Michael Xiao, Raj Dasani, Salil Goyal

Predicting COVID Vaccination in Africa by the End of 2022

Background

Since the start of the vaccination rollout, a key question has been: who will get the vaccine first and when? In the US, the answer was seemingly straightforward: it would be given to the elderly and those who were labeled high-risk, and then progressively opened up to younger ages (Faden, Crane, and Omer 2021). But when we take this to a global scale, it is not as straight-forward. In the recent stages of the pandemic, the topic of equity of vaccine distribution throughout the world has been in the spotlight (Gill and Ruta 2022). Many public health leaders have noted that poor countries are currently being left behind in the race to get populations vaccinated while the CDC has emphasized that many low income communities across the world are less likely to get the vaccine as of now, and are encouraging countries to assist others in this fight for global immunity (BBC News 2021).

Currently, Africa has the slowest vaccination rate of any continent, with just 20.3% of the population having gotten at least one dose of the vaccine (Holder 2022). To put this into perspective, the Brookings Institution, a public policy think tank based in Washington, D.C., estimated at the end of January 2022 that Africa had received only 6% of all COVID vaccines produced, despite having 17% of the world's population (Sidibé 2022). Many African countries have vaccinated less than 1% of their populations, and Eritrea stands out as having yet to start a vaccination program (Our World in Data n.d.).

There are many hypothesized reasons that Africa has been struggling with getting COVID vaccination rates off the ground. These include poor health infrastructure, a lack of funding for medical staff, and vaccine shortages (Mwai 2021). Early vaccine shortages were caused in large part by wealthy nations draining the vaccine supply from manufacturers; more recently, this has been less of an issue, and now it seems that African nations are able to get doses but are having difficulty with the logistics of distributing them to their populations (Mwai 2021). Currently, there are two main sources of COVID vaccines: donations from countries that produce vaccines through the COVAX program, and bilateral deals that exchanges local vaccine manufacturing for vaccine access (Loembé and Nkengasong 2021). A country's innate ability to uptake vaccines is related to its financial well-being (Basak et al. 2022), government policies (Aborode et al. 2021), and health infrastructures (Williams et al. 2021). And as countries prepare to distribute vaccines to the public, the supply of vaccines from manufacturers becomes the limiting factor. The current estimate leaves a gap of 650 million doses of vaccines that are required to fully vaccinate 40% of the population in low-income countries in Africa.

This has inspired us to look into this further and understand the current rates at which vaccination is currently progressing in terms of vaccinations, and how that aligns with goals mentioned by leaders of African countries. Specifically, we seek to forecast the vaccination rate of Africa by the end of 2022.

Key Considerations

There were a few key considerations we had when creating our forecast. First, we wanted to consider basic metrics that quantify a country's ability to respond to a health crisis. Therefore, we wanted to cover three key branches we thought fit best: economic standing, healthcare infrastructure, and corruption level. For economic standing, we chose GDP, which roughly indicates how wealthy a country is and thus would indicate how quickly a country would be able to source vaccines and distribute them. We obtained the GDP data from the World Bank. A measure of healthcare infrastructure would point to how easily a country is able to store and distribute vaccines. We decided to use the Health sub-index of the 2021 Global Health Security Index as a measure of health infrastructures (GHS). Finally, an index of corruption would give a general indication of how effective a country's response to a crisis might be to its entire population, given resources and vaccination access. We used the Corruption Perceptions Index (CPI) from Transparency International as a measure of corruption.

With these metrics, we wanted to consider these alongside previous and current vaccination rates, with COVID as well as other diseases. With COVID, we considered the current and historical COVID vaccination rates across countries in Africa. Current vaccination rates would give good base rates for what we might expect in the coming year. Historical vaccination rates would provide insight into how vaccination rates have changed over time, and thus give us some insight for how they might change in the coming year. The data for vaccination was obtained from Our World in Data. But to expand more than that, as well as to give us a better understanding of the timeline for a full vaccine rollout, we considered some reference classes–namely, the trends in vaccination rates in Africa for other, vaccine-preventable diseases. Three diseases had easily accessible vaccine data: tuberculosis, HPV, and yellow fever, made available by UNICEF. Analyzing how vaccination rates for these diseases progressed across African countries would give us an idea of what to expect for COVID conditional on what has happened in the past.

