

BMJ Open Exploring COVID-19 vaccine hesitancy and uptake in Nairobi's urban informal settlements: an unsupervised machine learning analysis of a longitudinal prospective cohort study from 2021 to 2022

Nandita Rajshekhar,¹ Jessie Pinchoff ,² Christopher B Boyer,³ Edwine Barasa,⁴ Timothy Abuya ,⁵ Eva Muluve,⁵ Daniel Mwanga,⁵ Faith Mbushi,⁵ Karen Austrian ⁵

To cite: Rajshekhar N, Pinchoff J, Boyer CB, *et al.* Exploring COVID-19 vaccine hesitancy and uptake in Nairobi's urban informal settlements: an unsupervised machine learning analysis of a longitudinal prospective cohort study from 2021 to 2022. *BMJ Open* 2023;**13**:e071032. doi:10.1136/bmjopen-2022-071032

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-071032>).

Received 13 December 2022
Accepted 23 August 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Jessie Pinchoff;
jpinchoff@popcouncil.org

ABSTRACT

Objectives To illustrate the utility of unsupervised machine learning compared with traditional methods of analysis by identifying archetypes within the population that may be more or less likely to get the COVID-19 vaccine.

Design A longitudinal prospective cohort study (n=2009 households) with recurring phone surveys from 2020 to 2022 to assess COVID-19 knowledge, attitudes and practices. Vaccine questions were added in 2021 (n=1117) and 2022 (n=1121) rounds.

Setting Five informal settlements in Nairobi, Kenya.

Participants Individuals from 2009 households included.

Outcome measures and analysis Respondents were asked about COVID-19 vaccine acceptance (February 2021) and vaccine uptake (March 2022). Three distinct clusters were estimated using K-Means clustering and analysed against vaccine acceptance and vaccine uptake outcomes using regression forest analysis.

Results Despite higher educational attainment and fewer concerns regarding the pandemic, young adults (cluster 3) were less likely to intend to get the vaccine compared with cluster 1 (41.5% vs 55.3%, respectively; $p<0.01$). Despite believing certain COVID-19 myths, older adults with larger households and more fears regarding economic impacts of the pandemic (cluster 1) were more likely to ultimately get vaccinated than cluster 3 (78% vs 66.4%; $p<0.01$), potentially due to employment requirements. Middle-aged women who are married or divorced and reported higher risk of gender-based violence in the home (cluster 2) were more likely than young adults (cluster 3) to report wanting to get the vaccine (50.5% vs 41.5%; $p=0.014$) but not more likely to have gotten it (69.3% vs 66.4%; $p=0.41$), indicating potential gaps in access and broader need for social support for this group.

Conclusions Findings suggest this methodology can be a useful tool to characterise populations, with utility for improving targeted policy, programmes and behavioural messaging to promote uptake of healthy behaviours and ensure equitable distribution of prevention measures.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ A strength of modern statistical methods, such as K-Means clustering, is the ability to facilitate data-driven analysis, objectively revealing subgroups without the researchers' preconceived assumptions potentially biasing the analysis.
- ⇒ A strength of this study is its longitudinal prospective design, following respondents from 2 months after the pandemic was declared through to vaccine availability.
- ⇒ Some limitations to K-Means clustering include possible changes to the clustering of the data when run multiple times due to the use of random starting points and challenges in interpreting the data when distinct subgroups are not present.
- ⇒ Limitations in the study design include potential selection bias favouring respondents who had mobile phones as well as social desirability bias, whereby respondents may have answered questions to be socially acceptable to the interviewer.
- ⇒ Relatedly, the study has high attrition due to the repeat rounds of collection.

INTRODUCTION

WHO officially declared COVID-19, a disease caused by the novel coronavirus SARS-CoV-2, a pandemic on 11 March 2020.¹ The first case of COVID-19 in Kenya was reported shortly after on 13 March 2020. To curb transmission, the Kenyan Government swiftly instated lockdown policies including restrictions on travel and large gatherings, and business and school closures. Experts were concerned that due to limited resources for distancing and hand washing, that populations in urban informal settlements would be at high risk of transmission.² Many studies

regarding COVID-19 and other outbreaks, such as Ebola, have cited loss of income, food insecurity, gender-based violence, mental health and lack of access to healthcare needs as major downstream impacts of disease mitigation policies.³⁻⁵ In the years since the pandemic began, restrictions have eased and with the rollout of COVID-19 vaccines to the general public in early 2021, the focus has shifted to increasing vaccination coverage. While vaccination is critically important, during initial phases of the rollout, 82% of globally available doses went to high and upper middle-income countries, with only 0.2% delivered to low-income and middle-income countries, highlighting continued vaccine inequity and injustice.⁶⁻¹⁰ As of July 2023, 65.9% of individuals globally have taken both doses of the COVID-19 vaccine.¹¹

