Post Covid-19 health commodity shortage tracking using the Maisha Meds network and Indian export data
Authors: Daniel Rosen, Odette Melvin and Florence Wambui
September 2020

1.0 Summary

Introduction
The COVID-19 pandemic has triggered a number of health commodity supply chain challenges that have led to healthcare commodity shortages and price hikes. This paper aims to describe how Maisha Meds has used a combination of international customs trade data and retailer point of sale (POS) data to develop and test a prototype of a medicine shortage flagging tool, to identify potential essential medicine shortages across sub-Saharan Africa.

Methods
Maisha Meds used export data from India, import data from Kenya and Point of Sale data from Kenyan pharmacies to identify anomalies in the supply of medicines, view price changes and look for leading indicators of potential national level medicine shortages across sub-Saharan Africa. To make the trade data useful it first had to be transformed from semi-structured text, based on a free text data entry, to a structured data set. This is referred to in this methodology as the cleaning of the data.

Analysis 1
Export data from India to sub-Saharan Africa was cleaned to identify which medicines had experienced a rapid drop in exports after February 2020, indicative of a supply interruption resulting from Covid-19. Monthly exports are run through an ARIMA forecasting model looking at smoothed exports value from Jan 2017 – June 2020 (log of 2 month moving average), which flagged anomalies as those outside of a 95% prediction interval.

Anomalies showing significant drops in export values, based on a forecast of expected values are labelled as being at high risk of shortage. Anomalies for finished pharmaceuticals were identified based on value, not volume, and did not differentiate between different forms of medicine (tablets, injections etc...). Anomaly flagging for API exports was more accurate as export records did not include brand names and had a common form (KG). Medicine shortage risk flagging was performed for exports to all of SSA, exports to SSA excluding South Africa, as well as exports to Kenya specifically.

Maisha Meds sought to validate the shortages flagging work by comparing results of our flagging model with shortages as reported by private sector wholesalers working in SSA, that were interviewed as part of the wider project with CGD, conducted by IQVIA in August 2020.
Analysis 2

Maisha Meds looked to validate the significance of changes to Indian export data by establishing a connection between historic shortage flags in Indian exports and increases in the price of medicines in Kenyan pharmacies. Maisha Meds analysed data captured from ~250 pharmacies in Kenya, using the company’s point of sale system. Using a Chow Test with a Trend Break Analysis for the products in the Maisha Meds Pharmacies it was possible to identify 254 products that had seen a significant change in price. From the products identified only 64 had a significant price increase, and only 15 of these had a significant enough reliance on India (>50% based on Kenyan import data), and enough data in both the export data and the facility data to be analysed.

Results

Analysis 1

Using the ARIMA forecast model Maisha Meds identified 116 out of 525 finished pharmaceutical products that experienced a significant drop in exports from India to all of Sub-Saharan Africa ex. South Africa after Feb 2020 with a total of 121 anomalous months. This was an increase from 34 / 525 products experiencing a significant drop in exports indicative of a shortage from October 2019 – Feb 2020, before Covid-19’s impact on supply chains.

Interviews of wholesalers in Nigeria, Kenya, Ghana, Jordan and Pakistan identified 39/62 molecules as being in shortage in at least one country, of these Maisha Meds had insufficient data to analyse 2/39 molecules. The Maisha Meds ARIMA forecast model was able to flag 18/37 molecules as at risk of shortage to a 95% confidence interval. In the instances where there was sufficient API data the ARIMA model successfully flagged 11/13 products as at risk of shortage that were confirmed by local wholesalers (or 15/18 including the products highlighted as in shortage not included on the original list of tracer molecules).

Analysis 2

Maisha Meds also conducted an analysis to link historic drops in Indian exports to major price increases in Kenyan pharmacy prices, in an effort to prove a link between change in exports and change in prices for patients in Africa. In 8/15 molecules, Maisha Meds was able to identify a statistically significant change in the value of exports from India to Kenya, using an 80% confidence interval, prior to a trend break price increase in Kenya. Looking at the products that exhibited a relationship between exports and price increase in Kenya, a negative correlation between price and export value was observed with a lag of 4-5 months in half of the cases, in particular for products with a higher import dependence on India. This would indicate that with refinement and improved data inputs it might be possible to identify a shortage before it affects medicine supply in a country.

