OXFORD

Estimating a cost-effectiveness threshold for health care decision-making in South Africa

Ijeoma P Edoka* and Nicholas K Stacey

SAMRC Centre for Health Economics and Decision Science—PRICELESS SA, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, South Africa

*Corresponding author. SAMRC Centre for Health Economics and Decision Science—PRICELESS SA, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Room 230, 27 St. Andrews Road, Parktown, 2193 Johannesburg, South Africa. Tel: +27 11 717 2374; E-mail: ijeoma.edoka@wits.ac.za

Abstract

Cost-effectiveness thresholds are important decision rules that determine whether health interventions represent good value for money. In low- and middle-income countries, the World Health Organization (WHO) one to three times per capita gross domestic product (GDP) per disability-adjusted life years (DALYs) averted has been the most widely used threshold for informing resource allocation decisions. However, in 2016, the WHO withdrew recommendations for using this threshold, creating a significant vacuum in South Africa and many countries that rely on results of cost-effectiveness analyses for making resource allocation decisions. This study estimates a cost-effectiveness threshold that reflects the health opportunity cost of health spending in South Africa using a three-step approach. First, marginal returns to health spending was estimated as health spending elasticity for crude death rates using a fixed effect estimation approach. Second, the opportunity cost of health spending was estimated as DALYs averted. Finally, a cost per DALY averted threshold was estimated as the inverse of the marginal product of health spending. We show that 1% of total health spending in 2015 (equivalent to approximately ZAR 1.54 billion/USD 120.7 million) averted 1050 deaths, 34 180 years of life lost, 5880 years lived with disability and 40 055 DALYs. The cost-effectiveness threshold was estimated at approximately ZAR 38 500 (USD 3015) per DALY averted, ~53% of South Africa's per capita GDP in 2015 (ZAR 72 700/USD 5700) and lower than the previously recommended one to three times per capita GDP. As South Africa moves towards implementing universal health coverage reforms through National Health Insurance by 2025, the adoption of a threshold that reflects health opportunity costs will be crucial for ensuring efficiency in the allocation of scarce resources. This study provides useful insight into the magnitude of the health opportunity cost of health spending in South Africa and highlights the need for further research.

Keywords: health opportunity cost, cost-effectiveness threshold, marginal returns, cost-effectiveness analysis, health expenditure, decision-making, resource allocation

Introduction

Cost-effectiveness analyses are globally accepted as an important tool for making resource allocation decisions within health care systems. Cost-effectiveness analyses compare costs and health benefits of competing health interventions to assess whether an intervention represents good value for money. Cost-effective interventions are typically identified by comparing the incremental cost per health unit gained or incremental cost-effectiveness ratio (ICER) to a 'benchmark' value or reference ICER, the cost-effectiveness threshold. Interventions with ICERs below the threshold are considered cost-effective and represents good value for money. Thresholds are, therefore, important for making resource allocation decisions and have been widely adopted in several countries for this purpose (Leech *et al.*, 2018). However, in spite of their widespread use, there remains limited understanding on what they ought to represent and how they can be derived (Gafni and Birch, 2006).

Key Messages
- Cost-effectiveness thresholds are important decision rules that determine whether health interventions represent good value for money.
- A cost-effectiveness threshold for South Africa that reflects the health opportunity costs of health spending was estimated at ZAR 38 500 (USD 3015) per DALY averted, ~53% of South Africa's GPD per capita in 2015.

In many low- and middle-income countries (LMICs), the World Health Organization (WHO) recommendation of one to three times a country's gross domestic product (GDP) per capita per disability-adjusted life year (DALY) has been the most widely used criteria for determining whether interventions are cost-effectiveness (Hutubessy *et al.*, 2003; Leech *et al.*, 2018). However, in recent years, the application of this decision rule has been extensively criticized for lacking scientific underpinnings to justify their use in making resource allocation decisions, resulting in a growing call for empirically estimated context-specific thresholds (Newall *et al.*, 2014; Marseille *et al.*, 2015; Bertram *et al.*, 2016; Leech *et al.*, 2018). Decisions relying on inappropriately set cost-effectiveness thresholds can result in inefficient allocation of resources and consequently in net population health losses.

Different approaches have been proposed for empirically estimating thresholds, broadly classified based on what they represent (Vallejo-Torres *et al.*, 2016; Robinson *et al.*, 2017). These include thresholds representing the social demand for health and those representing the health opportunity costs of health spending. The former is either based on estimates of individual or societal willingness to pay for additional health benefits. They reflect the value placed on consumption goods society or individuals are willing to give up 'outside' the health system for additional health gains. In other words, they reflect the non-health opportunity costs of health spending and are often referred to as 'demand-side thresholds' (Vallejo-Torres *et al.*, 2016; Robinson *et al.*, 2017). The second category of thresholds is based on estimates of the marginal productivity of the health system (Claxton *et al.*, 2015; Woods *et al.*, 2016; Edney *et al.*, 2018; Ochalek *et al.*, 2018; Vallejo-Torres *et al.*, 2018). These are the so-called 'supply-side thresholds'. They represent the health opportunity cost of resource allocation decisions or the value of health that will be given up elsewhere in the health system when additional costs are imposed on the health system with a fixed budget (Claxton *et al.*, 2015; Woods *et al.*, 2016; Edney *et al.*, 2018; Ochalek *et al.*, 2018; Vallejo-Torres *et al.*, 2018).

Where the aim of the decision-maker is to maximize population health within a constrained budget, priority should be given to interventions that produce positive population net health benefits, i.e. to interventions producing health benefits that outweigh their health opportunity costs. Thus, thresholds determining whether a health intervention represents good value for money ought to reflect the value of health benefits that will be forgone elsewhere in the health system due to funding decisions. Demand-side thresholds including the one to three times per capita GDP threshold, fails to capture these trade-offs thus potentially leading to the implementation of health interventions that may result in population net health losses. Recognizing these limitations, the WHO in 2016 cautioned against the sole use of the one to three times per capita GDP threshold for prioritizing health interventions, creating a significant gap across many LMICs including South Africa that rely on results of cost-effectiveness analysis for making resource allocation decisions (Bertram *et al.*, 2016). The aim of this study is to estimate a cost-effectiveness threshold that reflects the health opportunity cost of health spending in South Africa.

