South African COVID-19 Vulnerability Index

Technical Report

stats sa
Department:
Statistics South Africa
REPUBLIC OF SOUTH AFRICA

Table of Contents

Table of Contents 2
1 Background 3
1.1 The Vulnerability Index in summary 4
2 Methodology 5
2.1 Methodological considerations 5
2.2 The Alkire-Foster method 5
3 Constructing the Vulnerability Index 6
3.1 Indicator selection methodology 6
3.2 Vulnerability cut-offs and weights 8
3.3 Computational aspects 9
3.4 Data source and Indicators 9
4 Mapping the Vulnerability Index 10
4.1 Data area output 10
4.2 Spatially linking the data 10
4.3 Developing the dashboard 11
5 Using the Vulnerability Index 13
5.1 Example 1: Municipal level analysis 13
5.2 Example 2: Placename level analysis 15
6 Conclusion 17
6.1 Limitations 17
7 References 18
Appendix 1: Indicators and code ranges 19
Appendix 2: Application of the Alkire-Foster Method 21

1 Background

Since mid-March 2020, following President Cyril Ramaphosa's declaration of national disaster measures to contain the spread of the COVID-19 virus, government interventions have been directed to ensure containment of the spread of the Coronovirus, and subsequently to direct strategies to protect vulnerable communities.

The Vulnerability Index (VIndex) was developed as a way of using the smallest area population data available and incorporating dimensions and indicators to statistically reflect localised risk factors that may contribute to the spread of COVID-19 for each individual against a set criteria being measured.

The index uses a natural counting approach to determine the **headcount** of vulnerable individuals within a population as well as measure their level of simultaneous vulnerability determining **intensity**.

'Vulnerability' in this context refers to the demographic and socioeconomic factors that affect the resilience of individuals and communities, and represent those more likely to be adversely affected when COVID-19 manifests itself. COVID-19 disproportionately affects the elderly and those with underlying health conditions, and the population at risk includes those exposed to increased health and social vulnerability. For example those who:

- are poor or live in deprived conditions which impacts on health and sanitation;
- live in crowded areas or informal settlements which impacts on social distancing; and
- live in multi-generational households and large extended families in a single dwelling.

The index aims to identify these populations that are at multiple risk to identify areas where the population are considered most vulnerable to COVID-19. The index used the Alkire-Foster (AF) method as the technique allows an informed vulnerability approach incorporating indicators under different dimensions and indicators which can be mapped via a small-area map-based interface.

The South African COVID-19 Vulnerability Index (SA CVI) aims to statistically and spatially identify vulnerable populations that are more likely to be adversely affected should COVID-19 spread in the affected area.

The Index uses Census 2011 data and eight indicators related to labour force activity, access to media, household services, overcrowding, multigenerational status, age and chronic illness to output a composite indicator to reflect the vulnerability status of citizens to the COVID-19 virus and where these individuals are located.

Index indicators

1.1 The Vulnerability Index in summary

Rationale	The Vulnerability Index considers eight indicators related to labour force activity, access to household services, overcrowding, multigenerational status, age and chronic illness to output a composite indicator to show the vulnerability status of citizens to the COVID-19 virus and where these individuals are located.
Coverage	The index considers the population in all Enumeration Areas (EA) as defined by Stats SA, normally used for Censuses and statistical surveys. *While it is the lowest level of reporting for Stats SA, the data remains anonymous and are aggregated to the total population within the specific EA.*
Data Source	The index uses the 2011 Census data produced by Statistics South Africa. While this data may not be current, it is the only data that reports at EA level and allows for a more granular picture of where the most vulnerable of populations are located.
Indicators used	Eight indicators were used for the index, focussing on labour force activity, access to services, and household status. These indicators are as follows: • Employment without ownership of a car • No access to any of the following: internet, radio, television • No access to water within 200m of the dwelling • No access to flush and chemical toilets • Overcrowding status of the household • Multigenerational households • Elderly (60+) and • Use of chronic medication
Method	An adaptation of the Alkire-Foster method, which is used to produce National Poverty Indices, was used. The index uses a natural counting approach to determine the headcount of vulnerable individuals within a population as well as measure their level of simultaneous vulnerability, thereby determining intensity. A composite index was calculated and produced based on 8 indicators. All indicators have equal weights and hence each have a share of 1/8 contributed to the total weight of 1. The vulnerable were identified using a vulnerability cut-off, which represents the proportion of minimum deprivation a person must experience in order to be identified as vulnerable. The intensity of the vulnerability was then computed based on the number of themes/indicators that an individual is vulnerable in. The Vulnerability Index (VIndex) score was then finally computed, which is the headcount multiplied by the intensity. The final score was then ranked from 0-5, with 5 being the most vulnerable, 0 the least.
Limitations	The index makes use of 2011 Enumeration Areas and 2011 Census data; this data is almost 10 years old and there have been many changes in living circumstances, population movements and growth, which is not reflected. For the Health themes, data sources are not available at EA level and a proxy indicator for use of Chronic Medication was considered for the health dimension.
Use	VIndex can serve as a tool to assist with targeted response planning as it highlights areas with high vulnerability to COVID, as defined by the indicators that are most prevalent.

2 Methodology

The methodology used to calculate the index was compiled according to Alkire-Foster Methodology. This approach was used as it serves as a global and general framework for identifying individuals who are multi-dimensionally poor. For the Vulnerability Index, it was adapted for identifying individuals who are multi-dimensionally vulnerable to the Coronavirus pandemic.

2.1 Methodological considerations

In order to identify the populations that are at multiple risk, the approach used to derive the South African Multidimensional Poverty Index (SAMPI) was adopted. The SAMPI is based on the Alkire-Foster Method, and determines a person's experience of deprivation through several factors such as poor education and health, lack of decent work, inadequate access to services and lack of assets (Stats SA, 2014).

The SAMPI assesses poverty at an individual/household level, and an individual is identified as poor if they are deprived in a third or more of eleven indicators captured under four equally weighted themes. Further to this, the intensity of their poverty is determined based on the proportion of deprivations (third or more) they are experiencing.

The multidimensional approach seeks to integrate a range of indicators to capture the complex nature of vulnerability. As indicated in the introduction, COVID-19 disproportionately affects the elderly and those with underlying health conditions who live in deprived circumstances. As a result, the index aims to integrate indicators related to age, health, access to services, potential exposure by travelling to work, and lack of awareness to determine the number of individuals who are vulnerable. The SAMPI methodology recommends that indicators used may vary but should ultimately fit the purpose of the analysis or the situation on the ground.

