deviations from the current epidemiological trend, and the output measure should be close to zero.

This definition of an output, if employed in our production process, has a specific interpretation: it essentially looks at efficient allocation of funds on the state of the pandemic at a specific time period in the future. That is, this measure assumes that funding allocated today will have an impact on the outcome 5 years from the date those funds were allocated. However, it may also be the case that funding allocated today could have an effect on the disease a year from now, two years from now, etc. To measure this "cumulative" effect, we need only aggregate our previous measure of output:

$$cumulative_{jit} = \sum_{l=1}^{5} output_{ji,t-l}$$

That is, our cumulative measure entails summing all output measures from the time the funding was allocated to the beginning of the projection period. In this case, that means summing the outputs over the five years between the allocation of funding and the evaluation period.[11]

A potential drawback to our measure is that it weights positive and negative deviations from the epidemiological trend in the same fashion. Thus, for this measure to be appropriate, policy makers must (at least partially) be interested in epidemiological stability as opposed to dramatic swings in prevalence and death rates. This approach is also appropriate if the trends in both prevalence

[10] The small size of our data set prohibits us from taking this approach. Additionally, one may want to avoid this approach if using estimates and/or projections for epidemiological conditions. The reason is that, when using projections, past projections/estimates are used to predict future projections. As such, by including current and past values as inputs, one may be automatically biasing the results towards efficiency.

[11] Note that there are many ways to adapt this calculation to the particular needs of the research at hand. One could, for example, extend the time frame of the analysis, so that it was longer than 5 years. One could also include the evaluation period in this calculation.

and death rates are consistently increasing (or the trends in inverse prevalence and death rates are decreasing) over time for all provinces in the data set (which is the case in our data). We also note in passing that it is a relatively simple exercise to adapt this approach to other epidemiological states where these trends may not be occurring.

EMPIRICAL RESULTS

Table 1 contains some descriptive statistics for the variables used in our analysis. Examining the inputs, we see that the average HDI score is approximately 0.57 (both at the mean and the median), which is relatively low in magnitude. The standard deviation is also quite small, at 0.06. These statistics imply that not only are socio-economic conditions low, but they are also consistent over provinces and time. The statistics also show that provinces over time allocate approximately 20 percent of their budgets to health related programs. Again, the mean, median and standard deviations imply that this finding is relatively consistent over provinces and time. Perhaps more interestingly, there is a considerable amount of variation in both HIV spending per capita across provinces as well as in the ratio of discretionary to conditional grant funding. This implies two, related things. First, some provinces are making a more concerted effort to address the pandemic than others. Secondly, some provinces are also able to exercise much more flexibility in allocating HIV/AIDS funds than others, since they have access to a larger portion of discretionary funds relative to conditional grants.

The second portion of this table contains descriptive statistics used to create the outputs for the analysis. If we combine these variables, we see that the mean prevalence rate is approximately 15 percent – a relatively high figure. However, the standard deviations, minimum and maximum values for these variables also indicate that there is a substantial amount of variation over provinces and time. The death rate exhibits a similar pattern, with mean and median values exhibiting much smaller magnitudes, at approximately 1 percent, respectively.

Perhaps more importantly, the descriptive statistics in Table 1 provide some initial results from our benchmarking analysis. Examining the non-cumulative output efficiency scores, we see that the average province was about 41.5 percent efficient in allocating funds. Moreover, had those funds been allocated optimally, the prevalence rate could have been reduced 13.5 percent and over 40,000 AIDS-related deaths could have been saved or postponed.[12] Using the cumulative measure of output, we see that the mean efficiency of fund allocation is only about 36 percent. On the surface, this implies that, in the long run, funds are allocated in a less efficient manner. However, if we compare the median values, we see very similar efficiency measures across both methods. Thus, we can tentatively conclude that both output measures provide relatively consistent results.

Because we found inefficiency in funds allocation, we also looked further to determine whether this inefficiency varies i) over time and ii) across provinces. Table 2 presents efficiency results broken down by province and year for each type of output (cumulative and non-cumulative).[13] Examination of the table

[12] Two comments are in order here. First, because we have normalized our data, the potential improvement scores are stated in terms of the transformed (chi-square-based) output. So to state potential improvement in terms of prevalence and death rates, we simply work the transformation process in reverse. Second, the results in Table 1 do not contain potential improvement scores for the cumulative output because the cumulative nature of these outputs make them difficult to interpret. For example, it would not be unreasonable to expect cumulative potential reductions in the prevalence rate to exceed 20 percent, even when prevalence rates are 25 percent, because potential reductions in one year would be added to the next, etc.