There are different approaches proposed for estimating a supply-side threshold (Thokala *et al.*, 2018). An example is the league table approach, which ranks interventions based on their ICERs, from the lowest to highest ICER. The threshold is defined as the ICER of the last intervention that can be funded within an existing budget (Weinstein and Zeckhauser, 1973; Gold *et al.*, 1996). This approach requires the identification of the costs and benefits of all interventions. However, in practice, most countries, particularly LMICs, lack data required to estimate ICERs for all potential interventions. Furthermore, methodological differences in the estimation of ICERs for different interventions could limit the interpretation and application of league tables (Drummond *et al.*, 1993). Other previously proposed approaches include the identification of interventions previously disinvested to fund new interventions that imposed additional costs to the health care system and the identification of the health consequences associated with these decisions (Appleby *et al.*, 2009). Challenges with identifying all possible interventions that may have been withdrawn as a consequence of new funding decisions as well as the health impact of these decisions have limited the application of this approach in many settings (Vallejo-Torres *et al.*, 2016).

More recent approaches to estimating a supply-side threshold rely on the identification of health benefits associated with marginal changes in health spending. These thresholds are derived from estimates of the marginal productivity of the health system—i.e. the quantity of health a health system is capable of producing for marginal changes in health spending. A growing body of evidence is now emerging, particularly in high-income countries, on thresholds estimated using this approach (Claxton *et al.*, 2015; Edney *et al.*, 2018; Vallejo-Torres *et al.*, 2018). These studies show that supply-side thresholds are typically lower than demand-side thresholds. In two studies focusing on LMICs, similar findings have been reported. Supply-side thresholds were shown to be considerably less than the one to three times per capita GDP threshold for many LMICs, including South Africa (Woods *et al.*, 2016; Ochalek *et al.*, 2018). These studies estimated supply-side thresholds using either marginal productivity estimates obtained from multi-country data analysis (Ochalek *et al.*, 2018) or from UK estimates of health opportunity costs adjusted by the income elasticity of the demand for health in each country (Woods *et al.*, 2016). While these studies provide useful insights into where thresholds for LMICs are likely to fall and demonstrate the inappropriateness of previously adopted universal decision rules, they largely rely on multi-country data, potentially failing to account sufficiently for contextual nuances and heterogeneity between health systems (Edoka, 2019).

This study, the first to estimates a supply-side cost per DALY threshold for an upper middle-income country using subnational level data, contributes significantly to strengthening the growing body of evidence on appropriate methods for estimating supply-side thresholds in LMICs. It provides a context-specific estimate and approach to estimating cost-effectiveness thresholds using subnational health expenditure and epidemiology data.

Public health spending in South Africa

Every year South Africa's general tax revenue is divided equitably between the following three spheres of government: the national government, 9 provincial governments and 257 municipalities. The vertical division of general tax revenue across government spheres explicitly takes into account the functions assigned to each sphere and each sphere's ability to raise revenue internally. The provision of health care services is a competency of provincial governments, who have limited means of raising revenue and primarily rely on transfers from the national government to finance provincial health care systems.

Transfers to each province are either in the form of unconditional transfers or conditional grants. Unconditional transfers are determined by a provincial equitable share formula (National Treasury, 2016a) which consists of six components—education, health, basic, institutional, poverty and economic components. It explicitly takes into account the relative demand for public services between provinces in determining the size of total transfers to each province. Collectively, the education and health component account for ~75% of total transfers to provinces (National Treasury, 2016a). Thus, variations in total transfers between provinces are largely driven by variations in the demand for basic education, health care need (variations in health risk profile of the uninsured population) and demand for health services (variations in primary health care utilization rate and hospital caseload). Although the provincial equitable share formula is largely meant to reflect national priorities across sectors, each province has complete autonomy over how it allocates resources across its functions.

In addition to unconditional transfers, provincial governments receive conditional grants from the national government to deliver specific services and undertake capital investment as stipulated by the conditions of the grants. Conditional grants, which account for ~20% of total transfers to provinces, are instruments used by the national government to ensure that national priorities in certain disease areas, such as human immunodeficiency virus (HIV) are met (National Treasury, 2016a). The need for such services determines the size of each province's conditional grants.

Methods

Our study adopts recent approaches to estimating supply-side thresholds that rely on estimates of marginal productivity of the health system (Claxton *et al.*, 2015; Edney *et al.*, 2018; Ochalek *et al.*, 2018; Vallejo-Torres *et al.*, 2018). We describe this approach in three steps. In the first step, the effect of health spending on crude death rates was estimated using a fixed effect estimation approach for long panel data (Cameron and Trivedi, 2005). In the second step, the health opportunity cost of health spending was estimated as the number of deaths and DALYs averted by a change in health spending. Finally, a cost per DALY averted threshold, reflecting the health opportunity cost of health spending, was estimated.

Estimating marginal returns to public health spending in South Africa

To estimate marginal returns to public health spending in South Africa, we use a panel data of all provinces pooled over a 14-year (2002–15) period. We exploit provincial-level variations in health spending and crude death rates to identify marginal returns to health spending. The approach controls for a wide range of factors that are likely to simultaneously affect both the level of health spending and

crude death rates as well as unobserved province- and time-period fixed effects:

$$\ln(M_{it}) = \beta_0 + \beta_1 \ln(HE_{it}) + \beta_2' \sum_{l=1}^{3} \ln(HE_{it-l}) + \beta_3' \ln(X_{it}) + \rho_i + \tau_t + \varepsilon_{it},$$

(1)

where M_{it} is age-standardized crude death rates (per 100 000 population) in province i at time t and HE_{it} is health spending per capita in province i at time t. In South Africa, resource allocation at the provincial level is largely done on an historical basis such that health spending from previous years determines contemporaneous health spending. Given that lagged health spending is also likely to affect contemporaneous death rates, we control for lagged health spending (HE_{it-l}). In addition, we controlled for a wide range of observable characteristics of provinces that are likely to be correlated with contemporaneous health spending and death rates (represented by the vector X_{it}). These include sociodemographic and economic factors that may predict population mortality rates (education, age–gender composition of the population and access to safe drinking water and sanitation (Filmer and Pritchett, 1999)), and other factors that may simultaneously affect the size of the total transfers to provinces and mortality rates (medical aid coverage, HIV prevalence as well as primary health care utilization rates and hospital bed occupancy rates) Table 1 provides a comprehensive list of all control variables included in the analysis. ρ_i represents panel fixed effects, which captures unobserved time-invariant province-specific characteristics such as quality of governance, administrative structures and the level of corruption within provincial health system, which may affect the efficiency of health spending. τ_t represents general time trends, which captures unobserved time-varying factors that may affect health spending and/or mortality rates uniformly in all provinces, such as health technological advancements, or nationwide macro-economic shocks. Finally, ε_{it} is the random independent and identically distributed error term.