2.2 The Alkire-Foster method

The Alkire-Foster (AF) method serves as a general framework for identifying individuals who are multidimensionally poor. It essentially uses a natural counting approach to determine the headcount of poor individuals within a population, as well as measure their level of simultaneous deprivations. The framework is also useful for other phenomena and users can hence adapt the general approach with their own specifications according to the objective of the index computed (UNDP, 2019).

The AF method considers a population (n), with the number of dimensions ($d \geq 2$) to form an achievement matrix $n \times d$, where a typical entry y_{ij} denotes the achievement of an individual $i = 1 \cdots n$ in a dimension $j = 1 \ldots d$. In other words, each row in the vector represents the achievement of an individual i, while each column represents the distribution of dimension achievements j across the number of individuals. The cut-off under which an individual is considered deprived in a dimension j is denoted as z_j. The row vector z then represents the deprivation cut-offs specific to each dimension. For any vector v, $|v|$ represents the sum of all the elements in vector v, while $\mu(v)$ represents the mean of the vector v (Alkire and Foster, 2009).

The deprivation status of an individual is determined through an identification function $\rho(y_{ij}, z)$ of the individual y_{ij} and cut-off z through a dichotomous output, where $\rho(y_{ij}, z) = 1$ indicates that a person is deprived/poor and $\rho(y_{ij}, z) = 0$ indicating that the person is not deprived. Application of this function to each of the achievement vectors in y produces the set $Z \subseteq \{1, \cdots n\}$ of persons who are deprived in y given z. Application of the identification function ρ then produces a result of multidimensional deprivation for the matrix y and the cut-off z resulting in a functional relationship of $M(y; z)$ also known as an index of multidimensional poverty (Alkire and Foster, 2009).

3 Constructing the Vulnerability Index

3.1 Indicator selection methodology

Due to the limited set of data values/variables, a series of consultations and research was done to determine the indicators to be used in calculating the VIndex centred on circumstances and risk factors that makes users most vulnerable to COVID-19. Four themes or domains were selected as follows:

Population, Household Services, Household Composition and Health.

Each theme contains relevant indicators as indicated in Figure 1. Indicators and vulnerability cut-offs were selected as suitable for the production of the Index. Some indicators were also selected as proxy indicators due to lack of data for preferred indicators.

Figure 1: Vulnerability Index themes and indicators

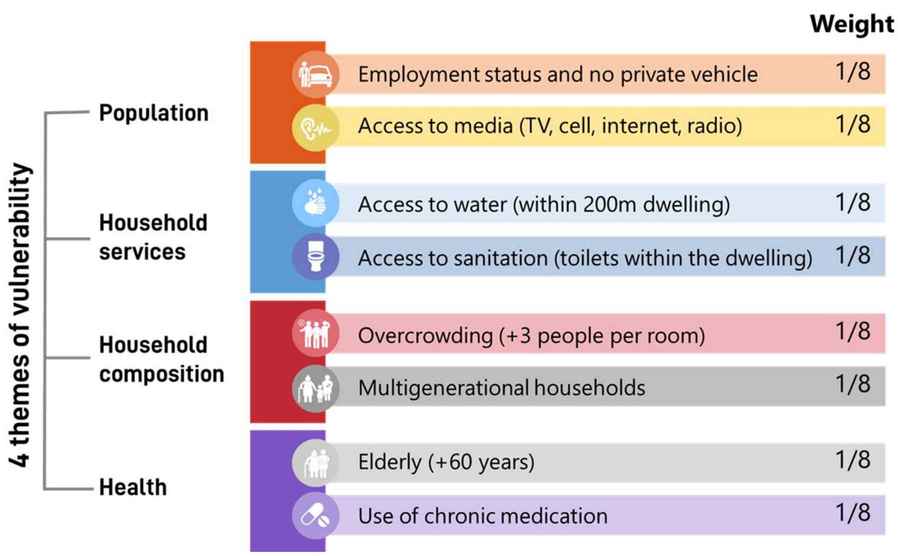


Figure 2: Vulnerability indicator cut offs

Table 1 outlines in detail the indicators and rationale for the selection of indicators, which included identifying older adults, persons with underlying or serious medical conditions, households with overcrowding, multi-generational households, access to media and access to private vehicles.

Table 1: Indicators and rationale for inclusion in the Vulnerability Index

Employed and access to private vehicle: Employment without ownership of a car Serves as a proxy to determine the likelihood of an individual making use of public transport or car-pooling as a means of transport to travel to work, thereby increasing the risk of exposure. Having to travel to work on a daily basis with reliance on shared or public transport exposes individuals to others not only while making use of the public transport. but also while waiting in queues
Access to Media: No access to any of internet, radio, television and cellphone Access to information plays a critical role in ensuring that households understand the importance of social distancing, personal hygiene and status of infections in a specific location.
Access to water: No access to water within 200m of the dwelling Captures population with access to water and relates to preventative measures individuals have access to, determining how levels of hygiene can be maintained through washing of hands. This is especially important when households may be overcrowded.
Access to sanitation: No access to flush and chemical toilets Captures population with access to adequate sanitation. Access to sanitation is also particularly important if toilet facilities are being shared within the household, or making use of public facilities. Sharing of toilet facilities limits the effectiveness of social distancing and preventative hygiene.
Overcrowding: Overcrowding household status Captures population at risk for spreading the virus due to living in overcrowded housing. This indicator was used as overcrowded households poses a health risk should someone in the household be affected. Sharing of rooms also means that self-isolation would be a challenge.
Multigenerational households: Extended household members Captures vulnerable population living in households which have both children and elderly living as extended families. Elderly living in households with children increases the risk of infection as children are inadvertent carriers of the virus and less alert in terms of preventative measures that should be maintained. Also, the likelihood of children being infected while at school is high.
Age: Elderly (60+) Captures population at risk as 80% of COVID-19 fatalities are among people over the age of 60. Older persons are therefore more susceptible to contracting the virus due to several underlying conditions, thus posing a risk of infection.
Chronic medication: Use of chronic medication Serves as a proxy to determine the likelihood of individuals having comorbidities, but this variable cannot consider the specific type of comorbidity in relation to determining COVID-specific comorbidities.

For two indicators, due to lack of data, proxy indicators were used i.e. car ownership and use of chronic medication. The variable on “car ownership” serves as a proxy to determine the likelihood of an individual making use of public transport or car-pooling as a means of transport to travel to work. The indicator therefore does not consider those who would be walking to work, as it does not solely focus on transport used to travel to work. Similarly, the indicator on “use of chronic medication” was used as a proxy to determine the likelihood of individuals having comorbidities. This indicator would therefore not consider the type of comorbidity as it is not reported with the intended purpose of determining COVID-specific comorbidities.