Discussion

As an early stage prototype to show what might be possible in building a medicine shortage flagging tool, this analysis had a large number of constraints including resource availability, issues in the accuracy and breadth of the data cleaning algorithm, lack of differentiation of medicines by form, lack of consistent methodology, incomplete data inputs, lack of granularity on smaller products and difficulties analysing individual country trends using trade data.

With these caveats in mind, Maisha Meds believes that the results were encouraging for an early stage prototype. From 37 shortages identified by wholesalers the model was able to flag 18 as being at risk of shortage based on a 95% confidence interval. Furthermore, the analysis was able to show
examples of a link between interruptions in supply from India and an increase in pharmacy prices in Kenya on average 4-5 months later, for a small number of products. This would indicate that it is possible to design an early warning system that would actually give countries enough time to act on a potential shortage.

The analysis was able to show the impact of Covid on the supply of certain essential medicines to the region. Furthermore, by going into greater depth on the data on price per standard unit it was possible to differentiate between global shortages and temporary supply disruptions such as the one caused by the Indian export ban of March 2020.

Accuracy of data capture and inconsistent capture of information on pack size, medicine strength and medicine form represents the largest risk to the viability of an early warning medicine shortages database. Most other issues identified in this analysis would be possible to overcome with refinement of data cleaning and processing, but the quality of data inputs with respect to these components would require improvements in the trade portals where this data is captured.

Conclusion

With more export and import data inputs it theoretically should be possible to build an early warning system to predict when a medicine is at risk of a shortage for all countries in SSA. This would require investment, increased availability of export data particularly from the EU and China, and an effort by customs officials in the longer term to improve the quality of data capture on pharmaceuticals. This is a particular issue for medicine pack size, strength and form where the data suffers from significant issues in the Indian data set.

While it is possible to build a warning system that uses data from pharmacies, central medical stores or even wholesalers in a country, often by the time you see a shortage in this data it is going to be too late to respond to it. Import and export data, particularly if you look further upstream to trade in API and key starting materials (KSM) may be more likely to provide actionable insights with respect to predicting medicine shortages before they happen.

For a system like this to be effective it would also likely need to sit under a global public health body such as the WHO to be able to aggregate data, disseminate results and centralise expertise for policy response.
2.0 Introduction

In addition to a multitude of social and economic turmoil created by the Covid-19 virus, the disease has also had a significant impact on the global health commodity supply chain. Covid-19 has led to export bans, a slowdown in the manufacture of active pharmaceutical ingredient (API) in China and logistic disruptions in India that culminated in disruptions to exports in March and April 2020.

Sub-Saharan Africa was particularly at risk for essential medicine supply shortage because of the lack of price elasticity in the population and the likelihood of procurement agents being outbid by more developed, richer countries. In addition, the relatively nascent local manufacturing industry has no API manufacturing capabilities and is unlikely to have stockpiled significant reserves of API due to capital constraints.

Using export data from India Maisha Meds showed that at an absolute value level the continent was mostly affected by the fall in India exports. Exports from India to the SSA region in March and April 2020 were around US$60mn lower each month than the average of the previous six months, though exports rebounded to near pre-Covid levels in May and June. Exports of API to local manufacturers in Africa also fell substantially to almost half of pre-Covid levels in April 2020 before also rebounding.

Figure 1: Change in finished pharmaceutical monthly exports from India to each global region Jan 2017 – June 2020

Source: Indian export data

Having seen medicine shortages and price surges in the pharmacies within the Maisha Meds network the organisation wanted to see if it was possible to use our data science capabilities, existing data assets and external data assets to predict when medicine shortages might occur in Kenya, and across the continent. Looking beyond the immediate impact of Covid, the aim was also to understand if such a system could represent a global public good for future development.