[13] As mentioned earlier, a more technical manner of proceeding is to use nonparametric statistical tests, or perhaps Spearman (nonparametric) correlation

provides a number of interesting findings. First, the consistency across efficiency scores between the cumulative and non-cumulative output specifications is, in general, retained in the disaggregated results. Additionally we see a definite correspondence in the pattern between a province's efficiency and its potential prevalence and AIDS death reductions. Given the nature by which these outputs and their potential improvement scores were calculated, these patterns are not surprising. As such, when identifying which provinces constitute a benchmark for the others, we will limit our discussion solely to the efficiency scores.

Another interesting finding is that the efficiency scores are relatively stable across time. Limpopo has the highest efficiency scores, with perfect (relative) efficiency in 3 of the 4 years. The Western and Northern Cape provinces each have high efficiency scores initially, but the scores drop dramatically over the course of time. The Gauteng and Free State provinces have the most inefficient allocation of funds over time. Kwa-Zulu-Natal's efficiency is particularly intriguing. If one looks at the non-cumulative output, it's efficiency scores increase slightly over time. However, using the cumulative output the opposite pattern emerges. Thus, while not the pattern overall, the time frame of the analysis does appear to have a minor role in defining efficiency.

Having found a clear benchmark for allocating funds, it is necessary to determine what factors make Limpopo efficient, as well as the factors that make provinces such as Gauteng very inefficient. To that end, Table 3, also breaks down several inputs and outputs by province and time. First, we note that, between 2007-2010, Limpopo is not expected to have the lowest prevalence and death rates in South Africa. Instead, the Northern and Western Capes have this honor. Additionally, Limpopo's prevalence and death rates are expected to rise slightly. Conversely, Gauteng does not have the highest prevalence and death rates, nor does its prevalence rate increase over time. Thus, Limpopo's efficiency and Gauteng's inefficiency are not solely due to where these provinces are on their epidemiological time paths.

What does distinguish these two provinces, in particular, is the total amount of funds used to fight the disease, as well as the mix of conditional grant and discretionary funds. Limpopo has the lowest percent of its provincial budget allocated to health, as well as the lowest per capita HIV spending levels. Its mix of conditional grant to other funding is close to the median for all provinces combined. Gauteng, on the other hand, has some of the highest spending levels in all three categories. Not surprisingly, Limpopo simply appears to be doing more with less – the classic notion of efficient production.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The purpose of this paper was to show how DEA, a tool commonly used in management science to measure efficiency, can be applied in a straightforward and practical manner to help policy-makers in the fight against HIV/AIDS. As an initial example, we applied DEA to a panel of South African provinces to determine which, if any, of these local governments were allocating funds in a manner that maximized reductions in prevalence and death rates 5 years after those funds were allocated. Our findings identify the Limpopo province as the clear benchmark, and Gauteng as the province most in need of improvement.

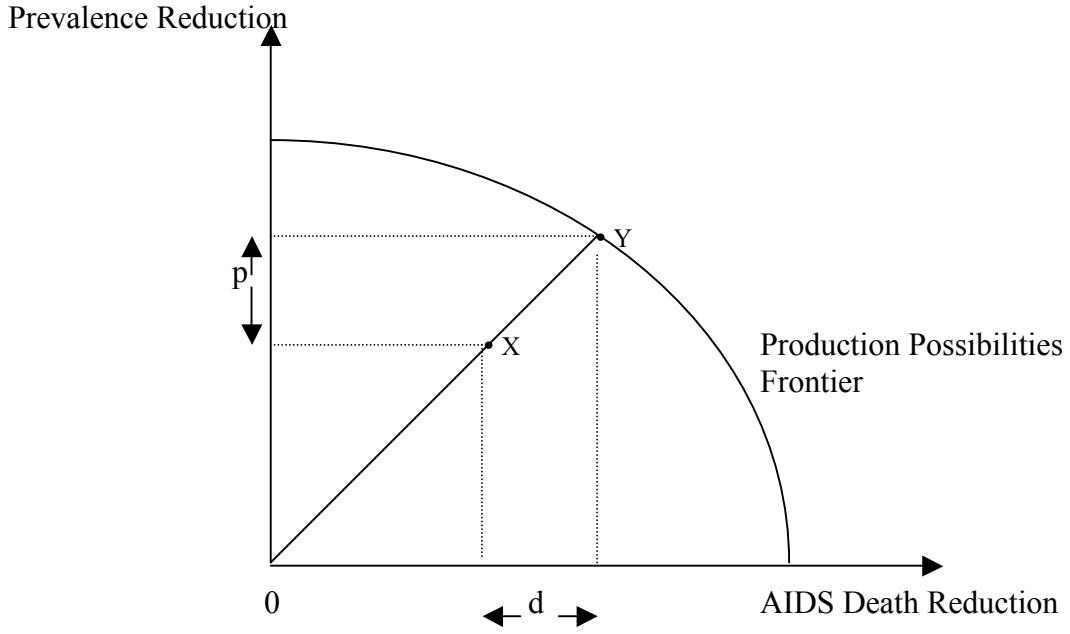
Having determined how provinces rank in their ability to allocate HIV/AIDS funding, the question arises as to what the results actually *mean*. That is, what factors, other than those variables specified in the analysis, actually