3.2 Vulnerability cut-offs and weights

According to the Alkire-Foster Methodology, all selected indicators are to be computed using the same unit of identification. This therefore indicates that indicators that do not have the same unit of identification have to be transformed to accommodate the unit of identification (UNDP 2019). As a result, all household indicators considered for the Vulnerability Index are reported at an individual level, where all persons inside a household are assigned the same result derived at a household level.
Table 2 outlines the indicators and cut-offs used to construct the index.

Table 2: Indicators[1] and vulnerability cut-offs for the Vulnerability Index

Theme	Indicator	Vulnerability cut-off
Population dynamics	Employment status and car ownership	If any person was employed and does not own a car
	Media (TV, Radio, Cellphone, Internet)	If any person did not own any one of television, radio, cellphone and has no access to internet
Household service dynamics	Access to water	If any person did not have access to piped water within 200m of the dwelling
	Access to sanitation	If any person did not have access to flush/chemical toilet in dwelling
Household status dynamics	Multigenerational households	Children (0-15 years*) and elderly (60+) in household
	Overcrowding	More than 3 persons per functional room
Health	Age	If any person was 60 years and older
	Chronic medication	If any person used Chronic Medication

**Children 0 – 15 years was used since the compulsory schooling age is up to 15 years. The official definition for children is 0 – 14 years.*

(Santos, 2019) indicates that the procedure for weighting of indicators can be done based on either statistical (which includes Principal Component Analysis or Factor Analysis) or normative approaches. The use of statistical approaches often limits the ability to compare such indices over time, as the weighting are limited to the data set used. The use of normative approaches are used to set weights based on the value that the indicator brings to the Index, as long as such a weight is justifiable. The normative approach has been the dominant approach thus far, as it is transparent and stable across units and over time. All National MPIs, including SAMPI has made use of this approach.

(Santos, 2019) also advises that in some cases participatory studies (to determine the contribution or value each dimension brings to the Index) are considered as alternative measures for setting weights although these methods have not yet been considered when constructing MPIs. Perhaps this weighting approach can be considered with the expansion of the Vulnerability Index going forward as it was already used for the selection of indicators.

Like the SAMPI, the Index follows a nested weighting approach. The Vulnerability Index used equal indicators across the themes and hence all indicator weights are equal across the themes. The weights are therefore as follows in Table 3:

[1] See Appendix 1 for variables and code ranges

Table 3: Indicators and weights

Theme	Indicator	Weight
Population dynamics (1/4)	Employment status and car ownership	1/8
	Media (TV, Radio, Cellphone, Internet)	1/8
Household service dynamics (1/4)	Access to water	1/8
	Access to sanitation	1/8
Household status dynamics (1/4)	Multigenerational households	1/8
	Overcrowding	1/8
Health (1/4)	Age	1/8
	Chronic medication	1/8

3.3 Computational aspects

A composite index was calculated and produced based on the 8 indicators. The intensity of the vulnerability was then computed based on the number of themes/indicators that an individual is vulnerable in. Headcount is defined as the proportion of persons that are considered to be multi-dimensionally vulnerable. Intensity is defined as the average proportion of indicators in which multi-dimensionally vulnerable persons are deemed vulnerable. These values are combined to calculate the final VIndex.

The Vulnerability Index (VIndex) score was computed as the headcount multiplied by the intensity. For the VIndex a person is considered vulnerable if the following cut-off criteria is not met:

- Unemployed and had access to one of the following media (tv, cellphone, radio, internet)
- Had access to water (within 200m from the dwelling) and has sanitation (Flush or Chemical Toilet)
- Did not reside in a household with both children (0-15) and elderly (60+) and has less than 3 persons per functional room
- The individual was younger than 60 and does not make use of chronic medication

For the VIndex all indicators have equal weights and hence each have a share of 1/8 contributed to the total weight of 1. A vulnerability score was computed for each person, defined as the weighted sum of deprivations. Using this score the vulnerable are identified using a second cut-off, the vulnerability cut-off, which represents the proportion of minimum deprivation a person must experience in order to be identified as vulnerable.

The Alkire-Foster method prescribes a cut-off (k) across dimensions/indicators, which translates to the weighted proportion of deprivations (in this case vulnerability) that the individual (unit of measurement) needs to experience to be considered vulnerable (Santos, 2019). For the Vulnerability index, a cut-off of k = 0.25 is considered. An individual is then defined as vulnerable[2] if they are deprived in at least two (2) of the eight (8) indicators or one full theme. All individuals who are only vulnerable in one indicator will be reduced to not being vulnerable in any indicators.

3.4 Data source and Indicators

Census data is the default data source given its ability to disaggregate information to the lowest possible geographic area. Using variables found in Census 2011, indicators can be generated at person level. Eleven (11) variables[3] reported in the Census were used to contribute to the index. Two (2) and four (4) indicators were combined to derive composite indicators for employment and lack of access to media respectively. For the employment indicator the two indicators used are employment status of an individual and ownership of a

[2] See Appendix 2 for Example using Alkire-Foster Method
[3] See Appendix 1 for Census Variables used.

motorcar. For the lack of access to media, the four variables access to television, radio, cellphone and internet were combined. The age variable are used to represent two indicators, one for multigenerational households and another for elderly. The resultant number of indicators for the index are therefore eight (8).

As previously indicated, all variables used are to be reported at the same unit of identification. For this reason, all individuals in a household were assigned the same result derived at a household level in terms of the variables access to water, access to sanitation, multigenerational households and overcrowding.

4 Mapping the Vulnerability Index

The mapping approach used to visualise the index was designed to be an interactive dashboard, allowing the user to visualise the Vulnerability Index by selecting different geographic levels (municipality, placename). Due to the small area nature of the index, static maps would not yield the geographical resolution required and interpretation would be limited, therefore a web-based dashboard interface was chosen.

4.1 Data area output

The Index makes use of Enumeration Area (EA) boundaries, as the smallest geographic statistical unit available, as main rationale for the index was to statistically identify and map vulnerable populations at the lowest possible geographic level. While placename or ward boundaries may have been a more appropriate level to visualise the index from a user perspective, they are not designed to be demographically homogenous and do not have optimum population sizes, especially when mapped thematically. While thematic maps give a good visualisation of differing values over space, they have disadvantages due to the challenge that statistical variations over a large area can be hidden as they are generalised over the entire area; for this reason, smaller units are better than large ones.

EA boundaries were therefore selected as the data output layer. EA level data cannot be published due to confidentiality concerns, but they can be mapped and presented with a ranked number and understood in the context of a placename, which will be essential to guide dissemination of the final product. The aim was for the user to be able to select the geographic boundary by municipality and placename, and the mapped results would show the spatial variations of vulnerability amongst EAs within the placename. The data was therefore deliberately not aggregated to reflect the smaller area spatial variation within a placename. While it is mapped at EA level, it is queried at placename level as the user is familiar with this selection. Using the tooltips, the user can further explore specific regions of interest. Tooltip text is displayed when the mouse pointer hovers over an area, and allows the user to get more information in relation to the VIndex score, headcount and intensity contributions to the final score, the main and sub-placename, and the settlement type of the area.