Figure 2: Change in supplier prices and rate of item stockouts in Maisha Meds wholesaler suppliers
3.0 Methods and Data Sources

3.1 Data sources

This project used three main data sources: Indian export data, Kenyan import data and Maisha Meds point of sale data. In addition to this Maisha Meds was also able to benefit from the work done by IQVIA in interviewing local pharmaceutical wholesalers.

1. Indian export records for HS codes 30 and 29 were acquired from a data vendor. Export data for HS code 29 was only for Jan 2019 – June 2020. Export data for all other codes was for Jan 2017 – June 2020. Data was filtered to look only at exports to sub-Saharan Africa.

2. Kenyan import records for HS codes 30 and 29 were acquired from a data vendor. Import data was only acquired for the full year of 2019.

3. Maisha Meds is a technology enabled healthcare company with the goal of improving access to affordable and high-quality medicines across the global south. Our software is live in over 300 pharmacies and clinics in Kenya, Uganda, Tanzania, Ghana, and Nigeria, where healthcare providers support over 130,000 patients each month using our systems. Data was used from ~250 pharmacies in Kenya to understand how global trade affected local supply.

4. Products at risk of shortage were validated through wholesaler interviews in a small number of cases. Interviews were conducted by IQVIA as part of their work also looking into shortages of medicines in LIC / LMICs.

3.1 Methodology

The methodology for this project broke down into data cleaning and labelling, export anomaly identification, further investigation of anomalies and analysis of the lag between exports and medicine price change in Kenya.

1. International Nonproprietary Name (INN) labelling: International trade data is captured with a series of fields, intended to help regulators and customs officials to monitor the flow of goods and the appropriate taxation levels for products. Shipment manifests are uploaded at the point of import and export to a portal that captures a series of fields such as Date, Item Description, HS Code, Quantity, Units, Price per Unit, Country / Port of Origin, Country / Port of Destination. Because HS Codes for pharmaceuticals fail to capture enough detail on each product the most important field for the purpose of this analysis is the Item Description. Unfortunately Item Description is an open text field which must be manually entered, and as such is prone to misspellings, omissions and errors.

Figure 3: A hypothetical example of a typical shipment’s data is below.

<table>
<thead>
<tr>
<th>Date</th>
<th>Item_description</th>
<th>Quantity</th>
<th>Unit Type</th>
<th>unit_value_in_int</th>
<th>export_destination</th>
<th>HS Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/06/2019</td>
<td>ARTEMETHER 20MG LUME FANTRINE 120MG 30X6X4 BLISTER PACK</td>
<td>246150</td>
<td>BOX</td>
<td>1256.88</td>
<td>KENYA</td>
<td>30046000</td>
</tr>
</tbody>
</table>

From this example the first piece of information that is needed is the INNs present in the medicines shipment, in this case Artemether and Lumefantrine. However, as one can see Lumefantrine is misspelled, as such it is necessary to also take account of potential misspellings. In another shipment the INN may not be given, just the brand name, so for these
one must map brand names to INN. When building the library of INN and brand name search terms, the analysis started with the highest value shipments and worked down.

Data cleaning was performed using Python and a string search looking for medicines INN names, brand names and misspelled INNs. In the event that the algorithm was able to tag INN names (misspelled or otherwise) it would firstly remove any duplicates then arrange INNs alphabetically. In the event that no INN names were available the algorithm would use brand name and then cross reference these to the molecule contents of the brand

The same methodology was applied to unfinished API exports under trade code 29 and to Kenyan imports of finished and unfinished pharmaceuticals. By looking at the proportion of each molecule that was imported to Kenya from India as a percentage of total value it was possible to calculate Kenya’s relative dependence on India per molecule.

2. **Export anomaly identification:** Having mapped the exports to INN molecules and molecule combinations next Maisha Meds created a historic time series to look for significant changes in export patterns, both historically and those following the start of the Covid-19 outbreak. Export data by its nature is quite erratic, volumes and values can rise and fall significantly month-on-month.

To control for these effects Maisha Meds used an autoregressive integrated moving average (ARIMA) model to forecast for expected values based on the time series using a two month rolling average of exports per molecule, to account for export volatility, on a logarithmic scale to stabilize the variance of the series for forecasting. Anomalies were identified using a 95% confidence interval.

