

## Developing a rural index for district hospitals in South Africa

The purpose of this index is to rank South Africa's District Hospitals on a scale from most rural to most urban using a quantitative measure. Such a measure will allow us to compare the allocation of resources (e.g. healthcare workers or financial resources) or health care outputs and outcomes between District Hospitals based on their relative rurality.

## Development of the method used in constructing the index

This index is constructed using the method developed by the RHAP, in collaboration with the National Department of Health and researchers from the School of Public Health at the University of the Witwatersrand. The method is detailed in full in a chapter of the 2014/15 South African Health Review (Eagar *et al*, 2015).

## A framework for developing a functional definition of rural

As a starting point, we drew on the RHAPs review of approaches to developing functional definitions of rural for the purposes of policy and planning (RHAP, 2014). In this review we explored research and literature on approaches to defining rural for health care in order to develop a framework that could be used to develop a fit-for-purpose definition for the South African health care context. That framework identifies four categories of variables that can be used in developing a functional definition of rural and which, as we show later, can be used in developing a quantitative index of relative rurality. The four variable categories included in the framework include:

1. Measures of health need: these include utilisation, clinical and epidemiological measures or proxies of need such as measures of relative deprivation.
2. Measures of geographic accessibility: longer distances in rural areas not only make it more difficult and expensive to access services for rural patients, but also add to the cost of providing such services.
3. Population measures: rural areas tend to have smaller populations with a greater geographic spread, meaning lower population densities.
4. Policy measures: contextual factors relating to historically neglected and marginalised groups in rural settings (based on race or ethnicity) who may have high levels of unmet need.

## Data sources and the selection of variables

For each category discussed above, we selected variables that were readily available in reliable and publically accessible databases. We narrowed down those sources to ones that are likely to be used in decision making processes within the public health care system. In the end, all variables selected

were available in either the South African District Health Barometer (DHB) data files published online by the Health Systems Trust ([www.hst.org.za](http://www.hst.org.za)) or those sourced from Statistics South Africa (StatsSA) from their online portal.

From these sources the following four variables were selected:

1. **Deprivation index:** due to concerns about the quality of facility-based output data, concerns around different levels of utilisation (unmet need), and the dearth of epidemiological data consistently and reliably collected from all district hospitals across the country, we settled on relative deprivation as our proxy for need. We chose relative deprivation because it repeatedly been shown to be an effective proxy for health care need in instances where epidemiological data are insufficient or unreliable (Barnett *et al*, 2002).

The South African Index of Multiple Deprivation is broadly considered to be the most reliable measure of deprivation in South Africa. The SAIMD draws on data collected as part of the national census and periodic household surveys. Since the SAIMD uses census and household survey data, it covers the entirety of the South African population at the small area level (i.e. down to sub-districts). This meant that we could easily determine levels of relative deprivation for communities in the immediate vicinity of each district hospital.

The data from the SAIMD for sub-districts was drawn from the 2014/15 District Health Barometer data file, which provides a SAIMD for each health sub-district.

2. **Average distance to clinics from district hospitals:** district hospitals play an important role in supporting service provision at the clinic level through outreach, supply chain management and the referral of patients to higher levels of care. Longer average distances to clinics potentially increase the cost of providing support and the time needed for outreach. We worked on the assumption that average distance to clinics from the district hospital could also be used as a proxy indicator for population dispersion within a district hospital's geographic area of responsibility.

The average distance from a district hospital to clinics within its area of responsibility is not a variable that is provided in any publically available database or report. For the pilot version of our index we constructed this measure using the GPS coordinates for each facility, which we mapped using Google maps. Once facilities were mapped on our custom Google map, we could then calculate the shortest distance from each clinic to its referring district hospital by road

using the Google maps directions feature. We then averaged out all distances for each district hospital.

While this measure is useful as a standalone variable in the analysis of output and outcome variables, it is extremely time consuming to develop. Since this variable could be excluded from the construction of the pilot index (see method immediately below) it was excluded from the construction of our index here.

3. **Distance from each district hospital to the nearest regional hospital:** distance to the regional hospital was used for two reasons; firstly, district hospitals refer complex cases to regional facilities, such that longer distances have implications for administrative support, how care is managed and the kind of support that clinicians at rural facilities can expect from senior medical officers and specialists. Secondly, since regional hospitals are located in large towns or cities, this measure could be used as a proxy for distance from urban centres.