analysis. We employ a simple, non-statistical comparison of median values on the grounds that it tells much the same story. Additionally, because our data set can be considered as a population in and of itself, the use of hypothesis testing is inappropriate.

contribute to Limpopo being the benchmark and Gauteng being the most inefficient government? Alternatively, how can policy makers actually make changes in a government's allocation process to increase its efficiency?

Unfortunately, the methodology and results presented here do not answer these questions. Instead, it is the role of both policy makers and future researchers to address these issues. Usually, however, the answers can be found by comparing features in the benchmark and the inefficient province that *do not* show up in this analysis. Factors such as cultural differences, whether a province is urban or rural, whether it is geographically isolated, the amount of government bureaucracy and a host of other socio-economic factors influence the allocation of HIV/AIDS funding. Choosing a different time frame for the analysis, or defining the outputs of disease reduction differently may also change the findings presented in this study. In any case, the contribution of this methodology is that it provides a straightforward, unbiased approach to begin this discussion, which hopefully leads to a reduction in the pandemic and an increase in the lives and quality of lives for individuals in that society.

Figure 1: A Graphical Explanation of Efficiency



Appendix 1: Using DEA to Measure Efficiency[14]

Efficiency and potential improvement scores are usually calculated with a firm or an individual as the decision-making unit. In output-oriented DEA, the efficient frontier is created using a linear program that maximizes the inverse efficiency score, with weighted outputs and weighted inputs set to unity. The program also includes a constraint that allows for variable returns to scale. Efficiency scores and target outputs (which can be used to identify the potential improvement scores) are subsequently calculated by comparing each firm's observed level of output to the efficient frontier.

Formulating of the output-oriented, linear programming model is relatively simple. Let y_i be a vector of m outputs and x_i a vector of k inputs for the ith individual. If we have data over n individuals, then X is a kxn matrix of input data for all individuals and Y is a mxn matrix of output data. Then the *envelope*, or efficiency frontier, is produced by solving the following constrained, linear optimization problem:

$$\max_{\phi,\lambda} \quad \phi_i$$
$$\text{subject to} \quad -\phi_i y_i + Y\lambda \geq 0$$
$$x_i - X\lambda \geq 0$$
$$\Omega'\lambda = 1$$
$$\lambda \geq 0$$

where Ω is an nx1 vector of ones, λ is a nx1 vector of constants and ϕ_i is a scalar. The value $1/\phi_i \leq 1$ is the (technical) efficiency score for the ith individual, taking a value of 1 if the individual is on the frontier, and thus efficient. The problem is solved once for each observation in the sample, giving efficiency scores for each.

Identifying the potential improvement scores can be slightly more problematic. If the sample of data is well behaved, or if the data set is very large (i.e., as the sample size approaches infinity), potential improvement scores can be calculated residually as $\left(\frac{1}{E} - 1\right)$*the observed level of output, where E is the efficiency score for the decision maker. However, for small or ill-behaved data sets, the empirical production possibilities frontier may not approximate a smooth, well behaved curve (whether linear, bowed in or bowed out).[15] In this case, the aforementioned equation will not accurately measure the potential improvement scores.

Empirical irregularities in constructing a PPF are commonly accounted for by introducing the concept of an *output slack*, which is essentially a residual parameter(s) that determines whether the entity in question is operating on this irregular part of the curve, and how much (in)efficiency results. The output slack values are subsequently subtracted from the potential improvement scores to adjust for the possibility of irregularities in the PPF. In its simplest case, these output slacks are calculated as $-y_i + Y\lambda$. However, several DEA software packages have routines that improve upon this approach. One such improvement (which we utilize in this paper) is the two-stage method outlined in Coelli (1996). However, since the details involved in this computation are quite extensive, we refer the reader to Coelli (1996) for a discussion of this issue.

[14] This section borrows heavily from Coelli (1996, 1997) and Rosenman and Friesner (2004).