The government of Kenya launched a phased rollout of COVID-19 vaccination from March 2021, starting with essential workers such as healthcare providers, then the elderly and those with comorbidities. In June 2022, the Kenyan Ministry of Health expanded their reach and aimed to vaccinate 27 million eligible adults and 5.8 million teenagers by the end of the year.¹² Certain jobs require vaccination such as civil servants, teachers and some private employers.¹³⁻¹⁶ Ongoing campaigns aim to increase vaccination coverage, assuage concerns about vaccine safety and promote uptake to protect Kenyans from severe outcomes and death as well as to protect against new and emerging variants. Vaccination is one of the most effective interventions to control the ongoing pandemic but vaccine acceptance rates around the world vary.¹⁷⁻¹⁹

Vaccine hesitancy is a major ongoing global concern as it is likely there will continue to be new vaccines or boosters required as the pandemic evolves. A study across 23 countries worldwide (including Kenya) found that soon after the vaccines were available (June 2021), over three-quarters (75.2%) of respondents reported vaccine acceptance, meaning they would get the vaccine. Reasons for vaccine hesitancy related to lack of trust in COVID-19 vaccine safety and science, and scepticism about its efficacy.¹⁹ Other factors included misperceptions regarding individual-level risk of contracting COVID-19, the severity of infections¹⁹⁻²⁴ and fear of side effects.²⁵ Some people surveyed reported a general lack of trust in scientific institutions or health authorities which can also increase vaccine hesitancy.¹⁹

Looking closer at COVID-19 vaccine hesitancy in Kenya, an early study in four Kenyan counties found hesitancy ranged from 10.2% to 44.6%, with Nairobi County having the highest proportion that reported they intended to get the vaccine, particularly among those who had received training from the Ministry of Health.²⁶ A 2022 study from six Kenyan health facilities found that while 81% reported it was important to get the vaccine, 40.5% also reported concerns, mainly regarding side effects.⁶ This study also found that hesitancy was higher in government and faith-based health institutions compared with private ones.⁶ Another study conducted in February 2022 found

that >45% of individuals eligible for vaccination in Kenya had not taken a single dose.^{19 27 28}

To increase vaccine uptake, it is important to address hesitancy by identifying sources of information, perceived trustworthiness of sources and how messaging can be adapted to drive positive behaviour change. Studies have shown that individuals who report receiving COVID-19 information from social media, primarily Facebook, have the highest rates of vaccine hesitancy.^{6 26} An Africa CDC report found that among those surveyed in Kenya, 65% reported having seen or heard at least some misinformation about COVID-19 from social media.²⁹ Overall, the potential for social media to contribute to misinformation is concerning, as the information shared is not scientifically filtered or reviewed. Other sources commonly reported for COVID-19 information include television (TV), SMS from government agencies and health providers. An African CDC report found that in Kenya, 78% of those surveyed say that TV is a trusted source of information.²⁹ In Nairobi, a study revealed that government health messages through TV, radio and SMS were among the most common sources of information for residents in urban informal settlements at the initial onset of the COVID-19 pandemic.³⁰ In particular, it is important to understand how young adults receive and interpret information regarding COVID-19, as some studies suggest this age group may be extremely hesitant because of perceived low risk of severe outcomes, mistrust in authority and fear regarding side effects especially around infertility and pregnancy outcomes.³¹⁻³³ A global study found young people were most likely to search for COVID-19 and other health information from social media, raising concerns about exposure to misinformation.³⁴

This study analyses data from a sample of individuals residing in urban informal settlements in Nairobi, surveyed in 2021 and 2022, before and after the distribution of the first COVID-19 vaccine. An exploratory analysis was implemented to understand how the characteristics of respondents could point to vaccine acceptance/hesitancy (prior to availability) and uptake (after the vaccine was available). We explored the utility of K-Means clustering to characterise participants based on demographics, knowledge, perceptions, risks and other factors, to determine if certain archetypes or subgroups are present in the cohort; and if so, how likely they are to want to take the COVID-19 vaccine and ultimately get it. We selected K-Means analysis because it is a data-driven approach, meaning that the patterns are derived from the data itself, a less biased method to characterise 'types' of participants. K-Means have been used in previous studies to group together participants in a dataset to predict health prevention and treatment strategies for each group.³⁵ We compared this statistical approach with a more basic one, to highlight the utility of K-Means clustering to understand unmeasured characteristics of the groups. Ultimately, K-Means clustering identified three subgroups in the dataset with implications for COVID-19 vaccination policy and messaging.