Products were only included for analysis if they had less than 5 zero values in the two month moving average and more than 21 months with finished pharmaceuticals export data and more than 9 months for API export data.

3. **Analysis of the lag between exports and medicine price change in Kenya:** Maisha Meds analysed data captured from 250 pharmacies in Kenya, using an android based point of sale system. First this analysis identified the medicines that, since Jan 2018, have experienced a rapid increase in price, indicative of a supply shortage. Based on the experience of analyzing pharmacy data in Kenya after Covid-19, Maisha Meds found rapid price changes to be the most reliable indicator of shortage in our dataset as it was less dependent on seasonal factors affecting volume fluctuations.

Using a Chow Test with a Trend Break Analysis for the products in the Maisha Meds pharmacy network it was possible to identify 254 products that had seen a significant trend breaks in price. From the products identified only 64 had a significant price increase, and only 15 of these had enough reliance on India (>50%), and enough Indian exports and Point of Sale data at the facility level, to be analyzed.

Having identified those products that had experienced a trend break in weekly average price Maisha Meds next compared these to the instances where Indian exports to Kenya for each molecule had experienced an anomaly, as defined by our ARIMA model. Lastly, because of the lag between medicine shipments from India for the supply chain to deliver medicines to
pharmacies in Kenya Maisha Meds analyzed the lag time between a change in Indian exports and a shift in the price of pharmaceuticals in Kenya.

This analysis was only done for finished pharmaceutical exports from India to Kenya as there were not enough months of data for unfinished exports to build a meaningful time series.

4.0 Results

Similarly to the methodology section, the results of the project breaks down into according to the same four core analyses.

1. **International Non-proprietary Name (INN) labelling:** The trade data labelling by INN molecule was successful in labelling US$545mn out of $597mn (91%) of all unfinished API exports from India to every country in sub-Saharan Africa by value. The trade data labelling by INN and Brand was successful in labelling US$7.75bn out of US$8.66bn (90%) of all finished pharmaceutical exports from India to every country in sub-Saharan Africa by value. The coverage of data labelling could have been improved with more time and resources. However, the analysis was attempting to attribute INNs to over 800,000 rows of data with open text fields, so achieving 100% coverage became a matter of diminishing returns the smaller the shipment rows become in value.

The accuracy of the labelling system was not as good as might have been hoped for, in particular for combination therapies. This became more apparent when the data was analyzed on a row-by-row basis in the later analyses and mistakes in the labelling system were identified and corrected for. The main issue was when a combination medicine had one drug name spelt correctly and the other misspelled, these typically showed up as a monotherapy of only the initial drug. For example losartan hydrochlorothiazide may be written such that the Losartan is correctly spelt but hydrochlorothiazide is not. As such it would appear as a monotherapy.

These issues were common across the dataset, though they could be corrected for with more resource. Unfortunately the team did not measure the number of incorrect rows, but for products commonly found in combination therapies an estimate of 5-10% of rows being mislabelled seems likely.

2. **Export anomaly identification:** The analysis ran the processed export records through the ARIMA forecast model, excluding any product with more than 5 zero values in the dataset, looking for anomalies with a 95% confidence interval. Importantly when looking at multiple countries in aggregate it becomes easier to spot significant trends in export change that is harder to spot in individual country data that by its nature will have more months with zero imports and will be more erratic, making anomalies harder to spot. This is why Kenyan data was only viable to evaluate 91 products and SSA exports could be used for 535 products.

*Figure 5: Number of anomalies identified before and after Covid – ARIMA log of 2 month average 95% confidence interval*

Maisha Meds did experiment with using lower confidence intervals. Findings from this are listed below. The feeling was that it was better to focus on those anomalies that stood out most.

Figure 6: Exports to all of Sub-Saharan Africa excluding South Africa (Finished pharmaceuticals)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>80% confidence</td>
<td>106</td>
<td>274</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>90% confidence</td>
<td>63</td>
<td>178</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>95% confidence</td>
<td>37</td>
<td>121</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to looking at the number of products experiencing a shortage, Maisha Meds also wanted to understand what proportion of shortages were caused by the Indian export ban on medicines in March 2020 and what percentage affected other molecules. Molecules on the Indian export ban were Chloramphenicol, Clindamycin, Erythromycin, Neomycin, Ornidazole, Paracetamol, Metronidazole, Azithromycin, Progesterone, Pyridoxine / Thiamine, Pyridoxine, Thiamine, Tinidazole and Chloroquine.