[15] The most common irregularity is that some portion of the empirical PPF runs parallel to one or more of the graph's axes.

Table 1: Descriptive Statistics

Variable	Mean	Std Deviation	Minimum	Median	Maximum
Inputs					
Per Capital HIV Spending Relative to Minimum Province	1.773	0.770	1.000	1.481	3.767
Ratio of Discretionary to Conditional Grant Spending	0.691	0.768	0.000	0.423	2.722
Percent of Provincial Budget Spent on Health Programs	21.371	4.463	16.200	19.325	31.330
HDI	0.577	0.064	0.463	0.571	0.697
Output Information					
Total Population 5 Years from Funding Date	5357041.276	2803596.699	1060967.558	4856629.002	9673612.620
Total HIV Infections 5 Years from Funding Date	805576.055	516520.535	111131.529	644518.312	1788586.653
AIDS Deaths 5 Years from Funding Date	80043.657	57582.461	7919.737	61822.162	194689.526
Prevalence Rate 5 Years from Funding Date	0.145	0.039	0.056	0.158	0.185
Death Rate 5 Years from Funding Date	0.007	0.004	0.001	0.007	0.015
Non-Cumulative Outputs and Efficiency Scores					
Prevalence Output 5 Years from Funding Date	0.039	0.063	0.000	0.018	0.311
Death Rate Output 5 Years from Funding Date	18.695	22.914	0.569	10.697	112.552
Efficiency Score	0.415	0.321	0.086	0.275	1.000
Potenital Gain in Prevalence	0.135	0.038	0.050	0.145	0.180
Potential Deaths Saved	40479.026	34289.304	3891.202	33880.645	157854.335
Cumulative Outputs and Efficiency Scores					
Prevalence Output 5 Years from Funding Date	2.790	4.487	0.076	1.066	23.000
Death Rate Output 5 Years from Funding Date	581.532	672.347	63.296	344.668	3372.868
Efficiency Score	0.364	0.326	0.044	0.219	1.000
Number of Observations	36.0				

Table 2a: Efficiency Scores for the Non-Cumulative Output

	2002	2003	2004	2005
Eastern Cape	1	0.576	0.29	0.182
Free State	0.331	0.17	0.092	0.211
Gauteng	0.185	0.105	0.121	0.26
Kwa-Zulu-Natal	0.197	0.138	0.316	0.516
Limpopo	1	1	0.85	1
Mpumalanga	0.194	0.112	0.086	0.199
Northern Cape	1	0.612	0.366	0.207
North West	0.363	0.212	0.152	0.219
Western Cape	1	0.751	0.539	0.384

Table 2b: Efficiency Scores for the Cumulative Output

	2002	2003	2004	2005
Eastern Cape	0.779	0.42	0.272	0.181
Free State	0.333	0.178	0.111	0.073
Gauteng	0.211	0.124	0.074	0.044
Kwa-Zulu-Natal	0.221	0.137	0.085	0.052
Limpopo	1	1	0.831	1
Mpumalanga	0.193	0.124	0.089	0.056
Northern Cape	1	0.585	0.353	0.217
North West	0.363	0.222	0.173	0.085
Western Cape	1	0.678	0.467	0.362

Table 2c: Potential Prevalence Reduction in 5 Years (Non-Cumulative Output)

	2002	2003	2004	2005
Eastern Cape	0.131	0.128	0.132	0.139
Free State	0.172	0.180	0.163	0.161
Gauteng	0.167	0.156	0.144	0.144
Kwa-Zulu-Natal	0.180	0.166	0.167	0.167
Limpopo	0.124	0.129	0.131	0.135
Mpumalanga	0.169	0.169	0.156	0.157
Northern Cape	0.091	0.091	0.095	0.101
North West	0.156	0.164	0.152	0.147
Western Cape	0.050	0.051	0.053	0.056

Table 2d: Potential AIDS Death Reduction in 5 Years (Non-Cumulative Output)

	2002	2003	2004	2005
Eastern Cape	44007	41104	43691	49405
Free State	18131	19918	22322	33728
Gauteng	45838	52358	71988	106319
Kwa-Zulu-Natal	70090	83860	124619	157854
Limpopo	34033	42175	44801	58612
Mpumalanga	20746	23434	28468	40904
Northern Cape	4058	3891	4372	4897
North West	21955	24996	29949	40897
Western Cape	9678	9874	11283	12991

Table 3a: Prevalence Rates in 5 Years[16]