Table 1 Cohort demographics for round 1 (n=2009) respondents from five informal settlements in Nairobi, Kenya April 2020

Variable	Frequency (%)
Age (mean (SD))	36.5 (11.3)
Age in categories (years)	
18–29	576 (28.7)
30–49	1184 (59.0)
50+	248 (12.4)
Female gender	1258 (62.8)
Education	
Primary or less	866 (43.2)
Secondary	878 (43.9)
Higher	257 (12.8)
Marital status	
Married	1170 (58.5)
Single	502 (25.1)
Divorced/Separated	328 (16.4)
Vaccine acceptance (2021)*	799 (72.1)
Vaccine uptake (2022)†	797 (71.1)
*Question added in round 5 (n=1108).	
†Question added in round 6 (n=1121).	

METHODS

Sample and survey design

The Population Council, in collaboration with the Kenya Ministry of Health, conducted a longitudinal prospective cohort study across five informal settlements (Kibera, Mathare, Kariobangi, Huruma and Dandora) in Nairobi, Kenya to understand knowledge, attitudes and practices around COVID-19. Participants were sampled from two previous longitudinal cohorts, Adolescent Girls Initiative-Kenya (AGI-K) (n=2565) and *Nisikilize Tujengane* (NISITU): engaging men and boys in girl-centred programming (n=4519). For AGI-K and NISITU surveys, household listings were generated and eligible households contained at least one adolescent member were sampled. For AGI-K and NISITU, sample size calculations were conducted and samples selected accordingly.

For the COVID-19 survey, 3465 households were randomly sampled from the AGI-K and NISITU cohorts and stratified by informal settlement, so they are somewhat representative but had to have at least one adolescent household member (eg, a household with only one adult member would not have been eligible for inclusion). For the COVID-19 surveys, we were aiming for a sample size of 2000 or 400 per informal settlement.³⁰ Of the random sample from AGI-K and NISITU (n=3465), 24% of the numbers were no longer in use, but refusals were quite low at about 1%. The resulting cohort for this COVID-19 study includes 2009 adult household members interviewed on 30 March 2020 and 31 March 2020 just after the pandemic was declared. Repeated mobile phone

surveys were completed in April (n=1768), May (n=1750), June (n=1525) of 2020, February 2021 (n=1117) and March 2022 (n=1121). Attrition was high given the frequent repeat nature of the survey and possibility of mobile phone numbers being discontinued, but given the unknowns early in the pandemic, the possibility of attrition was weighed against gathering critically needed information.

Survey questions include demographics, knowledge and awareness of COVID-19 transmission and symptoms, perceived risk, socioeconomic effects of the pandemic, health and mental health indicators, gender-based violence and uptake of various protective behaviours such as masking, isolating if sick, testing and vaccination (see questionnaires in online supplemental files 1 and 2). All interviews were conducted by phone by a team of 77 Kenyan surveyors to adhere to national physical distancing policies to prevent the spread of COVID-19. Respondents gave informed consent over the phone before commencing the survey. The same approach was used for all surveys at each time point. Only the questionnaire changed, with questions added or adapted between rounds.

Measures of variables

Relevant variables were selected based on how likely they are to influence behaviour and vulnerability to the effects of COVID-19 and missing values were imputed using the mice R package. The included demographic and behavioural variables were age, gender, educational attainment, marital status, slum, perceived risk, knowledge of symptoms, what myths they believe, disease prevention measures taken, symptoms experienced, social and economic impacts, household size, government assistance received and fears around COVID-19. These variables were used to construct subgroups using unsupervised machine learning, a variable description and summary statistics are included in online supplemental table 1.

Data analysis

The data were analysed using R V.4.1.2. To identify potentially relevant data-dependent subgroups, K-Means clustering was applied. This is an unsupervised, data-driven machine learning method of exploratory analysis often used to determine the number of 'clusters' that naturally exist within a high-dimensional space formed by a set of possible covariates. K-Means clustering was run, and three clusters were identified, even with repeated attempts, suggesting distinct subgroups. Silhouette plots (online supplemental figure 1) were visualised to find the appropriate number of clusters, and cluster means of each variable were calculated and tabulated (online supplemental table 2) to display the characteristic breakdown of each cluster.