Forecasting and Simulations

Naive Estimator

To get a general sense of the projected vaccination rate, we used Facebook's *Prophet* software to generate predictions for vaccination rate in Africa as a whole. *Prophet* is an open-source, fully automated package for time series forecasting. It uses a generalized additive regression model that accounts for seasonality, change in trends, as well as trend anomaly in the data to create roughly a linear trend (Figure 1). Simply using the vaccination rate data since the start of the rollout for the entirety of Africa, *Prophet* projects an 80% confidence interval of [32.71, 37.96] with mean 35.41.

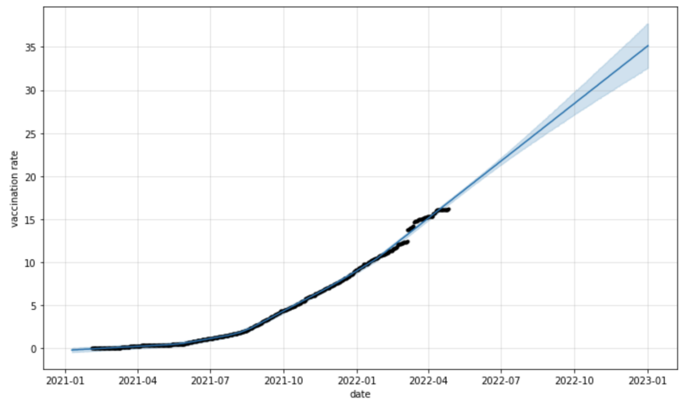

Figure 1. *Prophet* forecast

Curve Fitting, Distributions

As we viewed *Prophet* as a "black box" algorithm that requires minimal human input, we decided to produce a human-generated forecast on Africa as a whole and combined the two forecasts to incorporate our considerations. Our strategy was to identify the distribution of vaccinated individuals, given all the data we had, and seeing if it can be fit to be roughly normal, log-normal, power law, or another type of distribution.

We first simulated the trajectory of vaccination over the entire continent of Africa to get a general sense of the trend. To simulate vaccination, we pooled the daily number of vaccinations from February 1 to April 25 in 2022 and found that a log-normal distribution over the data best fit (Figure 2a). From there, we sampled the data 250 times (one for each day to the end of 2022) to generate a simulated trend. We repeated this process 1000 times and took the average and 80% confidence interval for each day to get the

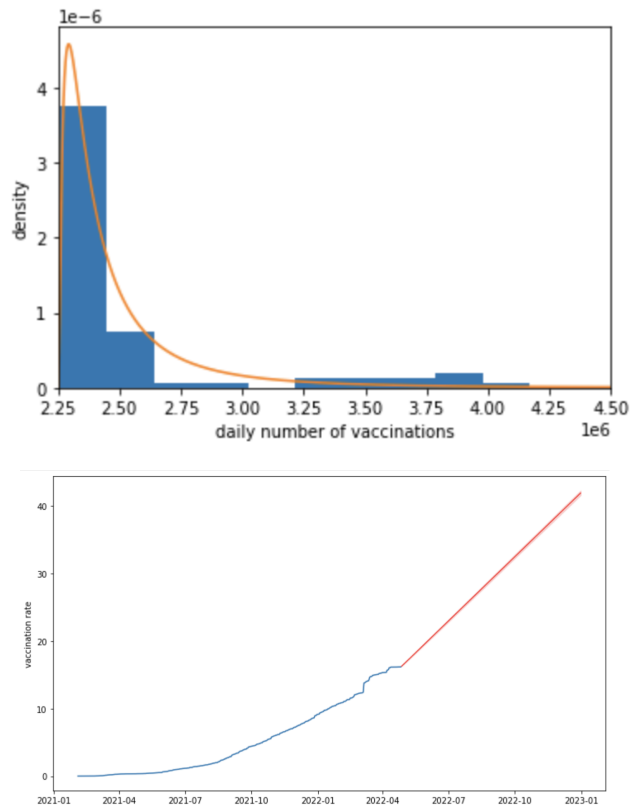

Figure 2. a.(top) log-normal distribution fitted over vaccination data **b.**(bottom) simulated trend