Figure 7: Number of anomalies identified after Covid from Indian export ban – ARIMA log of 2 month average, 95% confidence interval

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exports to SSA ex. South Africa (n=XX)</td>
<td>102</td>
<td>19</td>
<td></td>
<td>Exports to SSA ex. South Africa (n=XX)</td>
<td>34</td>
</tr>
</tbody>
</table>
The change in molecule values resulting from the Indian export ban is particularly easy to spot in products with large and regular monthly shipment values. This sort of disruption can be identified with little statistical methodology required.

*Figure 8: Changes in the export of medicines in March 2020 as a result of the Indian export ban*

In addition to looking for significant changes in export volume this analysis also looked for significant changes in the exported price per KG of the API’s in our database. This was made more difficult because for API exports the analysis had only 18 months of data, this meant that the number of products identified as experiencing a significant change to a 95% confidence level is lower. One product that seems low risk to call out from this dataset is Chloroquine Phosphate. Prices of this API spiked particularly high at the start of the Covid outbreak with an almost 500% price increase in the average price per KG as demand for the unproven Covid treatment spiked in Feb – June 2020.

Maisha Meds has looked to avoid mentioning other medicines affected by Covid that may not already be common knowledge to control for adverse effects such as potential hoarding or price gouging.

*Figure 9: Exports of Chloroquine Phosphate to Sub-Saharan Africa Jan 2019 – June 2020*
Figure 10: Anomalies identified in exports of API looking for significant price increases to a 95% degree of confidence.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exports to SSA ex. South Africa (n=76)</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Exports to Kenya (n=12)</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

3. **Further investigation of anomalies**: Having identified anomalies resulting from potential supply chain disruption from Covid, in partnership with IQVIA and CGD, Maisha Meds looked in greater depth at the pricing dynamics associated with these products. To do this it was necessary to manually calculate the pack sizes, forms and strength associated with around 5,000 lines of export data. It was found that many of the products identified as being at risk of shortage, while they had experienced a very real disruption in the supply chain, did not appear to be in global shortage as the price per unit was unchanged. In particular this applied to the products on the Indian export ban list, such as paracetamol. The disruption to the supply chain was real but the underlying cause was not likely to be a global shortage.

Figure 11: Disruption to Paracetamol trade in March had a large effect but price indicates no shortage

![Paracetamol charts](chart)

However, of the 60 products that were examined in depth twelve were identified as experiencing a significant increase in price since March 2020. Of these, only eight had sufficient monthly data to extend the analysis to Jan 2019, to look at whether the price increase in 2020 fit within a band of normal pricing for the product. In 6/8 cases the price increase in 2020 was not significant compared to usual price fluctuations. However, in two cases, the price increase was significantly above the average price per unit expected from Jan 2019 – May 2020. Combined with the drop in monthly exports these products appeared to be at real risk of regional shortage, as opposed to suffering a supply chain interruption as a consequence of Covid-19. One medicine that was highlighted as in potential shortage based on this data was confirmed as in shortage by wholesalers in Kenya in August 2020. The other was confirmed as being in shortage by an NGO implementer working on medicine supply chains in October 2020 [Edit: Addition to initial draft].

4. **Analysis of the lag between exports and medicine price change in Kenya**: As discussed in the methodology Maisha Meds had only 15 molecules with enough data to analyse the
relationship between Indian exports and medicine price increases in Kenya. In 8/15 of these price breaks Maisha Meds was able to identify a statistically significant change in the value of exports from India to Kenya prior to a price rise in Kenya.