This measure was constructed in much the same way as the measure for average distance from clinics to the nearest district hospital. Using GPS data (longitude and latitude) for each district hospital and regional hospital in South Africa, which we obtained from the DHB 2014/15, we developed a custom Google map of these facilities. We then identified the nearest regional hospital to each district hospital using the directions function in Google maps.

It is important to note here that by nearest regional hospital we mean the nearest hospital in the same province as the district hospital is located. We took this decision because we suspect that, in most instances, district hospitals are likely to only link with regional facilities in their own province of authority even if there is a closer regional hospital located in another province. This must still be verified though.

As we discuss in more detail later, this approach has direct implications for where some district hospitals are located on our index. For example, there are several facilities located in the western parts of Mpumalanga that are closer to regional facilities in Gauteng than they are to regional facilities in their own province. The same is true for the Northern Cape, where district hospitals in the South Western parts of the province are closer to regional facilities in the Western Cape than they are to those located in their own province.

4. **Catchment population:** the size of the catchment population was used as our population measure in our pilot index. In some contexts, rural facilities are classified as small hospital facilities due to the comparatively small and sparse populations they serve. The per capita cost of providing services to a small catchment population is higher than that required for a large

population because these facilities do not benefit from economies of scale in terms of fixed overhead costs.

There is no publically available data for a district hospitals catchment population. We therefore estimated this figure using sub-district uninsured population data from the National Census as a proxy for the catchment population of the district hospital located in that sub-district. This is a problematic approach since some sub-districts have two district hospitals within their borders and some sub-districts did not contain a district hospital. What this meant is that where more than one district existed within a sub-district we had to average out the population across each district hospital, regardless of the respective capacity of each facility. Then, by excluding sub-districts without a district hospital, we inevitably underestimated the catchment populations of most facilities.

Problems with this variable mean that until more reliable catchment population data can be obtained, it should not be viewed as reliable in any analysis. Fortunately, when we developed our pilot index we found that this variable did not add value to the construction of our index and could be excluded.

### Constructing the index using our selected variables

Before constructing the index using our selected variables, each variable needed to be normalised. By normalised we mean converting each variable with a different unit measure into a uniform unit measure so that they are comparable. This is necessary in this instance because our index is an aggregate measure of combined variables. If we did not normalise the variables it is likely that a variable using a large scale would unduly bias the output of the calculation. By standardising our variables to the same scale, we can be sure that the influence each variable has on the output is not caused by bias introduced by differing unit measures. We do not want the deprivation index, which has a large range, to have undue influence over the index, while distance measures, which are small in comparison, are undervalued.

We normalised our values by calculating the Z-score for each observation within a variable. This is done by subtracting the mean of each variable from each observation and dividing the result from the standard deviation of the variable. The output of this calculation ensures that all our observations are in the range between +3 standard deviations and -3 Standard deviations from the mean, which is now zero.

The index was calculated by using Principal Component Analysis (PCA). PCA is a statistical method commonly used in testing and selecting variables in the construction of composite indices; the

Human Development Index and indices of relative deprivation are good examples of its application. PCA is used to test how variables are associated and change in relation to one another and is a useful approach in assessing the suitability of data and excluding variables that do not contribute meaningfully to the index. PCA allowed us to determine statistically what the best combination of selected variables was.

We started our analysis using all four variables. Our initial analysis found that the average distance from the district hospital to the clinic added little statistical value to the measure and so could be excluded.

We then analysed the construction of the index using the deprivation, distance to the closest regional hospital and catchment population variables. In this instance we found that there was good statistical evidence for the use of all three variables in the index.

One of the limitations of the three variable index was that the catchment population variable was likely to have a confounding effect if we used it in an analysis with input, output or outcome measures that included population as a component. For instance, the three variable index could not be used to compare per capita expenditure measures between facilities because both measures included a population component.

We also found the catchment population data to be the least reliable of the three measures due to difficulty in ascertaining with any accuracy if captured figures were true reflections of the size of the population that actually made use of each facility (discussed earlier).

We therefore also tested the construction of the index using only two variables: the distance to the closest regional hospital and the relative deprivation indices for the sub-district where the facility is located. We found that there was no significant difference in the statistical fit between the three and two variable index. This meant that the two variable index was as reliable as the three variable index and so could be used with confidence in any analysis.