final trend (Figure 2b). The results projected a 80% confidence interval for the number of people fully vaccinated out of 100 people by the end of 2022 of [41.62, 42.29] with mean 41.92.

Trend Extrapolation

Our next methodology was inspired by the curve fitting and its variability: a manual trend extrapolation. We thought this method would be useful in addition to the *Prophet* model, because although it is industrial-grade, the *Prophet* model is sort of a black box whose inner workings we do not (have the time to get to) know. We felt having more control of a trend extrapolation forecast would do well as a complement. After looking at the differences among countries, we thought it would be best to group countries together via the factors we identified above: economic standing, healthcare infrastructure and corruption. For economic standing, we used the public data of GDP for all African countries. For healthcare infrastructure, we found a dataset created by the WHO, which gives a number between 0 and 100 for each country representative of the healthcare infrastructure of each country (GHS). For corruption, we used Transparency International's Corruption Perceptions Index, which ranks the perceived corruption in each country on a scale from 0 to 100 (CPI).

To group these 54 African countries via these three stats, we used a Bayesian hierarchical clustering algorithm, which produced four categories with roughly equal sizes. (We found this to be better than KMeans because of the data normalization among each metric as well as creating equal sized groups). This led to four clusters (Figure 3).

Cluster	#	GDP	CPI	GHS	Examples
0	14	High	High	Avg	South Africa, Algeria
1	14	Avg	Avg	High	Kenya, Nigeria
2	12	Low	Avg	Low	Somalia, Chad
3	13	Low	Avg	Avg	Ethiopia, Madagascar

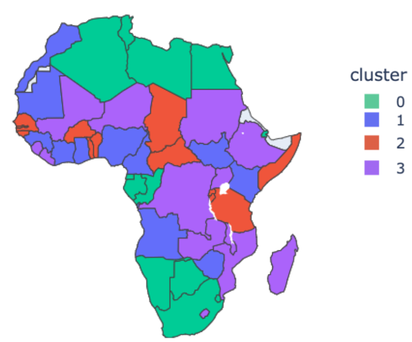

Figure 3. Results of clustering by GDP, CPI, and GHS

With these four clusters, we seeked to find the vaccinations among each country per day, but we ran into the problem where there was data for Africa as a whole in the dataset but not for every country for every day. Especially for smaller countries, we noticed gaps that were a week or a couple weeks long. To fix such, we linearly interpolated so we could make trends for each cluster. We then found the elbows of each cluster to identify where we can extrapolate a linear relationship within the data (Figure 4). We assumed a linear relationship as inspired by Facebook's *Prophet* linear projections as well as for the ease of forecast. Some of these elbows were more clear then others, such as Cluster 2, which we understand to be the next wave of countries to get up to a relatively high vaccination rate, and probably catch up with Cluster 1 and 0. Cluster 2 here is probably the most variable, and may or may not have a significant elbow within the next 6 months. We took this information and considered it in the variability when combining these.