Figure 12: Relationship between export disruption to Kenya and price change in Kenyan pharmacies

For the products that exhibited a significant drop in Indian exports in combination with an increase in Kenya pharmacy prices, the resulting price spike occurred 4-5 months later in the Kenyan pharmacies in half of cases. This would indicate that it is possible to use export data to predict shortages before they happen in the country. However, the absolute number of products where Maisha Meds had enough data that overlapped sufficiently with Indian exports to analyse a trend break in price was limited. Furthermore, the number of alternative potential explanations for why the price of a medicine in a country may vary means the strength of the relationship between exports and pharmacy prices should be taken with some degree of caution.

5.0 Discussion

The methodology and results above represent only a proof of concept. Maisha Meds believes it might be possible to build a better and more complete model to inform countries of potential medicine shortages as and potentially before they happen, but only with investment, more data sources and continuous validation of shortage predictions.

5.1 Key findings

The current model catches around half the shortages reported, more when API data is available

When comparing the shortages identified by private wholesalers in interviews in August with the predictions made by the model to a 95% confidence rating, the model identified 18/37 molecules as at risk of shortage to a 95% confidence interval. This rises to 20/37 if you include the molecules where there were 2 outlier periods at an 80% confidence interval. Relaxing the confidence rating to 80% captures 30/37, but given that it also flagged around half of the 521 molecules as being at risk of shortage following Covid, it isn’t really practical to use.

The analysis only had access to 18 months of data for Indian API exports, compared to 40 months of data for finished pharmaceuticals, as this limited the ability to build a time series model and it was only possible to analyse 13/37 molecules in this manner. However, in the instances where we did have sufficient API data the ARIMA model successfully flagged 11/13 products as at risk of shortage
that were confirmed by local wholesalers (or 15/18 including the products highlighted as in shortage not included on the original list of tracer molecules).

Overall flagging 18/37 confirmed shortages isn’t a fantastic result. However, when seen in the context of India only representing 33% of total exports to the sub-Saharan Africa continent, and given the known issues present in the labelling algorithms and API data availability, Maisha Meds views this as an encouraging result for a prototype with limited data inputs and limited time to clean those data inputs.

For the molecules where no wholesaler identified the products as in shortage the ARIMA model identified 4/16 as being at risk of shortage. Interestingly of these two were strongly flagged as at risk (two periods at a 95% confidence interval), and shortages for these products were reported in the UK and Canada. This may indicate that wholesaler interviews also carry risks of missing shortages for certain product types.

Most flags weren’t shortages

A key finding from this work was that although the model observed a large number of products that experienced a severe supply shock in March or April 2020, the number of products upon investigation that appeared to exhibit a genuine regional shortage was relatively small based on wholesaler interviews and more detailed analysis of export data.

It would seem that despite significant interruption to the flow of goods there was enough supply in the system, either from India or elsewhere, for many essential medicines to meet the demand. Demand may also have dropped as well with the pharmacies in the Maisha Meds network observing a ~20% decline in monthly sales after Covid.

That being said there was still a not insignificant number of products where this was not the case and where a “real shortage” was identified through a drop in export value and volume, an increase in average price per unit and confirmation of the shortage from wholesaler interviews on the ground or observed changes in pharmacy prices in Kenya.

Changes in exports typically precede changes in medicine prices in pharmacies by 4-5 months

Another key finding was that it is possible to link changes in export supply data to changes in prices at the pharmacy level. A lag period of 4-5 months from a supply disruption to affecting private sector pharmacies actually makes a lot of sense when one considers the re-ordering cycle of a typical private sector wholesaler in sub-Saharan Africa.

On average a private sector wholesaler in SSA holds between 3-4 months of stock and reorders every 3 months from their suppliers. Upon placing an order after it is manufactured, shipped by sea and cleared through customs it is typically 2 months to arrive with the wholesaler. This 5 month cycle would match with the typical time from a shortage in exports to the materialising of a shortage in the market.

Point of Sale Data can both spot historic shortages and changes in real time

Maisha Meds believes that the data capabilities displayed in this analysis help make the case for moving away from ad-hoc facilities surveys and DHIS-2 data entry that lacks granularity on medicine stock levels, price and daily sales to using digital Point of Sale systems across SSA. The data that results from these systems are able to spot spikes in medicine prices, drops in availability and changes in pharmacy behaviour indicating shortage in real time. Below is an example of a flagging
system built by Maisha Meds to highlight products at risk of shortage based on rapid price increases in Kenya each week.