The parameter of interest, β_1, describes the marginal instantaneous impact of health spending on crude death rates and is interpreted as an elasticity, i.e. the percentage change in age-standardized crude death rate for a 1% change in per capita health spending.

Given that health spending is unlikely to affect each age group uniformly, as a robustness check, we estimate the effect of health on crude death rates in 17 five-year age group categories (0–4, 5–9, ..., ≥80 years age group):

$$\ln(MS_{itc}) = \gamma_0 + \gamma_1 \ln(HE_{it}) + \gamma_2' \sum_{l=1}^{3} \ln(HE_{it-l}) + \gamma_3' \ln(X_{it})$$
$$+ \rho_i + \tau_t + \varepsilon_{itc},$$

(2)

where MS_{itc} represents age-specific crude death rates per 100 000 population for age group c in province i at time t. The parameter of interest, γ_1, is the health spending elasticity, interpreted as the percentage change in age-specific death rates for a 1% change in health spending.

Estimating the health opportunity cost of public health spending in South Africa

First, we estimate the number of deaths averted by 1% of total health expenditure in a given year by applying elasticities obtained from step 1 (1) to the total number of deaths recorded in the same year. We use 2015 as our primary year of interest and estimate deaths averted both for the entire population (3) and for each of the 17 age group categories (4) previously described:

Table 1 List of variables and data sources

Variables	Data source
Dependent variables	
Age-specific crude death rates	StatsSA Vital Statistics
	StatsSA Mid-Year Population Estimates
Independent variable	
Health expenditure	National Treasury Estimates of Provincial Revenue and Expenditure
Control variables	
GDP per capita	StatsSA
Percentage of population estimated to be HIV positive	Thembisa model
Average number of PHC visits per person per year	HST District Health Barometer
Average number of beds occupied	HST District Health Barometer
Proportion of population with medical aid	StatsSA General Household Survey
Proportion of population living in urban areas	StatsSA General Household Survey
Proportion of population with tertiary education	StatsSA General Household Survey
Proportion of population with safe drinking water supply	StatsSA General Household Survey
Proportion of population with access to safe sanitary systems	StatsSA General Household Survey
Proportion of female population over 50 years	StatsSA General Household Survey
Proportion of male population over 50 years	StatsSA General Household Survey
Other variables	
Population estimates	StatsSA Mid-Year Population Estimates
Years of life lost in total population	IHME Global Burden of Disease Study
Years lived with disability in total population	IHME Global Burden of Disease Study
Conditional life expectancy	WHO GHO database

StatsSA, Statistics South Africa; HST, Health Systems Trust; IHME, Institute for Health Metrics and Evaluation; WHO GHO, World Health Organization Global Health Observatory.

$$\text{Deaths averted}_t = \frac{\beta_1}{100} \times \text{Deaths}_t, \tag{3}$$

$$\text{Deaths averted}_t^c = \frac{\gamma_1}{100} \times \text{Deaths}_t^c, \tag{4}$$

where t is the year 2015; β_1 is the heath spending elasticity for crude death rates in the entire population; and γ_1 is the health spending elasticity for age-specific death rates.

Next, we estimate the number of DALYs averted. DALYs are calculated as the sum of the years of life lost (YLL) due to premature mortality and the years lost due to disability (YLD). Total DALYs averted by 1% of total health spending in 2015 can thus be estimated as the sum of the effect of health spending on survival (YLL averted) and the effect of health spending on morbidity (YLD averted). We do this in three steps.

First, we indirectly estimated the effect of health spending on survival (YLL) at time $t = 2015$ by applying deaths averted [estimated from (3)] to average conditional life expectancy (CLE) for the entire population in 2015:

$$\text{YLL averted}_t = \text{Deaths averted}_t \times \text{CLE}_t. \tag{5}$$

YLL averted for each age group was estimated similarly, using the number of deaths averted (estimated from equation 4) and CLE for each age group.

Second, the effect of health spending on disease morbidity (YLD) was similarly estimated indirectly. Health spending is likely to affect morbidity in two ways—positively, through disease prevention and/or treatment and, negatively, through lengthening survival and consequently, the number of years that could potentially be lived with disability. Both positive and negative effects of health spending on disease morbidity were estimated indirectly from the health spending effect on survival, based on the assumption that the effect of health spending on mortality and survival is proportional to the effect of health spending on morbidity (Ochalek *et al.*, 2018):

$$\text{YLD averted}_t = \underbrace{\text{YLL averted}_t \times \frac{\text{YLD}_t}{\text{YLL}_t}}_{\text{positive effect}}$$
$$\underbrace{- \text{YLL averted}_t \times \text{YLD per capita}_t}_{\text{negative effect}}, \tag{6}$$

where YLD_t is the disease morbidity burden for the entire South African population (Murray *et al.*, 2012)[1] at $t = 2015$ and YLL_t is the disease mortality burden for the entire population in the same year.

Finally, total DALYs averted by 1% of total health spending at $t = 2015$ was estimated as the sum of the health spending effect on survival (5) and the health spending effect on morbidity (6):

$$\text{DALYs averted}_t = \text{YLL averted}_t + \text{YLD averted}_t. \tag{7}$$

This represents the health opportunity cost of 1% of total health spending in 2015.

Estimating a cost per DALY averted threshold for South Africa

Using the health opportunity cost expressed as the number of DALYs averted by 1% of total health spending (or the marginal product of health spending), the cost per DALY averted was estimated as the inverse of the marginal product of health spending:

$$\text{cost per DALY averted}_t = \frac{1\% \times \text{total health spending}_t}{\text{DALYs averted}_t}. \tag{8}$$

Cost per deaths averted and cost per YLL averted were similarly estimated by replacing DALYs averted in (8) with deaths averted and YLL averted estimated from (3) and (5), respectively.