4.2 Spatially linking the data

The final Index file with 96,462 EAs with Headcount, Intensity and VIndex value was linked to the 2016 Municipal boundary EA master file of 103,576 EAs, which covers the entire country. A final file with the complete list of EAs for the country was created, including attribute data for municipalities, placenames and ward ID. A total of 7,114 EA records not reflecting data were assigned a 0 value on the assumption these are unoccupied.

For the final map classification of the VIndex, values from 1-5 were used on the basis of defined intervals (see Table 4). When mapping the SAMPI, only the headcount output is mapped as this is easier to communicate to stakeholders and for them to interpret. However, for SACVI the Vindex score (or Adjusted Headcount Ratio), the product of the headcount and the intensity was used and ranked as the priority is not knowing "how many" but rather "where is the risk higher or lower"? The VIndex is therefore mapped because it indicates the incidence of vulnerability, adjusted by the intensity, since high intensity means that people are vulnerable in more indicators and hence more susceptible to the risk. The Vindex score hence provides a clear picture for targeting overall vulnerability without necessarily focussing only on the total number of persons that are vulnerable, but also the intensity of that vulnerability. To add more value to interpretation the headcount and intensity values can be seen in tandem to the final index score through the use of the tooltips in the dashboard to disaggregate the final index.

Defined breaks were used for the classification to help the user differentiate from Highest – Lowest in terms of vulnerability. This ranked EA value was then used to thematically map on a colour graduated scale of 0-5. The classification was based as follows:

Table 4: Defined interval classifications

Classification	Value	Description
0,0 - 0,09	0	This result indicates an adjusted headcount ratio between 0 and 0,09, which indicates a headcount less than or equal to 36% (36% of persons in the EA are vulnerable) and an intensity level between 0,25 and 0,5 (vulnerable in up to four indicators)
0,1 - 0,19	1	This result indicates an adjusted headcount ratio between 0,1 and 0,19, which indicates a headcount between 25% and 76% and an intensity level between 0,25 and 0,50 (vulnerable in up to four indicators)
0,2 - 0,29	2	This result indicates an adjusted headcount ratio between 0,2 and 0,29, which indicates a headcount greater than 40% and an intensity level between 0,29 and 0,50 (vulnerable in up to five indicators)
0,3 - 0,39	3	This result indicates an adjusted headcount ratio between 0,3 and 0,39, which indicates a headcount greater than 64% and an intensity level between 0,25 and 0,50 (vulnerable in up to four indicators) *The difference between this and the previous class relates to the resultant ratio composition of the headcount and intensity.*
0,4 - 0,49	4	This result indicates an adjusted headcount ratio between 0,4 and 0,49, which indicates a headcount greater and equal to 98% and an intensity level between 0,49 and 0,66 (vulnerable in between four and five indicators on average)
0,5 +	5	This result indicates an adjusted headcount ratio greater than or equal to 0,5, which indicates a headcount greater and equal to 98% and an intensity level between 0,49 and 0,66 (vulnerable in between four and five indicators on average) *The difference between this and the previous class relates to the resultant ratio composition of the headcount and intensity.*

*since the intensity is based on the vulnerability cut-off of 2 or more indicators, the intensity will never be less than 0,25.

4.3 Developing the dashboard

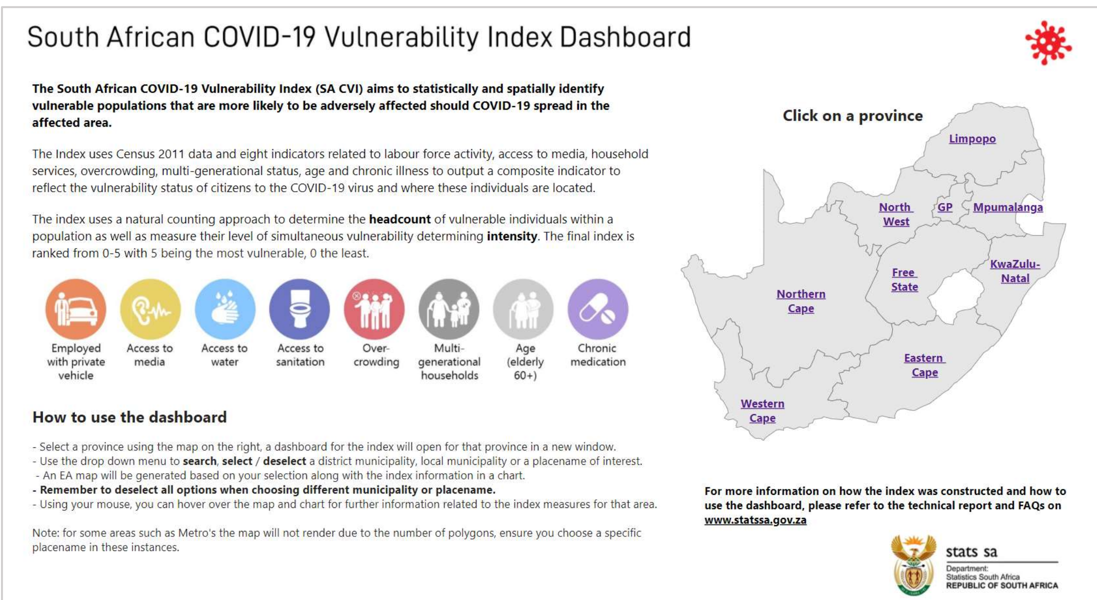
A static map product would not allow for effective data analysis at such a small spatial area of EA or placename, therefore an interactive dashboard was chosen as an effective way of delivering the final product for optimal data visualisation. Dashboard functionality allows a user to select and unselect with an auto-zoom function, providing an efficient way to explore data in relation to geographic features based on location from placename to municipality. This gives the end user the ability to visualise data on a very local basis.