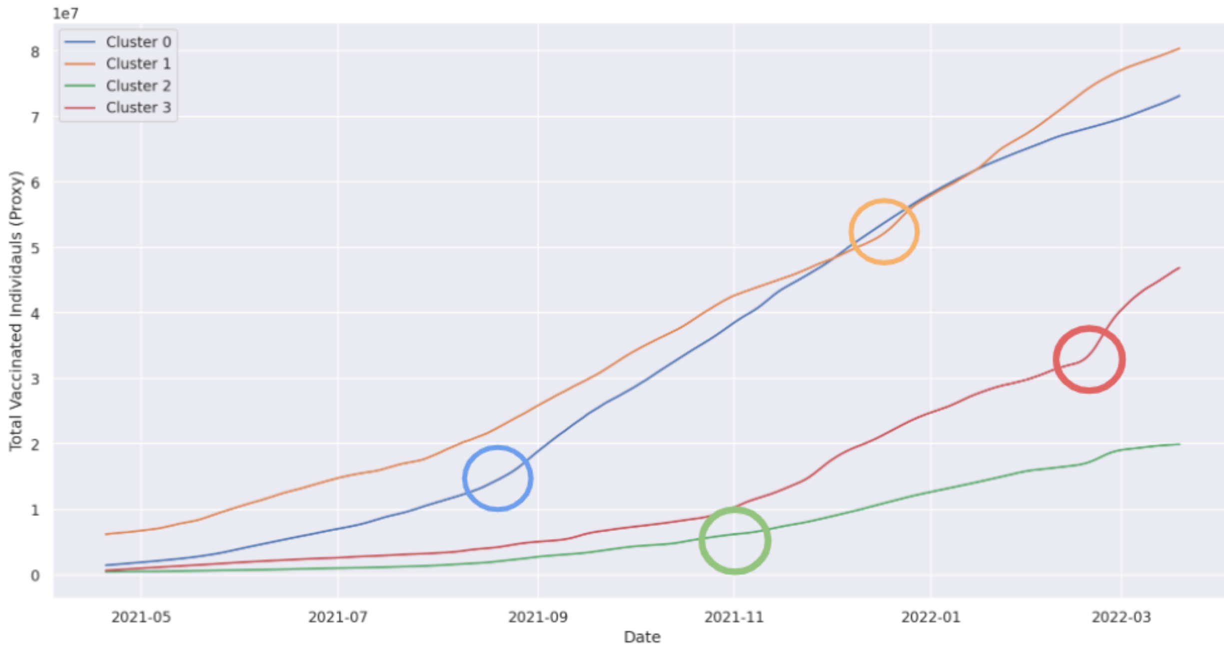

Figure 4. Vaccination trend by cluster. Elbows for each line are annotated with circles.

So, in order to find the projected vaccination rate for each cluster by the end of 2022, we found the vaccination rate at each of these elbows and the most recent day we had (March 19th, 2022) and used these to compute a slope and extrapolate till the end of the year (Figure 5).

Cluster	GDP	CPI	GHS	Examples	Curr Vacc %	Elbow	Slope (per day)	Mean Pred. (End 2022)	80% CI
0	High	High	Avg	South Africa, Algeria	29.32%	2021-08-15	276k	61.14%	[55%, 75%]
1	Avg	Avg	High	Kenya, Nigeria	16.71%	2021-12-15	308k	35.1%	[30%, 45%]
2	Low	Avg	Low	Somalia, Chad	10.24%	2022-11-01	100k	24.91%	[15%, 32%]
3	Low	Avg	Avg	Ethiopia, Madagascar	9.86%	2022-02-15	456k	37.34%	[20%, 50%]

Figure 5. Forecast results from clustering

With these, the 80% Confidence Intervals were created based on the GHS level of each cluster and a rough estimate of their previous highest and lowest slopes. Then using the means and CIs, we combined them, in ratio to each cluster's population to get our mean and 80% CI for the entire country as follows:

Mean: 39.08%

80% CI: [28.97%, 50.24%]