Equation (8) shows that cost per DALY averted is directly linked to the size of the total health budget. This suggest that the cost per DALY averted will increase (decrease) with respect to an increase

(decrease) in total health spending. Therefore, we estimate a cost per DALY averted threshold for each year from 2002 to 2015 by rerunning (3–8) using year-specific data on health spending, mortality and morbidity.

Data sources and variables
This study pools data from various sources to construct a panel of province-level observations over a 14-year (2002–15) time period.

Mortality
All-cause mortality (number of deaths) was obtained from the Department of Home Affairs death register published annually by Statistics South Africa (2002–2015b). This register contains individual-level data on deaths occurring in South Africa between January and December of each year, including information on the province where death occurred, the cause of death and sociodemographic characteristics (age, gender, marital status, occupation and education) of the deceased. We merged mortality data to mid-year population projection estimates from Statistics South Africa (Statistics South Africa, 2002–2015a) and a South Africa standard population estimate (Dorrington, 2013) to estimate province-level crude death rates per 100 000 age-standardized population from 2002 to 2015. Age-specific death rates were estimated for each 17 five-year age group categories using epidemiological and demographic estimates for each age group.

Health expenditure
Annually, the National Treasury publishes Provincial Budget Reviews, which contains detailed information on total spending in each province, disaggregated by different public sectors. Data on audited health expenditure for each tax year (1 April to 31 March) were extracted from Estimates of Provincial Revenue and Expenditure (National Treasury, 2016b). In the regression analysis, health spending for each year is reported in 2010 prices.

Other covariates
We include a wide range of variables that are likely to explain variations in both crude death rates and the level of health spending within provinces. These include the proportion of individuals with private medical insurance, average number of primary health care visits per person per year, hospital bed occupancy rates and HIV prevalence. The first three covariates determine the size of the health share within the equitable share formula while the last covariate determines the size of the HIV conditional grant transfers to provinces. Other variables controlled for include GDP per capita (which determines the size of total transfers to provinces through the economic component of the equitable share formula and may also have an indirect effect on mortality). These variables were obtained from various sources including the South African General Household Survey, the Health System Trust District Health Barometer and the Thembisa model of HIV epidemic in South Africa (Johnson *et al.*, 2016). Table 1 outlines all variables and data sources.

Other covariates including the proportion of households living in urban areas, the proportion of individuals with higher education, the proportion of males and females over the age of 50 years and the proportion of household with access to safe drinking water supply and sanitation were obtained from the South African General Household Survey (Table 1)

Finally, epidemiological data used for estimating DALYs averted including total population YLD and YLL were obtained from the Institute for Health Metrics and Evaluation Global Burden of
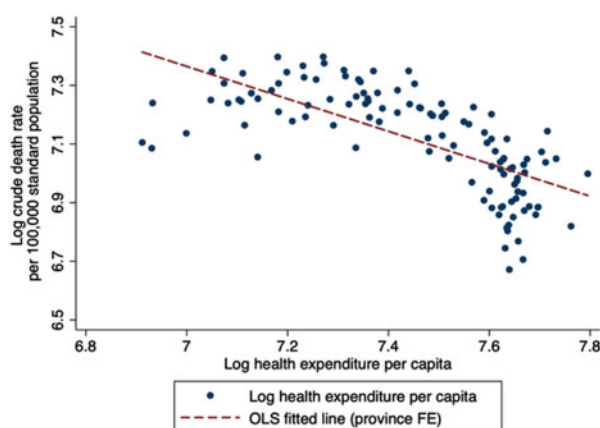

Figure 1 Relationship between health spending and crude death rates

Disease database while CLE were obtained from the WHO Global Health Observatory databases (Table 1).

Results
Figure 1 depicts the negative relationship between health spending and death rates: (log)crude death rates decline as (log)health spending increases. Although province fixed effects are accounted for, the relationship described in Figure 1 is likely to be confounded by temporal changes in other factors that affect both health spending and mortality. The fixed effect model adopted here accounts for these.

Using variations in health spending and crude death rates observed within each province and controlling for a wide range of potentially confounding factors including unobservable time-invariant province fixed effects and aggregate time trends, we estimated a health spending elasticity for age-standardized crude death rate of −0.223. This implies that on average, a 1% increase in health spending per capita will result in a 0.223% decrease in age-standardized death rates in the entire population. In 2015, approximately ZAR 154.07 billion was spent on health in South Africa. Therefore, 1% of total spending in 2015, equivalent to approximately ZAR 1.54 billion, averted ∼1050 deaths, 34 180 YLLs, 5880 YLD and 40 055 DALYs nationally (Table 2). Using these estimates, we estimated a threshold of approximately ZAR 38 500 per DALY averted/USD 3015 per DALY averted (Table 2). This represents ∼53% of the GDP per capita in 2015.

Robustness check
Given that health spending is unlikely to affect each age group uniformly, as a robustness check, we estimated the health opportunity cost of health spending and cost per DALY averted based on health spending elasticities for age-specific crude death rates. Table 3 summaries the elasticities for each age group. The effect of health spending is negative and significant for the majority of the age groups. The highest effect is observed in children under the age of 5 years—a 1% increase in health spending reduces under-five crude death rates by ∼0.52% (Table 3). Applying age-specific elasticities to 2015 total health expenditure, mortality and morbidity data, we estimated a marginally higher threshold of ∼ZAR 45 630 (USD 3577) per DALY averted (∼63% of South Africa's GDP per capita in 2015).

Cost per DALY averted and health spending
Table 4 shows cost per DALY averted for each year, estimated using the health spending elasticity for age-standardized crude death rate

Table 2 Health opportunity costs and cost-effectiveness thresholds estimated using health spending elasticities for age-standardized crude death rate[a]

	Elasticity[b]	Average CLE	Deaths averted	YLL Averted	YLD Averted	DALYs Averted
Age-standardised population average	−0.223**	32.66	1047	34 177	5878	40 055

Cost-effectiveness threshold	Cost per DALY averted	Cost per YLL averted	Cost per death averted			
ZAR	38 465.46	45 081.36	1 472 262.41			
USD[c]	3 014.77	3 533.30	115 340.56			
%GDP per capita[d]	53%					

[a]A full list of data inputs used in estimating cost per DALYs/YLL/Deaths averted is presented in Table 1A of the Appendix.