*Figure 13: Maisha Meds data was able to spot the products going into shortage in Kenyan pharmacies in March 2020 in real time*

Maisha Meds has developed an android based point of sale system that was designed for use in public and private facilities in the most low resource settings (low power and internet connectivity). Funding to increase the usage of digital point of sale systems in Sub-Saharan Africa would help to create a stream of real time data to spot both price spikes and shortages in public and private facilities in real time, as well as improve supply chain functionality in demand forecasting, logistics, price and consumption tracking.

5.2 Caveats and issues identified in the approach

**Resource constraints:** As an early stage prototype built with only a small grant from the Centre for Global Development (CGD) there were multiple constraints on the construction of this model. The labelling library and algorithm could have been significantly refined. The periods analysed for change could have been extended as more data was available, but there was not enough time to analyse it.

**Issues in the accuracy and breadth of the labelling algorithm:** The methodology used for labelling text strings from Indian export data is still in its relative infancy and had breadth and accuracy issues. Only 90% of finished pharmaceutical and API exports by value were labelled due to the diminishing returns of cleaning the data and the resource constraints of the project. In addition, there were also accuracy issues identified in the data labelling process. The main problems were identified in combination products where one of the medicines in the product may have been labelled but the second API may have been misspelled and thus the combination was identified as a monotherapy.

**Lack of categorisation of medicines by form:** The analysis did not look to differentiate medicines by form, as such it would also have missed if there was a shortage of a lower volume medicine form, such as an injectable or suppository. This doesn’t seem to have disadvantaged the analysis when looking at the products reported as in potential shortage by wholesalers but is still worth noting as a potential area for improvement.

**Incomplete data inputs:** Only having access to export data for India limited the accuracy of the analysis to predict shortages, as India accounts for only 33% of all exports to the SSA region by value. Maisha Meds’ pharmacy point of sale data for 300 pharmacies, while very granular in being able to identify real changes for patients in Kenya, is also relatively limited compared to continent wide
medicine consumption, though the organisation is looking to grow the number of pharmacies on the platform.

**Lack of consistent methodology:** When Maisha Meds began to create a medicine flagging tool the organisation did not know exactly what this would look like. As such many of the analyses and processes for the analysis of potential medicine shortages have been invented along the way. Examples of an inconsistent methodology include using differing time periods for API and Finished pharmaceutical analyses.

**Lack of granularity on smaller products and difficulties analysing individual country trends:** Smaller volume products and exports to individual countries weren’t always possible to analyse due to the lack of sufficient data to create a time series. As seen in the results Maisha Meds was able to analyse 525 products for all of sub-Saharan Africa excluding South Africa but when focussing just on Kenya it was only possible to analyse 91 products due to the erratic nature of export data. Taking a continent wide approach is a better way to smooth out erratic export data but makes the findings less obviously applicable to each individual country. One way to factor for this and improve the analysis in the future is if it were possible to split exports into public and private sector exports. As large public sector shipments may be for 12-18 months of supply, while private sector shipments are more likely for 3-4 months of supply for a country being able to disentangle these factors using the consignee names on the shipment manifests could significantly improve the analysis and reduce the “noise” in the data resulting from large public procurement tenders.

**Accuracy issues in the export data:** when looking at the trade data on a line-by-line basis it is common to find that not all the data needed to understand the shipment is recorded. For a row to be fully analysed it requires the pack size, strength and medicine form. In cases where this is not available often these pieces of information have to be inferred or the price per unit is based on the average of shipments from the same month. This represents a real problem for data quality and one that is not possible to solve in all cases through data science, the quality of data capture itself must also be improved.

5.3 Areas for improvement and development

**Addition of new data sources:** Indian export data is easy to access commercially because it can be purchased monthly from a variety of sources. Getting data for other countries is more difficult as each has different levels of data confidentiality. For example in the US one can buy granular data on all imports but not US export data. To make a more effective medicine shortage labelling tool would require not only data from India but also data from the EU and China.
Assuming it is possible to get the export, in 2018 India, China, South Africa, and EU accounted for 92% of records of India, China, and the EU (South Africa finished pharmaceutical exports and Indonesian trade data is available to purchase) the database would have coverage of ~92% of all finished pharmaceuticals going to sub-Saharan Africa.