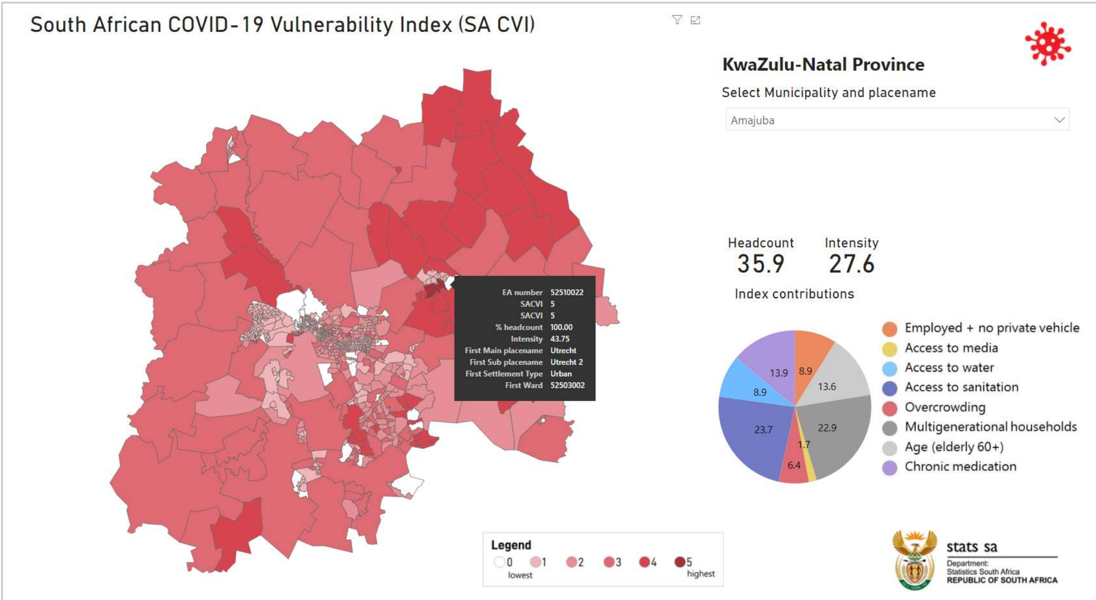
Mapping of the index at EA level and design of dashboards and dropdown menus were developed and tested. Due to the time and cost of development of a purpose-specific system, an off-the shelf solution from PowerBI was chosen using shapemaps. The 2011 EA boundaries were converted into a TopoJSON file and Power BI desktops shape map tool visual was selected to compare regions on a map using color. The limitation of this tool is that it cannot render more than 5,000 polygons – this affects the map at provincial and some metropolitan areas, but is resolved when lower geographic resolution is selected.

A dashboard for each province was designed with dropdown filters for Province, District, Local municipality and Placename to all drill-down selection. The spatial file was imported and linked it to the final index and mapped values 1-5 coded. The dashboard functionality was designed to allow a user to select and unselect with an auto-zoom function. As indicated by hovering on the map you be able to see the sub-placename, Ward detail and index, headcount and intensity. A landing home page allows the user to select the province of interest, and details how to use the dashboard with links to the technical report and FAQs.

Figure 3 shows examples of the data rendered at municipal level. Using the drop down menu, the required municipality and placename can be selected / deselected, and toggled using the arrows. The map will automatically zoom to the selected area of interest. By using the mouse and hovering over the map, information will appear related to that area in a tooltip. Using a combination of the final index, the headcount totals, intensity values as well as indicator contributions allows the user to understand the final index value at a very nuanced level. Additional information on placename, ward and settlement type also allows the user to understand the area in the context of the different spatial hierarchy.

Figure 3: Screenshots of data mapped in the dashboard





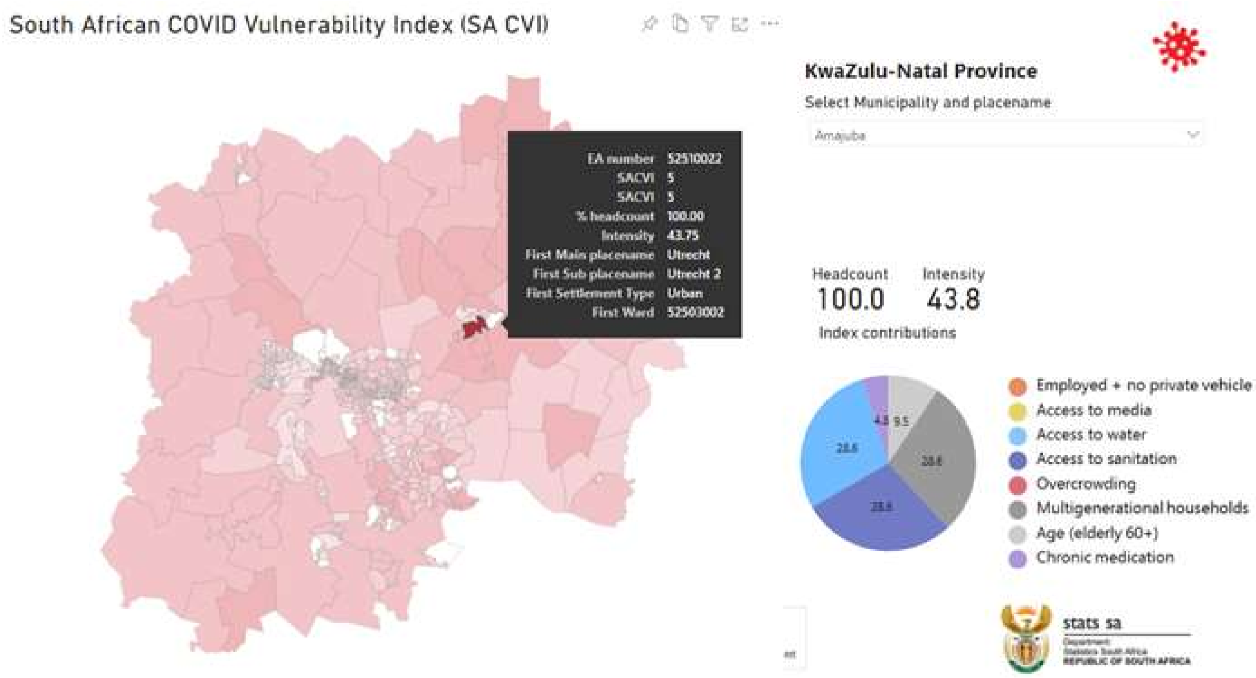
5 Using the Vulnerability Index

The VIndex can serve as a tool to assist with targeted response planning as it highlights areas which are vulnerable to COVID as defined by the rank and indicators that are most prevalent. The VIndex can also serve as means for evidence-based pandemic management planning to provide the best and most feasible location-focused response. There are various ways a user could approach the tool – looking first at municipal level, for example, and then identifying highly vulnerable areas, or examining a specific locality or placename in detail. These examples are detailed below using the KwaZulu-Natal province. Using the dashboard selection tools the headcount (the proportion of households that are considered to be multi-dimensionally vulnerable), intensity (the average proportion of indicators in which multi-dimensionally vulnerable households are deemed vulnerable) and index contributions linked to a specific area can be displayed.