[b]Elasticity estimated using a fixed effect estimation approach for long panel data. Control variables include GDP per capita, HIV prevalence, medical aid coverage, primary healthcare and hospital service utilization rates, proportion of the population: living in urban areas, with a higher education, with access to safe drinking water and sanitation and proportion of female and male population over the age of 50 years.

[c]USD-ZAR Exchange rate = ZAR 12.76/USD.

[d]2015 GDP per capita = USD 5700.

**p<0.05.

CLE = conditional life expectancy, YLL = years of life lost, YLD = years lived with disability, DALY = disability –adjusted life year.

Table 3 Health opportunity costs and cost-effectiveness thresholds estimated using health spending elasticities for age-specific crude death rates

Age group	Elasticity[a]	Conditional LE	Deaths averted	YLL Averted
<1–4 years	−0.519**	64	230	14 692
5–9 years	0.487	61.15	−15	−922
10–14 years	0.577*	56.45	−18	−1008
15–19 years	0.12	51.8	−8	−425
20–24 years	−0.365**	47.15	51	2393
25–29 years	−0.15	42.65	34	1452
30–34 years	−0.316**	38.4	90	3460
35–39 years	−0.240*	34.4	70	2409
40–44 years	−0.285**	30.75	84	2589
45–49 years	−0.242*	27.05	70	1900
50–54 years	−0.14	23.3	44	1035
55–59 years	−0.323***	19.7	109	2152
60–64 years	0.106	16.3	−38	−619
65–69 years	−0.101	13.25	35	460
70–74 years	0.244*	10.65	−76	−814
75–79 years	−0.122	8.35	38	314
80+ years	0.0756	5.575	−46	−258
Age-specific population average			653	28 809

Cost-effectiveness threshold:	Cost per DALY averted	Cost per YLL averted	Cost per death averted
ZAR	45 632.73	53 481.37	2 359 463.27
USD[b]	3 576.51	4 191.66	184 925.41
% GDP per capita[c]	63%		

[a]Elasticity estimated using a fixed effect estimation approach for long panel data. Control variables include GDP per capita, HIV prevalence, medical aid coverage, primary healthcare and hospital service utilization rates, proportion of the population: living in urban areas, with a higher education, with access to safe drinking water and sanitation and proportion of female and male population over the age of 50 years.

[b]USD-ZAR Exchange rate = ZAR 12.76/USD.

[c]2015 GDP per capita = USD 5700.

*p<0.1, **p<0.05, ***p<0.01.

LE = life expectancy, YLL = years of life lost, DALY = disability–adjusted life year.

(−0.223) and year-specific data on health spending, mortality and morbidity. To facilitate comparison across time, we present cost per DALY averted estimates in 2015 prices. Between 2002 and 2015, we observe an increase in cost per DALY averted (Figure 2). However, trends observed in real cost per DALY averted suggest subtle changes over time, especially over short time periods (3–4 years).

Table 4 Health benefits of one percent of total health spending and cost per DALY averted[a] (2002–2015)

Year	Deaths averted	YLL averted	YLD averted	DALYs averted	Cost per DALY averted (current ZAR)	Cost per DALY averted (2015 ZAR)
2002	1135	33 970	2608	36 578	9074.52	18 297.54
2003	1256	36 989	2090	39 079	9574.61	18 237.39
2004	1300	37 672	1908	39 580	10 236.07	19 230.90
2005	1337	38 363	1863	40 226	11 685.37	21 232.00
2006	1377	39 434	1888	41 321	12 983.12	22 543.60
2007	1366	39 207	2030	41 238	15 174.92	24 602.97
2008	1347	39 196	2284	41 480	18 209.60	26 469.43
2009	1297	38 418	2593	41 011	21 386.15	29 017.89
2010	1231	36 999	2984	39 983	24 501.26	31 887.08
2011	1141	34 862	3369	38 231	29 114.33	36 086.25
2012	1085	34 039	3965	38 004	32 242.24	37 824.74
2013	1042	33 282	4584	37 866	34 508.82	38 281.96
2014	1030	33 353	5205	38 557	36 534.68	38 210.98
2015	1047	34 177	5878	40 055	38 465.46	38 465.46

[a]Details of data inputs used in estimating health benefits and cost per DALY averted are outlined in Table 1A of the Appendix.

YLL = years of life lost, YLD = years lived with disability; DALY = disability-adjusted life year.
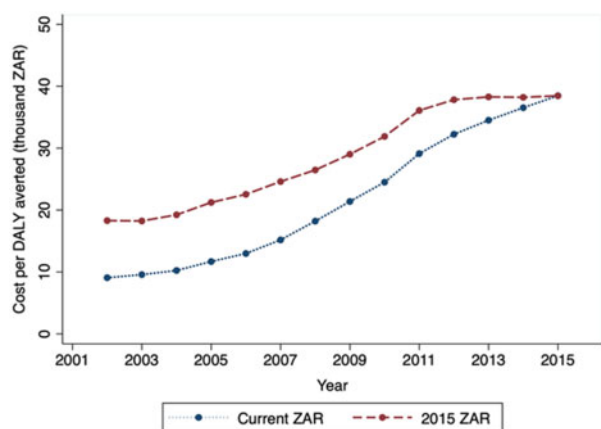
Figure 2 Trends in cost per DALY averted (2002–15)
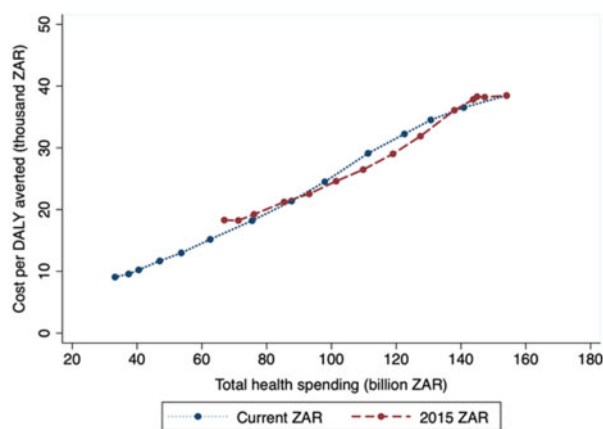
Figure 3 Cost per DALY averted and total health spending

Overall, as expected, the relationship between cost per DALY averted and health spending is positive (Figure 3).