For the purposes of this index we therefore use the two variable method in its construction.

One of the limitations of our approach is that we did not assign a weight to either variable and so assumed they held equal importance in the output of the index. This means that in our index the deprivation variable and distance to closest regional hospital variable held the same importance. For example, it may be that deprivation has a greater influence on outcome variables and should carry a greater weight in our analysis. It may be worthwhile to consider weighting variables in future. This would require careful consideration of objective and subjective factors in allocating weights though and is beyond the scope of this index.

## Continuous versus categorical classification of facilities as rural or urban

We considered classifying hospitals categorically into deep rural, rural, peri-urban and urban facilities for the purposes of analysis. When looking at the output of the index, however, there were no clear cut-off points in the index scores to meaningfully support four categories, and we felt that any such separation would be artificial and would make analysis using the index unreliable.

This means that facilities are not classified as urban or rural but rather in terms of their position relative to the most rural facility or the most urban facility

## Interpreting the rural index

After having decided on the variables to be used in the index, the two index variables nominal values—along with other key identifying characteristics including its location (province, district and sub-district)—were captured in an excel spreadsheet for each facility. The index variables were then normalised and aggregated into a rural index score for each facility. All the pertinent information, including the facility catchment population variable (for reference) were captured in the format indicated in figure 1 below.

The spreadsheet is formatted in such a way that the index can be sorted according to any of the categories identified by column headings. This is done by clicking on the column heading tab for the desired sort category (e.g. sort by province or sort by rural index score). In the example in figure 4 the data is sorted by highest rural index score to lowest rural index score (the image only includes the 8 most rural facilities on the index).

Province	District	Sub-district	Facility	Nearest Provincial/ Regional Hospital	Distance to nearest Provincial/Regional Hospital	Dep Index	Catchment Pop	Rural Index
Eastern Cape	A Nzo DM	Umzimvubu SD	Greenville Hosp	Mthatha General Hospital	262	3998.029	123557	4.07
Western Cape	Central Karoo DM	Beaufort West LM	Murraysburg Hosp	George Hospital	432	1716.341	25213	3.88
KwaZulu-Natal	Umkhanyakude DM	Umhlabuyalingana LM	Manguzi Hosp	Lower Umfolozi War Memorial Hospit	258	3877.931	80434	3.87
Mpumalanga	Nkangala DM	Thembsile Hani LM	KwaMhlanga Hosp	Philadelphia Hospital	400	2105.127	327145	3.87
Mpumalanga	Nkangala DM	Dr JS Moroka LM	Mmamethlake Hosp	Philadelphia Hospital	380	2245.943	261781	3.74
Northern Cape	Namakwa DM	Hantam LM	Calvinia Hosp	Dr Harry Surtie Hospital	402	1899.719	22052	3.66
KwaZulu-Natal	Umkhanyakude DM	Jozini LM	Mosvold Hosp	Lower Umfolozi War Memorial Hospit	244	3724.163	95793	3.49
Eastern Cape	Joe Gqabi DM	Elundini SD	T Bequest Hosp (Elu)	Mthatha	231	3689.157	69448	3.26

Figure 1 Layout of rural index: most rural facilities

For the purposes of ranking according to relative rurality, the most important column is the rural index column on the far left. In figure 1 it is sorted according to most rural to least rural and in figure 5 below it is sorted most urban to least urban.

Province	District	Sub-district	Facility	Nearest Provincial/ Regional Hospital	Distance to nearest Provincial/Regional Hospital	Dep Index	Catchment Pop	Rural Index
Gauteng	Tshwane MM	Tshwane 3 SD	Tshwane Dist Hosp	Steve Biko Academic Hospital	3	580.0161	312659	-3.65
Gauteng	Johannesburg MM	Johannesburg F SD	South Rand Hosp	Chris Hani Baragwanath Hospital	15	441.8133	639390	-3.63
Gauteng	Tshwane MM	Tshwane 3 SD	Pretoria West Hosp	Steve Biko Academic Hospital	4	580.0161	312659	-3.63
Western Cape	Cape Town MM	CT Southern SD	Victoria Hosp	Groote Schuur Hospital	11	573.0961	298210	-3.54
Western Cape	Cape Town MM	CT Tygerberg SD	K Bremer Hosp	Groote Schuur Hospital	21	675.2678	630263	-3.28
Western Cape	Cape Town MM	CT Southern SD	False Bay Hosp	Groote Schuur Hospital	30	573.0961	298210	-3.26
Western Cape	Cape Town MM	CT Eastern SD	Eerste River Hosp	Tygerberg Hospital	21	714.8003	266944	-3.23
Free State	Mangaung MM	Bloemfontein SD	National Dis Hosp	Universitas Hospital	5	970.0866	527926	-3.16