	2002	2003	2004	2005
Eastern Cape	0.152	0.155	0.156	0.156
Free State	0.185	0.182	0.177	0.171
Gauteng	0.171	0.167	0.160	0.153
Kwa-Zulu-Natal	0.185	0.179	0.173	0.167
Limpopo	0.136	0.136	0.135	0.134
Mpumalanga	0.177	0.174	0.170	0.166
Northern Cape	0.105	0.106	0.106	0.105
North West	0.168	0.165	0.160	0.155
Western Cape	0.057	0.057	0.057	0.056

Table 3b: Death Rates in 5 Years

	2002	2003	2004	2005
Eastern Cape	0.004	0.005	0.006	0.007
Free State	0.007	0.009	0.011	0.013
Gauteng	0.006	0.008	0.010	0.013
Kwa-Zulu-Natal	0.009	0.011	0.013	0.015
Limpopo	0.004	0.005	0.006	0.007
Mpumalanga	0.008	0.010	0.012	0.014
Northern Cape	0.002	0.003	0.004	0.005
North West	0.006	0.007	0.009	0.011
Western Cape	0.001	0.002	0.002	0.003

Table 3c: Per Capita HIV Spending Relative to Other Provinces

	2002	2003	2004	2005
Eastern Cape	1.618	1.300	1.507	1.560
Free State	1.321	1.577	1.808	1.470
Gauteng	1.871	2.254	3.723	3.767
Kwa-Zulu-Natal	1.789	3.424	3.585	3.163
Limpopo	1.000	1.000	1.000	1.000
Mpumalanga	1.697	1.357	1.360	1.476
Northern Cape	1.451	1.478	2.007	1.848
North West	1.484	1.457	1.265	1.436
Western Cape	1.476	1.560	1.456	1.280

Table 3d: Discretionary to Conditional Grant Ratio

	2002	2003	2004	2005
Eastern Cape	1.168	0.822	0.598	0.473
Free State	0.095	0.156	0.094	0.000
Gauteng	1.889	1.809	2.282	2.722
Kwa-Zulu-Natal	0.711	1.880	1.371	1.440
Limpopo	0.407	0.440	0.110	0.000
Mpumalanga	0.351	0.229	0.000	0.000
Northern Cape	0.000	0.000	0.000	0.000

[16] Reproduced from the ASSA's AIDS Model and from Hickey et al (2003).

North West	0.608	0.304	0.000	0.313
Western Cape	2.042	1.242	0.650	0.662

Table 3e: Percent of Provincial Budget Spent on Health

	2002	**2003**	**2004**	**2005**
Eastern Cape	17.4	18.3	19.1	19.1
Free State	21.7	22.4	22.2	22.0
Gauteng	31.3	30.0	29.2	27.6
Kwa-Zulu-Natal	25.6	24.5	23.8	23.1
Limpopo	16.4	16.2	16.4	16.2
Mpumalanga	17.5	18.5	18.5	18.3
Northern Cape	16.4	19.3	19.4	20.7
North West	17.1	17.9	17.7	18.4
Western Cape	27.0	27.5	26.8	25.9

REFERENCES

Coelli, T. (1996), "A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program," *Centre for Efficiency and Productivity Analysis Working Paper 96/08*.

Coelli, T. (1997), "A Multi-Stage Methodology for the Solution of Oriented DEA Models," *Centre for Efficiency and Productivity Analysis Working Paper*.

Hickey, A. et al. (2003), "Budgeting for HIV/AIDS in South Africa: Report on Intergovernmental Funding Flows for an Integrated Response in the Social Sector," IDASA – Budget Information Service Report. http://www.internationalbudget.org/resources/library/HIV.pdf.

Martin. G. (2003), "A Comparative Analysis of the Financing of HIV/AIDS Programmes in Botswana, Lesotho, Mozambique, South Africa, Swaziland and Zimbabwe," HSRC Publishers: http://www.synergyaids.com/documents/Finance_CompAnalys.pdf.

Rosenman, R., and Friesner, D. (2004), "Scope and Scale Inefficiencies in Physician Practices," *Health Economics*, Vol. 13, pp. 1091-1116.

Saxenian, H., and Schwab, N. (2003), "The ABC Model: Optimal Allocation of Resources for HIV Prevention," http://www.iaen.org/files.cgi/10177_schwab--Presentation_for_IAEN_April_24.pdf.

Simar, L. and Wilson, P. (2003), "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes," Discussion Paper 0307, Institute De Statistic at Universite Catholique de Louvain.

United Nations. (2002), "World Population Prospects: The 2002 Revision, Volume III: Analytical Report," United Nations Department of Economic and Social Affairs/Population Division: http://www.un.org/esa/population/publications/wpp2002/WPP2002_Vol3.htm.