Discussion

Cost-effectiveness thresholds are important criteria for assessing whether new health interventions that impose additional costs on the health system represent good value for money. Given budget constraints, funding new, more expensive health interventions may result in the displacement of existing interventions and their corresponding health benefits. The decision to fund a new health intervention should therefore be based on the value of health benefits that will be displaced elsewhere in the health system in order to fund the new intervention—i.e. the health opportunity cost of health spending. In many LMICs, including South Africa, the WHO one to three times per capita GDP threshold has been the most widely used benchmark for determining the cost-effectiveness of new health interventions. However, the WHO threshold, which is based on the value of a life year (World Health Organization, 2001), does not reflect the health opportunity cost of health spending, potentially resulting in population net health losses.

Using subnational data, we estimated a South Africa-specific cost per DALY averted threshold, which reflects the health opportunity cost of health spending. This threshold is based on estimates of the marginal productivity of the health system. Estimating marginal returns to health spending requires disentangling the effects of health spending from the effect of other determinants of population health, such as socioeconomic and environmental factors that are equally important in explaining population health (Edoka, 2019). This presents important methodological challenges due to the endogeneity of health spending that arises from unobserved heterogeneity and endogeneity arising from the reverse causal relationship between health spending and population health outcomes (Edoka, 2019). A failure to account for both sources of endogeneity will introduce an attenuation bias to estimates of the effect of health spending on population health outcomes. We adopt a fixed effect econometric approach to address some of these challenges, pooling data across all nine provinces in South Africa and exploiting province-year variations in health spending and crude death rates from 2002 to 2015. Our model accounts for heterogeneity across provinces by controlling for a range of observable time-variant factors that may affect both the level of health spending and population mortality rates within provinces. In addition, the fixed effect model controls for unobservable heterogeneity including unobservable time-invariant factors specific to each province and unobservable time-varying factors that affect all provinces uniformly.

Similar to previous studies, we observe a negative and statistically significant effect of health spending on death rates (Gallet and Doucouliagos, 2017). Our findings show that a 1% increase in health spending reduces crude death rates by ∼0.223%. This estimate is higher than estimates reported in other studies in similar settings. The majority of studies estimating health spending elasticities in LMICs have adopted a multi-country approach, exploiting cross-country variations in health spending and death rates. Elasticities estimated using multi-country data are typically observed to be lower than estimates from single-country analysis. For example, Ochalek *et al.* (2018) using an instrumental variable approach estimated elasticities for 127 countries including South Africa, ranging from −0.25 to −0.35 for under-five mortality, −0.08 to −0.19 for adult female mortality and −0.1 to −0.2 for adult male mortality. Using a similar estimation strategy and dataset, Bokhari *et al.* (2007) estimated a South Africa-specific elasticity of −0.34 for under-five mortality. Our corresponding elasticities for under-five mortality, adult male and adult female crude death rates were −0.55, −0.28 and −0.25, respectively. Differences in estimation strategies (fixed effect vs instrumental variable approach) may partly explain the differences between our findings and those of previous studies. However, a failure to fully account for unobserved heterogeneity in multi-country studies may result in an attenuation bias of estimates of health spending elasticities, thus explaining the lower estimates reported in previous studies.

In the absence of provincial-level data on DALYs, we indirectly estimated a cost per DALY averted threshold using health spending elasticities for crude death rates and indirectly estimated YLL averted, YLD averted and DALYs averted by 1% of the total health expenditure in 2015. The threshold, estimated at approximately ZAR 38 500 (USD 3015) per DALY averted, represents the cost of averting one DALY within the South African health system in 2015 or the 'shadow price' of public health expenditure. Therefore, new health interventions with incremental cost per DALY averted lower than ZAR 38 500 will generate positive population net health benefits, i.e. for every dollar spent on these interventions, more health will be gained in the population than would be displaced elsewhere when resources have to be reallocated to fund new interventions. Our estimate is equivalent to ∼53% of South Africa's GDP per capita in 2015, lower than the previously recommended threshold of one to three times per capita GDP. This is likely to affect decisions on what gets funded within the public health system as has been demonstrated in other settings where interventions regarded as cost-effective under the per capita GDP decision rules are not cost-effective when decision rules reflecting health opportunity costs are applied (Loganathan *et al.*, 2018).

Our study explicitly demonstrates the link between health budgets and the health opportunity cost of health spending. Given South Africa's 3-year budgeting system, we observe small changes in cost per DALY averted in the short run. For example, between 2012 and 2015, cost per DALY averted remains constant at approximately ZAR 38 000, after accounting for inflation. Short-term constraints in real growth of the health budget in South Africa may partly explain this. South Africa operates a Medium Term Expenditure Framework–a 3-year rolling budget system that provides the basis for annual budget planning over 3-year intervals. This implies that the budget is 'fixed' in the medium term (3 years). Although reallocation of resources between health programmes is permitted on an annual basis to reflect changing health systems priorities, this is often done within budget limits outlined in the 3-year budget plan (National Treasury, 2019). However, substantial expansions to the health budget in the long-term will warrant adjustments to the threshold. This has implications for the

magnitude of the cost-effectiveness threshold that could be applied for resource allocation decisions under the proposed National Health Insurance (NHI). Proposed reforms under the NHI are expected to be accompanied by substantial expansions to the health budget and, possibly, improvements in the productivity of the public health system (NDOH, 2017). Both changes will have contrasting effects on cost per DALY averted—on the one hand, higher productivity of the health system will result in a lower cost per DALY averted, ceteris paribus. Conversely, an increase in the size of the health budget will result in a higher cost per DALY averted threshold, all else held constant. However, it is unclear how the combination of both factors will play out in predicting, a priori, the direction of the magnitude of the cost per DALY averted threshold.