It would have coverage of ~90% of API going to local manufacturers, though mapping global API flows precisely is difficult.

It would cover 77.4% of medical device exports to the region. If export data for the USA (6%) and South Africa (6%) were added coverage would also be ~90%. There is very little manufacture of medical devices in Africa, barring a few factories producing non-durable items such as syringes.

The ideal methodology to complete this task would be for UN Comtrade to grant access to the granular level trade data for all imports and exports in every country. However, this may not be possible. In this event it would be necessary to apply to each country and organization to request ongoing monthly access to the data. In the case of the EU this would mean applying to Eurostat –

**Eurostat**

**Principle 5: Statistical confidentiality**

“The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes are absolutely guaranteed.” Indicator 5.1: Statistical confidentiality is guaranteed in law defined in the European Statistical Law.

External users [can be] individually approved and sign agreements governing their access to microdata. For the most sensitive data, access is allowed only via a computer terminal in a secure environment. The microdata cannot be removed from this environment.

Alternatively it might be necessary to apply to individual countries based on their specific regulations.


**SubSection 20(6)**

[Legal disclosure of information can occur] ‘To a person exercising public functions in relation to public safety or public health, for the purposes of those functions.’

Explanation: This is designed to cover disclosures to public bodies with a responsibility for public health or public safety. It enables disclosure of information on threats to public health or public safety from any source, including individuals. The threat need not be imminent or serious, although the proportionality and necessity of any disclosure must be carefully considered in every case...

When considering disclosure under Section 20(6), HMRC must know how the information to be provided will enable the recipient to carry out their functions in relation to public health or public safety, and there must be a realistic expectation that the recipient will act on the information disclosed.

**Wholesaler data:** The only source of global wholesaler data is IQVIA, the work already done in parallel to this project by IQVIA has yielded very strong results in spotting potential shortages using wholesaler data. If it were possible to integrate this approach it could only make the eventual risk rating for spotting shortages by country stronger.
Refining of the medicine labelling algorithm: This is quite straightforward, though time consuming. Currently Maisha Meds has a list of ~5,000 drug names and product brand names that it uses to identify product shipments in the Indian export data. By continuing to expand the list of brand names and products, as well as the use of fuzzy matching algorithms to take account of spelling mistakes it will be possible over time to gradually improve the quality of the medicine labelling algorithm.

Creation of a machine learning or AI algorithm for pack size, form and strength: Having identified the medicine in each line of export or import data it is next necessary to understand the pack size, strength of the medicine and the form (injectable, tablet, cream etc...), in order to understand real volumes and price per unit. This is difficult to do just using a labelling algorithm and will require a more complex approach. Export data is often incomplete in how it is written, often elements of a shipment have to be inferred based on other information. For example, if one brand of medicine is 100mg / tablet, in a pack of 30 and usually ships at a price of $1 for 30 tablets, when you find that same brand with a price of $1 per pack in the same month you can assume it is probably 100mg strength and 30 in a pack. However, if it were 6 months later with no additional information one probably couldn’t make the same assumption. This interplay of factors will require a machine learning model with significant algorithm training on each molecule to be able to accurately tag the data going forwards.

Calculating the average number of treatments per kilogram of API: Work may already exist or may need to be done to give an estimate on the total quantity of essential medicine that would be likely created per kilogram of API. This would allow a company to combine the API trade data and the finished pharmaceutical trade data into a single common risk factor.

Going upstream of API: There is a final analysis to go even further upstream that has not been done as part of this project, that is to look at the Indian imports of API and chemicals under trade codes HS 29&28. This would examine whether there are changes in the price and volume of API from China to India. If the Indian finished pharmaceutical exports are able to spot a potential medicine shortage four months before it happens in the country, it is possible that this methodology could spot potential shortages even sooner, though would be even further removed from the country level perspective.