Reference Classes

For our third methodology, we used several reference classes to adjust our predictions. The most accessible vaccination data was available through UNICEF. We first gathered data on the tuberculosis vaccine BCG. There were several abnormalities in the data we found for this, though. This vaccine was developed in the early 1900s, but the WHO only apparently started seriously pursuing vaccination in Africa starting in the 1970s [5]. Further, the UNICEF website wouldn't allow us to download more than 34 years' worth of data, so our dataset was restricted to 1986-2020. Even still, much of the data before the late 1980s and early 1990s for many countries was missing. Then, to get an estimate of the trends of BCG vaccination over the years in Africa, we took a weighted average of the vaccination rates across all African countries for each year (where the weights were the populations of each country), and ended up with one Africa-representative number for vaccination rate for each year for which we had data. Then, we averaged the difference in vaccination rates between successive years and arrived at one number representing the average percentage increase in vaccinations per year in Africa. For BCG, this number turned out to be 0.68%, mostly because the differences in vaccination rates between some years were negative. The UNICEF dataset gave no additional information, so we hypothesized that this was because some years might have had less vaccines given relative to the number of births. For bounds, we took the maximum and minimum difference in vaccination rates between successive years, for which we got 6.42% and -7.82%, respectively. As a sanity check, the negative vaccination rate doesn't make a lot of sense given the push for COVID vaccines in Africa, so we instead rounded the lower bound up to 0%.

We repeated the same process for the yellow fever and HPV vaccines. The data for the yellow fever vaccine were similarly plagued with missing value issues; the data only really started from the mid-2000s. Processing it as above yielded a value of 1.07%, meaning on average, the yellow fever vaccination rate in Africa increased by 1.07% every year. The bounds ended up being 9.63% and -6.98%, and we similarly rounded up the lower bound to 0%.

For the HPV vaccine, data was only available for the years 2015-2020; furthermore, there were abnormally low numbers for 2020, likely because of the COVID pandemic. Hence, we dropped 2020 from the analysis here. Even still, processing the HPV vaccine yielded -2.78%. We are not completely sure as to why this is the case; however, a quick sanity check tells us that the HPV vaccination rates won't be very useful as a reference class, because a negative COVID vaccination rate doesn't make sense in the current context. Hence, we discarded the HPV vaccination rates as a reference class.

Just using the BCG and yellow fever vaccine data as reference classes, and assigning them equal weight, we ended up projecting an 80% confidence interval for percentage people vaccinated of [20.3, 28.33], with mean 21.20.

Combining Forecasts

Finally, we combined our four forecasts by assigning each one a weight and then taking a weighted average. We assigned the *Prophet* forecast a weight of 0.35 because it is a state-of-the-art industrial forecast software, but it uses solely historic vaccination data as input. For the forecast generated by simulation, we assigned a weight of 0.15. The approach seems plausible in general, but when we sampled from the distribution, we assumed that the number of vaccinations each data is independent, which is not necessarily true. We were most confident about the trend extrapolation we computed, so we assigned that a weight of 0.48. To finish up, we assigned the reference classes a weight of 0.03, because the forecasts ended up not making much sense (likely partially due to subpar quality data). The forecasts and weights are summarized below:

- Prophet: [32.71, 37.96] with mean 35.41. Weight: **35%**
- Curve-fitting: [41.62, 42.29] with mean 41.92. Weight: **15%**
- Trend extrapolation: [28.97, 50.24] with mean 39.08. Weight: **47%**
- Reference classes: [20.3, 28.33], with mean 21.20. Weight: **3%**

We compute a weighted average of the means and 80% confidence intervals to arrive at a final mean forecast of 37.96 and 80% confidence interval [31.92, 44.09] for the percentage of the African population vaccinated at the end of 2022.

Summary

To forecast the percentage of fully vaccinated people in Africa by the end of 2022, we combined four forecasting approaches: machine-generated trend extrapolation, distribution estimation and simulation, clustering trend extrapolation, and use of reference classes. *Prophet*, a forecasting package, provided an industrial-grade, automated forecast of general trends. Sampling from a fitted log-normal distribution accounted for the variation in daily differences of number of vaccinations. Clustering enabled us to examine the trends in closer scales with respect to the capabilities of countries to distribute vaccines. Reference class provided the perspective of the historically low vaccination uptakes in Africa. Combining these approaches via weighted average, we arrived at the final prediction of an 80% confidence interval of 31.92% to 44.09%, with a mean of 37.96%.