The choice of the health outcome measure also has implications for the magnitude of the health opportunity cost of health spending. Our findings show that, for the same budget or level of health spending, the health opportunity cost is higher when health gains are valued on one health dimension (mortality or YLL) compared with health measures that capture more dimensions of health (mortality and morbidity—DALYs). For the latter, this implies that methodological differences in estimating DALYs and QALYs may result in different magnitudes of the health benefits of health spending and by extension, differences in the magnitude of the cost-effectiveness threshold (Gold *et al.*, 2002; Bevan and Hollinghurst, 2003; Airoldi and Morton, 2009; Augustovski *et al.*, 2017). Therefore, the cost per DALY averted threshold estimated here may not necessarily be appropriate for informing resource allocation decisions for interventions expressed in terms of incremental cost per QALY gained. Further research is needed to understand how differences in health values captured in DALYs and QALYs may affect estimates of the health opportunity cost of health spending. This is particularly relevant for LMICs such as South Africa where both DALYs and QALYs are used interchangeable in economic evaluations of new health interventions (NDOH, 2012).

Overall, the estimate of cost per DALY averted reported here is consistent with estimates reported elsewhere for South Africa (Woods *et al.*, 2016; Ochalek *et al.*, 2018). For example, using a similar approach but based on country-specific elasticities derived from multi-country analysis, Ochalek *et al.* (2018) reported a cost per DALY of approximately USD 3300 per DALY averted, ∼58% of South Africa's GDP per capita [DALY1 in Ochalek *et al.* (2018)]. Difference in health spending elasticities may explain the differences between our findings. Furthermore, while our estimates are based on average effects of health spending across the entire population, estimates reported in Ochalek *et al.* (2018) are based on extrapolations of health spending effects for under-five and adult mortality rates to health spending effects for the entire population. Nevertheless, our findings are comparable, suggesting that the cost per DALY averted for South Africa is likely to fall within ∼50–70% of South Africa's GDP per capita.

While our study provides further evidence of the magnitude of the health opportunity cost of health spending in South Africa, some limitations warrant further research to strengthen our findings. Our econometric strategy may not have fully accounted for all sources of endogeneity. Endogeneity may arise due to unobserved heterogeneity that affects both the level of health spending and death rates within the population and from reverse causality (Edoka, 2019). The fixed effect approach we adopted here, while accounting for a wide range of observable and unobservable heterogeneity between provinces, may not have fully accounted for endogeneity due to reverse causality. Reverse causality may arise when population

mortality rates from previous years affect the levels of resources allocated to the health sector. This can be seen as a form of omitted variable bias, such that (1) and (2) are mis-specified due to the omission of lagged death rates (e.g. M_{it-1} and M_{it-2}). Lagged death rates are likely to be correlated with both contemporaneous health spending and contemporaneous mortality, thereby introducing an attenuation bias to estimates of health spending elasticities. This has implications for the magnitude of the health opportunity costs and the cost-effectiveness threshold—an underestimation of health spending elasticities will result in an underestimation of the health opportunity cost (all else held constant) and an overestimation of the threshold. Therefore, the threshold estimated in this study may have been overstated. Future research should consider applying estimation strategies that accounts for the dynamic effect of mortality, such as the lagged dependent variable approach (Cameron and Trivedi, 2005). This approach will require sub-regional level data such as district- or municipal-level data.

Conclusion

As South Africa moves towards implementing universal health coverage reforms through NHI, health technology assessment and the use of cost-effectiveness analysis is expected to play an increasingly important role in informing the inclusion of new health interventions into the health benefit package (NDOH, 2017). While ethical considerations such as equity in the distribution of resources will also play a role in informing 'who gets what' within the public health system, cost-effectiveness analyses and the use of an appropriately set cost-effectiveness threshold will be crucial for ensuring efficiency in the allocation of scarce resources. This study estimates a cost-effectiveness threshold that explicitly takes into account trade-offs inherent within resource allocation decisions when health budgets are constrained. Our estimated cost per DALY averted threshold reflects the opportunity costs of health spending and is observed to be lower than the previously recommended one to three times per capita GDP decision rule[2].

By explicitly taking into account both the productivity of the health system and the size of available health budget, the framework for estimating supply-side thresholds presented here will allow the periodic updating of thresholds to reflect significant changes to the health budget, the productivity of the health system and other factors that may significantly affect the productivity of the health care system. However, further research is needed to refine estimates of the marginal productivity of the health system and to better understand the interplay between factors that determine the magnitude of the cost per DALY averted threshold.

Notes

1. YLD was obtained from the IHME Global Burden of Disease Study, which estimated the magnitude of health loss from diseases and injuries, disaggregated by age, gender and location. By applying disease/injury-specific disability weights to the number of individuals living with the disease/injury (Murray et al., 2012).
2. The proportion of GDP per capita reported here should not be interpreted as a guide to inform future thresholds that are based on GDP per capita estimates but is meant to highlight the difference between supply-side thresholds and previously suggested decision rules.

References

Airoldi M, Morton A. 2009. Adjusting life for quality or disability: stylistic difference or substantial dispute? *Health Economics* **18**: 1237–47.

Appleby J, Devlin N, Parkin D, Buxton M, Chalkidou K. 2009. Searching for cost effectiveness thresholds in the NHS. *Health Policy (Amsterdam, Netherlands)* **91**: 239–45.

Augustovski F, Colantonio LD, Galante J et al. 2017. Measuring the benefits of healthcare: DALYs and QALYs—does the choice of measure matter? A case study of two preventive interventions. *International Journal of Health Policy and Management* **7**: 120–36.

Bertram MY, Lauer JA, DE Joncheere K et al. 2016. Cost-effectiveness thresholds: pros and cons. *Bulletin of the World Health Organization* **94**: 925–30.

Bevan G, Hollinghurst S. 2003. Cost per quality-adjusted life year and disability-adjusted life years: the need for a new paradigm. *Expert Review of Pharmacoeconomics & Outcomes Research* **3**: 469–77.

Bokhari FA, Gai Y, Gottret P. 2007. Government health expenditures and health outcomes. *Health Economics* **16**: 257–73.

Cameron A, Trivedi P. 2005 *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

Claxton K, Martin S, Soares M et al. 2015. Methods for the estimation of the NICE cost effectiveness threshold. *Health Technology Assessment* **19**: 1.

Dorrington RE. 2013. Alternative South African Mid-Year Estimates, 2013. In: *Centre for Actuarial Research Monograph 13.* Centre for Actuarial Research, University of Cape Town.

Drummond M, Torrance G, Mason J. 1993. Cost-effectiveness league tables: more harm than good? *Social Science & Medicine* **37**: 33–40.