5.1 Example 1: Municipal level analysis

This provides a fairly high level analysis. The user will select the district or local municipality and then, using the mouse, can hover over areas of interest to get tool tip information, or by selecting a polygon the contributions to the index can be established at a glance. Figure 4 shows this type of analysis – when the chosen EA is selected (as shown in this example from Utrecht), the headcount, intensity and contribution chart will display for that EA. It can be seen from the tooltip that this is an urban area. The selected EA has a high Vindex ranking at 5, the headcount is 100% which indicates that all persons residing in the EA were vulnerable to some extent and the intensity is 43,8% meaning that on average these persons were vulnerable in 43,8% of the eight indicators used to compile the index. The pie chart shows which of the indicators contribute the most to the vulnerability. It can also be seen that the biggest contributor to vulnerability was access to water and sanitation and access to media (28,6% respectively).

Figure 4: Municipal level analysis



5.2 Example 2: Placename level analysis

This provides a lower level analysis. The user will select the district or local municipality and then the placename of interest. Using the example of Matabetule in eThekwini district, 6 EAs were selected each with a different ranked score based on the Vindex to use as an example of how to use the Index. Figure 5 shows this type of analysis – the selected EAs in the placename Matabetule in eThekwini district area selection are shown in Figure 5 below, highlighting the different EAs and their index values. Different EAs can be selected to compare the different contributors.

Figure 5: Matabetule mapped example of VIndex

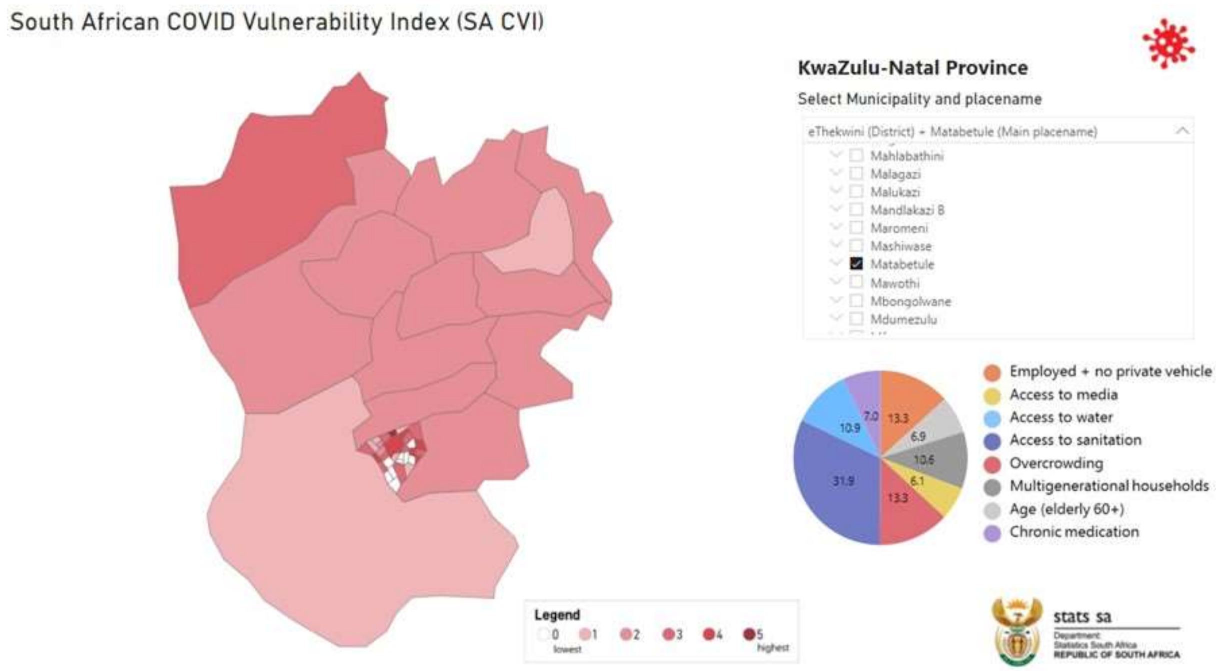


Table 5 shows attribute data (the same data which is presented graphically in the different pie charts and tooltips) which shows the percent headcount and intensity for a variety of EAs with different rankings in the placename. EA 59915049 has a high VIndex ranking at 5, the headcount is 100% which indicates that all persons residing in the EA were vulnerable to some extent and the intensity is 44%, meaning that, on average, these persons were vulnerable in 44% (which translates to, on average, 3,55 indicators) of the eight indicators used to compile the composite indicator.

EA59915042 has a low VIndex ranking at 1, the headcount is 20% which indicates that on average 20% of persons residing in the EA were vulnerable, and the intensity is 28% meaning that on average these persons were vulnerable in 28% of the indicators.

Table 5: Matabetule example of VIndex

EA number	Sub placename	Main placename	District	Ward	Settlement Type	% headcount	Intensity	VIndex	Rank
59915049	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	100,00	43,75	0,44	5
59915039	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	68,22	45,55	0,31	4
59915032	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	65,70	38,99	0,26	3
59915038	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	48,03	28,24	0,14	2
59915042	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	20,30	25,00	0,05	1
59915033	Lower Manaza	Matabetule	eThekwini	59500003	Traditional	0,00	0,00	0,00	0

Table 6 shows the attribute data in terms of the indicator contributions to the final score that is displayed graphically in the dashboard. For EA59915049, ranked at 5 on the VIndex, the variables for chronic medication, sanitation and access to media are the largest contributions to the index value (29% respectively). This information could assist with response planning which focuses on advocacy programmes in areas where lack of media is a high contributor to the index result. For EA59915032, ranked at 3, the main contributor to the VIndex was over-crowding and sanitation. This information could assist pandemic management.

For EA59915042, ranked at 2 on the VIndex, sanitation is the main contributing variable at 50% along with no access to private vehicle.

Table 6: Matabetule example of VIndex index contributions

EA number	VIndex	Rank	Age	Employed	Chronic	Water	Sanitation	Media	Over-crowding	Multigen
59915049	0,44	5	14,28	0,00	28,57	0,00	28,57	28,57	0,00	0,00
59915039	0,31	4	9,88	0,00	5,94	21,53	1,96	21,53	19,56	19,56
59915032	0,26	3	1,76	1,76	17,12	0,00	32,06	0,00	30,29	16,98
59915038	0,14	2	8,27	6,20	26,33	0,00	38,39	0,00	0,00	20,78
59915042	0,05	1	0,00	50,00	0,00	0,00	50,00	0,00	0,00	0,00
59915033	0,00	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

6 Conclusion

The aim of the SA-CVI is to measure and map at the lowest geographic level possible the vulnerability of persons to COVID-19 based on risk factor identification to determine the extent to which persons or places are likely to be affected within the context that data limitations provide. The Vulnerability Index strength is that it measures simultaneous dimensions of vulnerability, the aim to statistically identify vulnerable populations that are more likely to be affected by the pandemic in terms of who will suffer most if a COVID-19 infection occurred in that specific area.