We would also like to acknowledge some limitations of our investigation. Firstly, a lot of our data was sourced from the UN/WHO/UNICEF, but these data often had lots of missing values. Second, we assumed roughly constant vaccine rollout trends. Third, we never looked at the supply side of vaccines, but rather only assessed the distribution side. Fourth, we did not look at vaccine receptiveness among the general populations or at government-sponsored vaccine messaging across countries. Most of these limitations were due to time constraints on our end, and it is self-evident that a forecast of this nature would be improved with these additional considerations.

References

Aborode, Abdullahi T., Oluwatosin A. Olofinsao, Ekwebelem Osmond, Abiokpoyanam P. Batubo, Omowonuola Fayemiro, Onigbinde Sherifdeen, Luqman Muraina, Babatunde S. Obadawo, Shoaib Ahmad, and Emmanuel A. Fajemisin. 2021. “Equal Access of COVID-19 Vaccine Distribution in Africa: Challenges and Way Forward.” *Journal of Medical Virology* 93 (9): 5212–15. https://doi.org/10.1002/jmv.27095.

Basak, Palash, Tanvir Abir, Abdullah Al Mamun, Noor Raihani Zainol, Mansura Khanam, Md Rashidul Haque, Abul Hasnat Milton, and Kingsley Emwinyore Agho. 2022. “A Global Study on the Correlates of Gross Domestic Product (GDP) and COVID-19 Vaccine Distribution.” *Vaccines* 10 (2): 266. https://doi.org/10.3390/vaccines10020266.

BBC News. 2021. “Covax: How Many Covid Vaccines Have the US and the Other G7 Countries Pledged?” *BBC News*, September 23, 2021, sec. World. https://www.bbc.com/news/world-55795297.

Faden, Ruth R., Matthew A. Crane, and Saad B. Omer. 2021. “Opinion | The Best Vaccination Strategy Is Simple: Focus on Americans 65 and Older.” *Washington Post*, February 8, 2021. https://www.washingtonpost.com/opinions/2021/02/08/best-vaccination-strategy-is-simple-focus-americans-65-older/.

Gill, Indermit, and Michele Ruta. 2022. “Why Global Vaccine Equity Is the Prescription for a Full Recovery.” *Brookings* (blog). February 11, 2022. https://www.brookings.edu/blog/future-development/2022/02/11/why-global-vaccine-equity-is-the-prescription-for-a-full-recovery/.

Holder, Josh. 2022. “Covid World Vaccination Tracker - The New York Times.” April 24, 2022. https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html.

Loembé, Marguerite Massinga, and John N. Nkengasong. 2021. “COVID-19 Vaccine Access in Africa: Global Distribution, Vaccine Platforms, and Challenges Ahead.” *Immunity* 54 (7): 1353–62. https://doi.org/10.1016/j.immuni.2021.06.017.

Mwai, Peter. 2021. “Covid-19 Vaccinations: African Nations Miss WHO Target.” *BBC News*, December 31, 2021, sec. Reality Check. https://www.bbc.com/news/56100076.

Our World in Data. n.d. “COVID-19 Task Force Dashboard.” Accessed April 27, 2022. https://data.covid19taskforce.com/data.

Sidibé, Michel. 2022. “Vaccine Inequity: Ensuring Africa Is Not Left Out.” *Brookings* (blog). January 24, 2022. https://www.brookings.edu/blog/africa-in-focus/2022/01/24/vaccine-inequity-ensuring-africa-is-not-left-out/.

Williams, Victor, Bassey Edem, Marianne Calnan, Kennedy Otwombe, and Charles Okeahalam. 2021. “Considerations for Establishing Successful Coronavirus Disease Vaccination Programs in Africa.” *Emerging Infectious Diseases* 27 (8): 2009–16. https://doi.org/10.3201/eid2708.203870.