Figure 2 Layout of rural index: Most urban facilities

On the index there is no cut-off point for deep rural, rural, peri-urban and urban facilities. Instead, the index should be interpreted as facilities being ranked according to their position relative to the most rural end of the scale and the most urban end of the scale. On our index, Greenville Hospital in the Umzimvubu sub-district of the Eastern Cape has the highest rural index score of 4.07. On the other end of the scale, Tshwane district hospital has the lowest score of -3.65 on our rural index scale and could be interpreted as the most urban facility.

The rankings on the scale do, for the most part, intuitively make sense if one has a good idea of where many of the facilities are located in reality. The facilities at the most rural end of the scale all tend to be located in small towns or villages a long distance from large towns or cities. Those that fall in the middle of the scale, with scores between -1 and 1 on the rural index all tend to be located in larger towns but are still some distance from large cities and metros.

Caution must be used when reading the index though. In some instances facilities rank closer to either end of the scale because on one of the two measures they have scores outside the normal range. An example here is Murraysburg District hospital in the Western Cape, which is ranked as the second most rural facility on our scale. It receives its ranking because it is situated a significant distance from the nearest regional hospital (430km vs SA mean of 110 km). This facility is, however, located in a district with a relative deprivation score of 1716, which is significantly below the SA mean of 2374 (the higher the score the more deprived the sub-district).

Another example of where discretion in interpretation should be used is for all facilities in the Nyandeni sub-district (Canzibe, Isilimela, St Barnibus and Bambisana) These facilities are all ranked within the middle portion of the index despite being located in one of the most economically under developed sub-districts in the country. Their position on the ranking is largely because of their proximity to Mthatha hospital. The distance measure used here, however, does not account for the absence of transport infrastructure and difficulties likely confronting access to Mthatha from the sub-district.

If we consider the distance measure more closely we find that the decision to use distance to the nearest regional hospital in the same province, instead of the distance to the nearest regional hospital regardless of province, to have an influencing effect on ranking. The best examples here are the KwaMhlanga and Mmametlhake district hospitals in Mpumalanga. These facilities are amongst the most rural facilities on the index largely because they are located 400km and 380 km from the nearest regional hospital in the province respectively. While their deprivation scores certainly indicate a high degree of rurality, their final rural index score changes significantly if we measure distance to the nearest regional hospital regardless of province. At 75km and 42km respectively, both facilities are far closer to the Steve Biko Academic Hospital in Pretoria, than the nearest regional hospital in Mpumalanga. If we use the Steve Biko Academic Hospital as the referring facility, then KwaMhlanga and Mmametlhake's rural index scores drop from 3.87 and 3.74 respectively to -0.83 and -1.15 respectively. This change moves them from the most rural end of the index to the middle of the index. The solution here is to decide on what basis distance should be measured and then ensure this limitation is made clear to anyone using the index.

While these outliers are the exception rather than the norm, it may be worth considering adding a weighting to the deprivation measure to increase its contribution to the rural index. There is no obvious value for this weighting though and this would need to be a subjective judgement call.

## References

Barnett S, Roderick P, Martin D, Diamond I, Wrigley H. Interrelations between three proxies of health care need at the small area level: an urban/rural comparison. *J Epidemiology and Community Health*. 2002;56(10):754–61.

Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. *Am J Public Health*. 2005;95(7):1149.

Vermaak K. Definition of rural in policy and health research. Unpublished manuscript; 2006.

Humphreys JS. Delimiting “rural”: implications of an agreed “rurality” index for health care planning and resource allocation. *Aust J Rural Health*. 1998;6(4):212–6.

McGrail MR, Humphreys JS. Measuring spatial accessibility to primary care in rural areas: improving the effectiveness of the two-step floating catchment area method. *Applied Geography*. 2009;29(4):533–41.

Penno E, Gauld R, Audas R. How are population-based funding formulae for health care composed? A comparative analysis of seven models. *BMC Health Serv Res*. 2013;13(1):470.