6.1 Limitations

There are a number of limitations in relation to the final index constructed which are briefly outlined below in relation to statistical and spatial data, and the methodology of index construction:

Statistical data for index

Census 2011, which is almost 10 years old, was used as the only available data source. There have been many changes in living circumstances, population changes, movements and growth. It was the only source of data that would allow analysis at the lowest spatial resolution. For some of the themes, data sources were not available at EA level (e.g. no health data, age groups could not be computed as variables were not available). To overcome this proxy indicators were used (e.g. the proxy indicator for use of Chronic Medication was considered for the health dimension). Processing and coding 44 million records for the different variables took many hours, crashing the system and sometimes having to be run overnight.

Spatial data for index

Many of the EA classifications from 2011 will have changed over the last decade – for example, residential and housing developments as informal settlement changes will not be reflected in the data, some EAs are classified as unoccupied, EA classifications do not match the population numbers reflected in the data, for example, a farm EA with 2 persons, a Vacant EA with 2,000 persons. EAs with populations less than 50 with EA classifications of vacant or industrial or parks and recreation and farms also have "skewed" some of the final results and should be interpreted as such.

Methodology for index

There are many criticisms to the approach of aggregating a group of indicators into a single index and it was a challenge to determine the final measures. Although it is multidimensional, the index cannot take everything that is important into account. The index measures intensity (how many simultaneous indicators of vulnerability are experienced by an individual), but there is no measure of the "depth" of each vulnerability. Refinements were done and the index was recalculated several times to refine the index as much as possible, but data was a significant limitation.

The choice of weights for such an index is inherently problematic. The approach of weighted averaged was discussed and index method proposed using headcounts took some time to refine and develop the methodology to apply in this context. Compiling an index of this nature, which specifically relates to issues such as accessibility to water and sanitation services, household crowding status and lack of assets, means the results will be more skewed towards more rural areas where service delivery is lacking as opposed to other urban areas where service delivery is high.

While current trends in terms of COVID-19 cases may counter some of these results, it should be noted that the VIndex does not aim to predict the number of cases or identify possible hotspots, but serves as an identifier of areas where COVID-19 would have the most adverse effect should it become widespread in such an area, and some of the interventions that can be considered to mitigate its impact on the most vulnerable.

7 References

Alkire, S. and Foster, J. (2009) Counting and multidimensional poverty measurement. Journal of Public Economics, 95(7-8):476–487.

Stats SA. (2014) The South African MPI. Creating a multidimensional poverty index using census data. Statistics South Africa, 2014.

UNDP. (2019) How to Build a National Multidimensional Poverty Index (MPI): Using the MPIs to inform the SDGs. United Nations Development Programme (UNDP) and Oxford Poverty and Human Development Initiative (OPHI), University of Oxford, 2019

M.E. Santos. (2019) Challenges in designing national multidimensional poverty measures, Statistics series, No. 100 (LC/TS.2019/5), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2019.

De Kadt, J., Gotz, G., Hamann, C., Maree, G., and Parker, A. (2020) Mapping vulnerability to COVID-19 in Gauteng. GCRO Map of the Month, Gauteng City-Region Observatory, March 2020. https://www.gcro.ac.za/outputs/map-of-the-month/detail/mapping-vulnerability-to-covid-19/

Wolkin, A., J. Rees Patterson, S. Harris, E. Soler, S. Burrer, M. McGeehin, S. Greene. Reducing Public Health Risk During Disasters: Identifying Social Vulnerabilities. Journal of Homeland Security and Emergency Management. Volume 12, Issue 4, Pages 809–822. 2015.

Flanagan, B., E. Gregory, E. Hallisey, J. Heitgerd, and B. Lewis, 2011. A Social Vulnerability Index for Disaster Management. Journal of Homeland Security and Emergency Management 8(1). https://svi.cdc.gov/Documents/Data/2016_SVI_Data/SVI2016Documentation.pdf

Johns Hopkins University Center for Systems Science and Engineering COVID-19 Global Cases https://statescoop.com/5-ways-maps-can-help-communities-respond-to-covid-19/

World Health Organization Coronavirus Information https://coronavirus-disasterresponse.hub.arcgis.com/

Smith, S (2020) Managing Health and Economic Priorities as the COVID-19 Pandemic Spreads through Africa Spotlight, Africa Center for Strategic Studies, March 30, 2020.

Office for National Statistics COVID roundup https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19roundup/2020-03-26

Paris21 Addressing COVID-19: How are Statistical Offices doing? https://zoom.us/webinar/register/WN_ha9FP8FlR2y7vbJRKLNeUw

DeCaprio, D., Gartner, J.A., Burgess, T., Kothari, S., Sayed, S., McCall, C.t (2020) Building a COVID-19 Vulnerability Index doi: https://doi.org/10.1101/2020.03.16.20036723

Appendix 1: Indicators and code ranges

Census Variable	Codes
f02_age	**Valid value range:** Age: 000 to 120 in single years.
p13e_devmedchronic	All persons aged five years and older in households 1 = Yes 2 = No 3 = Do not know 9 = Unspecified . = Not applicable (institutional population and transients)

derp_employ_status_official	Derived from P23A_EMPLOYMENTSTATUS, 23A_EMPLOYMENTSTATUS,
	23A_EMPLOYMENTSTATUS,
	P24_TEMPWORKABSENCE, P25_UNEMPLOYMENT,
	P26_UNEMPLOYMENT, P27_REASONSNOTWORKING and P28_WORKAVAILABILITY
	
	status = 0;
	if P23A_EMPLOYMENTSTATUS = 1 or P23B_EMPLOYMENTSTATUS = 1 or
	P23C_EMPLOYMENTSTATUS =
	1 or P24_TEMPWORKABSENCE = 1, then status = 1;
	if status = 0 and P25_UNEMPLOYMENT = 1 and P28_WORKAVAILABILITY = 1, then status = 2;
	if status = 0 and P26_UNEMPLOYMENT = 1 and P28_WORKAVAILABILITY = 1 and
	P27_REASONSNOTWORKING = ('8', '10', '11'), then status = 3;
	if status = 0, then status = 4
	
	Final code list
	1 = Employed
	2 = Unemployed
	3 = Discouraged work-seeker
	4 = Other not economically active
	5 = Age less than 15 years
	6 = Eligible age but no response to employment questions
	9 = Unspecified
	. = Not applicable (persons younger than fifteen years, institutional population and transients)