Edney LC, Haji Ali Afzali H, Cheng TC, Karnon J. 2018. Estimating the reference incremental cost-effectiveness ratio for the Australian Health System. *PharmacoEconomics* **36**: 239–52.

Edoka IP. 2019. Returns to health spending in low- and middle-income countries. In: *Oxford Research Encyclopedia of Economics and Finance.* Oxford: Oxford University Press.

Filmer D, Pritchett L. 1999. The impact of public spending on health: does money matter? *Social Science & Medicine* **49**: 1309–23.

Gafni A, Birch S. 2006. Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Social Science & Medicine* **62**: 2091–100.

Gallet CA, Doucouliagos H. 2017. The impact of healthcare spending on health outcomes: a meta-regression analysis. *Social Science & Medicine* **179**: 9–17.

Gold M, Siegel J, Russell L, Weinstein M. 1996. *Cost-Effectiveness in Health and Medicine.* Oxford, NY: Oxford University Press.

Gold MR, Stevenson D, Fryback DG. 2002. HALYs and QALYs and DALYs, Oh My: similarities and differences in summary measures of population health. *Annual Review of Public Health* **23**: 115–34.

Hutubessy R, Chisholm D, Edejer TT. 2003. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost Effectiveness and Resource Allocation* **1**: 8.

Johnson LF, Chiu C, Myer L et al. 2016. Prospects for HIV control in South Africa: a model-based analysis. *Global Health Action* **9**: 30314.

Leech AA, Kim DD, Cohen JT, Neumann PJ. 2018. Use and misuse of cost-effectiveness analysis thresholds in low- and middle-income countries: trends in cost-per-DALY studies. *Value in Health* **21**: 759–61.

Loganathan T, Ng C-W, Lee W-S, Hutubessy RCW, Verguet S, Jit M. 2018. Thresholds for decision-making: informing the cost-effectiveness and affordability of rotavirus vaccines in Malaysia. *Health Policy and Planning* **33**: 204–14.

Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. 2015. Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization* **93**: 118–24.

Murray CJL, Ezzati M, Flaxman AD et al. 2012. GBD 2010: design, definitions, and metrics. *The Lancet (London, England)* **380**: 2063–6.

National Treasury. 2016a. *National Budget Review Annexure: W1—Explanatory Memorandum to the Division of Revenue*. Pretoria, South Africa: National Treasury.

National Treasury. 2016b. *National Budget Review. Chapter 6: Division of Revenue and Spending by Provinces and Municipalities*. Pretoria, South Africa: National Treasury.

National Treasury. 2019. *Medium Term Expenditure Framework: Technical Guidelines*. Pretoria, South Africa: National Treasury.

NDOH. 2012. *Guidelines for Pharmacoeconomic Submissions*. Pretoria, South Africa: NDOH.

NDOH. 2017. *National Health Insurance for South Africa: Towards Universal Health Coverage (NHI White Paper)*. Pretoria, South Africa: NDOH.

Newall AT, Jit M, Hutubessy R. 2014. Are current cost-effectiveness thresholds for low- and middle-income countries useful? Examples from the world of vaccines. *PharmacoEconomics* 32: 525–31.

Ochalek J, Lomas J, Claxton K. 2018. Estimating health opportunity costs in low-income and middle-income countries: a novel approach and evidence from cross-country data. *BMJ Global Health* 3: e000964.

Robinson LA, Hammitt JK, Chang AY, Resch S. 2017. Understanding and improving the one and three times GDP per capita cost-effectiveness thresholds. *Health Policy and Planning* 32: 141–5.

Statistics South Africa. 2002–2015a. *Mid-Year Population Estimates South Africa*. Pretoria, South Africa. STATS SA

Statistics South Africa. 2002–2015b. *Mortality and Causes of Death in South Africa*. Pretoria, South Africa. STATS SA.

Thokala P, Ochalek J, Leech AA, Tong T. 2018. Cost-effectiveness thresholds: the past, the present and the future. *PharmacoEconomics* 36: 509–22.

Vallejo-Torres L, García-Lorenzo B, Castilla I *et al*. 2016. On the estimation of the cost-effectiveness threshold: why, what, how? *Value in Health* 19: 558–66.

Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. 2018. Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Economics* 27: 746–61.

Weinstein M, Zeckhauser R. 1973. Critical ratios and efficient allocation. *Journal of Public Economics* 2: 147–57.

Woods B, Revill P, Sculpher M, Claxton K. 2016. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value in Health* 19: 929–35.

World Health Organization. 2001. *Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health*. Geneva: World Health Organization.

Appendix

Table 1A Data inputs for estimating cost per DALY averted (2002–2015)

Year	Total Population (millions)	Total Spending (current ZAR, billions)	Total Spending (2015 ZAR, billions)	#Deaths (thousands)	#YLL (millions)	#YLD (millions)	Average CLE	YLD per capita	YLD:YLL ratio	Health Spending Elasticity
2002	45.9	33.2	66.9	509	27.52	5.28	30	0.115	0.192	−0.223
2003	46.4	37.4	71.3	563	31.26	5.41	29	0.116	0.173	−0.223
2004	47.0	40.5	76.1	583	32.88	5.54	29	0.118	0.169	−0.223
2005	47.6	47.0	85.4	600	33.91	5.73	29	0.120	0.169	−0.223
2006	48.2	53.6	93.2	617	34.46	5.78	29	0.120	0.168	−0.223
2007	48.9	62.6	101.5	612	34.07	5.82	29	0.119	0.171	−0.223
2008	49.6	75.5	109.8	604	33.22	5.87	29	0.118	0.177	−0.223
2009	50.3	87.7	119.0	582	31.96	5.93	30	0.118	0.185	−0.223
2010	51.0	98.0	127.5	552	30.33	6.04	30	0.118	0.199	−0.223
2011	51.7	111.3	138.0	512	28.28	6.03	31	0.117	0.213	−0.223
2012	52.5	122.5	143.7	487	26.12	6.05	31	0.115	0.232	−0.223
2013	53.3	130.7	145.0	467	24.19	6.10	32	0.114	0.252	−0.223
2014	54.1	140.9	147.3	462	22.82	6.16	32	0.114	0.270	−0.223
2015	55.0	154.1	154.1	469	21.83	6.22	33	0.113	0.285	−0.223

CLE = conditional life expectancy, YLL = years of life lost, YLD = years lived with disability, # = number.