H13_MOTORCAR	**Final code list** 1 = Yes 2 = No 9 = Unspecified
H07_WATERPIPED	**Final code list** 1 = Piped (tap) water inside dwelling/institution 2 = Piped (tap) water inside yard 3 = Piped (tap) water on community stand: distance less than 200m from dwelling/institution 4 = Piped (tap) water on community stand: distance between 200m and 500m from dwelling/institution 5 = Piped (tap) water on community stand: distance between 500m and 1000m (1km) from dwelling/ institution 6 = Piped (tap) water on community stand: distance greater than 1000m (1km) from dwelling/institution 7 = No access to piped (tap) water 9 = Unspecified
H10_TOILET	**Final code list** 1 = Flush toilet (connected to sewerage system) 2 = Flush toilet (with septic tank) 3 = Chemical toilet 4 = Pit toilet with ventilation (VIP) 5 = Pit toilet without ventilation 6 = Bucket toilet 7 = Other 0 = None 9 = Unspecified . = Not applicable (Collective living quarters)

Census Variable	Codes
H03_TOTROOMS DERH_HSIZE	H03_TOTROOMS **Excludes bathrooms and kitchens and includes garages used as rooms** **Valid value range:** 1 to 25 = Number of rooms . = Not applicable (collective living quarters) DERH_HSIZE The derived variable has the following valid values 001:998 number of persons 999 or more persons. For all households, if the household has between 001 and 998 person records, then assign the derived variable HHSIZE to the number of person records; otherwise (999 or more person records found), assign derived variable HHSIZE = 999. **if total persons more than 3 per functional room then overcrowded**
H13_TV H13_RADIO H13_CELLPHONE H13A_INTERNET	**Final code list** 1 = Yes 2 = No 9 = Unspecified No Media formula if H13_TV and H13_RADIO and H13_CELLPHONE and H13A_INTERNET = 5 then Nomedia = 1 else Nomedia = 0

Appendix 2: Application of the Alkire-Foster Method

Suppose one is interested in analysing the vulnerability of a 3-person population considering the seven indicators: Age, Employment Status (ES), media (Media), access to water (WA), access to sanitation (SA), chronic medication (CM) and overcrowding (OC) in the household. The 3x7 matrix X indicates the achievements of the three persons in the seven indicators. A person is considered to be deprived in an indicator if the criteria of the cut-off for that indicator is not met. The cut-off vector is denoted as z and the cut-offs are as follows:

- Age = <60 years
- Use of Chronic Medication = No
- Employment status = Not employed
- Access to Water = Access within 200m of the dwelling
- Access to sanitation = Access to flush or chemical toilet
- Media = Yes
- Overcrowding = No

Achievement matrix

Age CM ES WA SA Media OC

$$X = \begin{bmatrix} 61 & Yes & Not\ employed & Water & No\ toilet & No & Yes \\ 50 & No & Employed & No\ Water & No\ toilet & Yes & Yes \\ 20 & No & Employed & Water & Toilet & Yes & No \end{bmatrix} = \begin{matrix} Person\ 1 \\ Person\ 2 \\ Person\ 3 \end{matrix}$$

Vector *z* determines who is vulnerable and who is not.

$$z = [< 60 \quad No \quad Not\ Employed \quad Water \quad Toilet \quad Yes \quad No \ ]$$

The vulnerability matrix g^0 is constructed by comparing each individual person's achievements to the vulnerability cut-off vector where a score of 1 is allocated to the indicator if the person is vulnerable and a score of 0 is allocated otherwise.

Vulnerability matrix

Age CM ES WA SA No media OC

$$g^0 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} = \begin{matrix} Person\ 1 \\ Person\ 2 \\ Person\ 3 \end{matrix}$$

Vector *w* denotes the weight vector and all indicators are weighted equally as follows:

$$w = [0.14 \quad 0.14 \quad 0.14 \quad 0.14 \quad 0.14 \quad 0.14 \quad 0.14]$$

The weights are then applied to the vulnerability matrix g^0 to obtain the weighted vulnerability matrix. The vulnerability score (c_i) for each person is computed based on the sum of the weighted scores.

Weighted vulnerability matrix

Age CM ES WA SA No media OC Vulnerability score c_i

$$\overline{g_0} = \begin{bmatrix} 0.14 & 0.14 & 0 & 0 & 0.14 & 0 & 0.14 \\ 0 & 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0 & 0 & 0.14 & 0 & 0 & 0.14 & 0 \end{bmatrix} = \begin{matrix} 0.57 \\ 0.71 \\ 0.29 \end{matrix}$$

Based on the above matrix, who will be considered as vulnerable in this 3-person population? Vulnerability is determined by a cut-off, denoted by *k*. For example if k = 0.4285 or 42,9%, then a person is vulnerable if vulnerable in 3 or more of the weighted indicators (i.e. if the Vulnerability score c_i is 0.4285 and above). This is the intermediate value approach for the cut-off. If cut-off is set to consider the person vulnerable when satisfying 1 indicator, it is known as the union approach and if the cut-off is set to consider all indicators before considering the person vulnerable, it is known as the intersection approach. All indices to date have made use of the intermediate value approach.

Once the intermediate cut-off has been applied, the weighted vulnerability matrix is censored (not applicable in the union approach) to focus only on persons who were identified as vulnerable and the weighted values for all persons who are non-vulnerable (in this case $k<0.4285$) is set to 0.

Censored vulnerability matrix

Age CM ES WA SA No media OC V score $c_i(k)$

$$g_0(k = 0.43) = \begin{bmatrix} 0.14 & 0.14 & 0 & 0 & 0.14 & 0 & 0.14 \\ 0 & 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0 & 0 & 0.14 & 0 & 0 & 0.14 & 0 \end{bmatrix} = \begin{matrix} 0.57 \\ 0.71 \\ 0 \end{matrix}$$

Headcount

The headcount ratio *H* is then computed as the proportion of people who are vulnerable, which in this case is 2 of the 3-person population as indicated in the matrix above. That is H = 2/3. This means that 66,7% of the people are multi-dimensionally vulnerable.

Intensity

The intensity A is the average share among those who are vulnerable which in this example will be the average of v score $c_i(k)$ of person 1 and person 2. That is (0.57+0.71) / 2. Hence A = 0.64= 64%. This therefore indicates that vulnerable individuals (i.e. person 1 and 2) are, on average, vulnerable in 64% of the weighted indicators.*

Index

Finally, the Index is obtained through the product of the headcount (H) and Intensity (A). That is HxA = 0.667x0.64 = 0.427. This hence indicates that multi-dimensionally vulnerable individuals (persons 1 and 2) experience 42,7% of the total vulnerability that would be experienced had all the people in the population (persons 1, 2 and 3) been deprived in all the indicators at the same time.