To assess the value of the K-Means algorithm against more traditional methods, we ran likelihood ratio tests. The likelihood ratio test compared the fit of a model

Table 2 Likelihood ratio test for vaccine hesitancy (Nairobi survey round 5; February 2021, prior to vaccine rollout in Kenya), where H0: outcome~intercept+covariate and H1: outcome~intercept+covariate+cluster indicator

Outcome: vaccine acceptance ("How likely are you to take the COVID-19 vaccine if it were offered today?")	
Covariate	Likelihood ratio test P value
Education	<0.0001
Marital status	<0.0001
Age	0.111
Household size	<0.0001
Concerned the pandemic will impact income	<0.0001
Likely to test if symptomatic, know if positive for COVID-19	<0.0001
Know at least three symptoms of COVID-19	<0.0001
Household gender-based violence risk	<0.0001
Eat less due to COVID-19	<0.0001
Loss of income experienced due to COVID-19	<0.0001

containing demographic covariates of interest alone versus a model with the addition of a cluster indicator. We conducted this analysis twice, once for the outcome of vaccine hesitancy (in 2021, prior to vaccine availability) and again for the outcome of vaccine uptake (in 2022, once the vaccine was widely available). For each of these outcomes of interest, p values were calculated for each model containing a demographic covariate of interest when nested (H0: outcome~intercept+covariate) and complex (H1: outcome~intercept+covariate+cluster

Table 3 Likelihood ratio test for vaccine uptake (Nairobi survey round 6, March 2022), where H0: outcome~intercept+covariate and H1: outcome~intercept+covariate+cluster indicator

Outcome: vaccine uptake ("Have you had at least one dose of the COVID-19 vaccine?")	
Covariate	Likelihood ratio test P value
Education	<0.0001
Marital status	<0.0001
Age	0.966
Household size	<0.0001
Concerned the pandemic will impact income	<0.0001
Likely to test if symptomatic, know if positive for COVID-19	<0.0001
Know at least three symptoms of COVID-19	<0.0001
Household gender-based violence risk	<0.0001
Eat less due to COVID-19	<0.0001
Loss of income experienced due to COVID-19	<0.0001

indicator), with significant p values indicating that the model with the cluster indicator (complex model) is a better fit for the data. Overall, significant p values for the likelihood ratio tests for each demographic covariate highlight that the cluster variable adds additional, unmeasured information about the subgroups in the dataset versus the demographic covariate alone. Separate models were fit for age, education, marital status, household size, likely to know positive COVID-19 status, knowledge of COVID-19 symptoms, household gender-based violence risk, economic impacts (food insecurity and income loss) and respondent concerns around loss of income due to COVID-19.

After creating the clusters, we used the newly defined cluster variable to compare vaccine hesitancy and vaccine uptake across the three groups using regression forest analysis, an approach which uses non-parametric statistical estimation based on random forests, to estimate the conditional mean of the outcomes of interest. The best-fit tree was found, and the results were visualised as forest plots using ggplot in R. P values were calculated for three-way and pairwise comparisons of the clusters for vaccine acceptance and vaccine uptake using Wald tests.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

RESULTS

Participants had an average age of 36.5 years (SD=11.3) with 59% of participants between ages 30 and 40 years, 28.7% of participants aged 18–29 years and 12.4% of participants aged 50+ years, over half were female (62.8%) and over half were married (58.5%) (table 1). In 2021, before the vaccine was widely available, most of the respondents (72.1%) said they would be willing to get a vaccine, and about this same percentage had received the vaccine in 2022 once it was available (71.1%). However, this means over a quarter (29%) still had not received the vaccine at the time of the most recent survey.

Based on the results of the K-Means clustering, each of the three clusters that emerged define slightly different 'types' of people. Cluster 1 contained older, married individuals who knew less about common COVID-19 symptoms, were more likely to have believed common myths around COVID-19, and lived in the largest households. Members of this cluster also had the most concern about potential economic harms (fear of food shortages and loss of income) and had a higher perceived risk of COVID-19 early in the pandemic. Cluster 2 primarily consisted of less educated, married or divorced, middle-aged women who were the most economically impacted (skipping meals, loss of income, lack electricity at home, lack social support) at the beginning of the pandemic. These individuals were also the most likely of the three groups to report a perceived risk for gender-based violence from increased

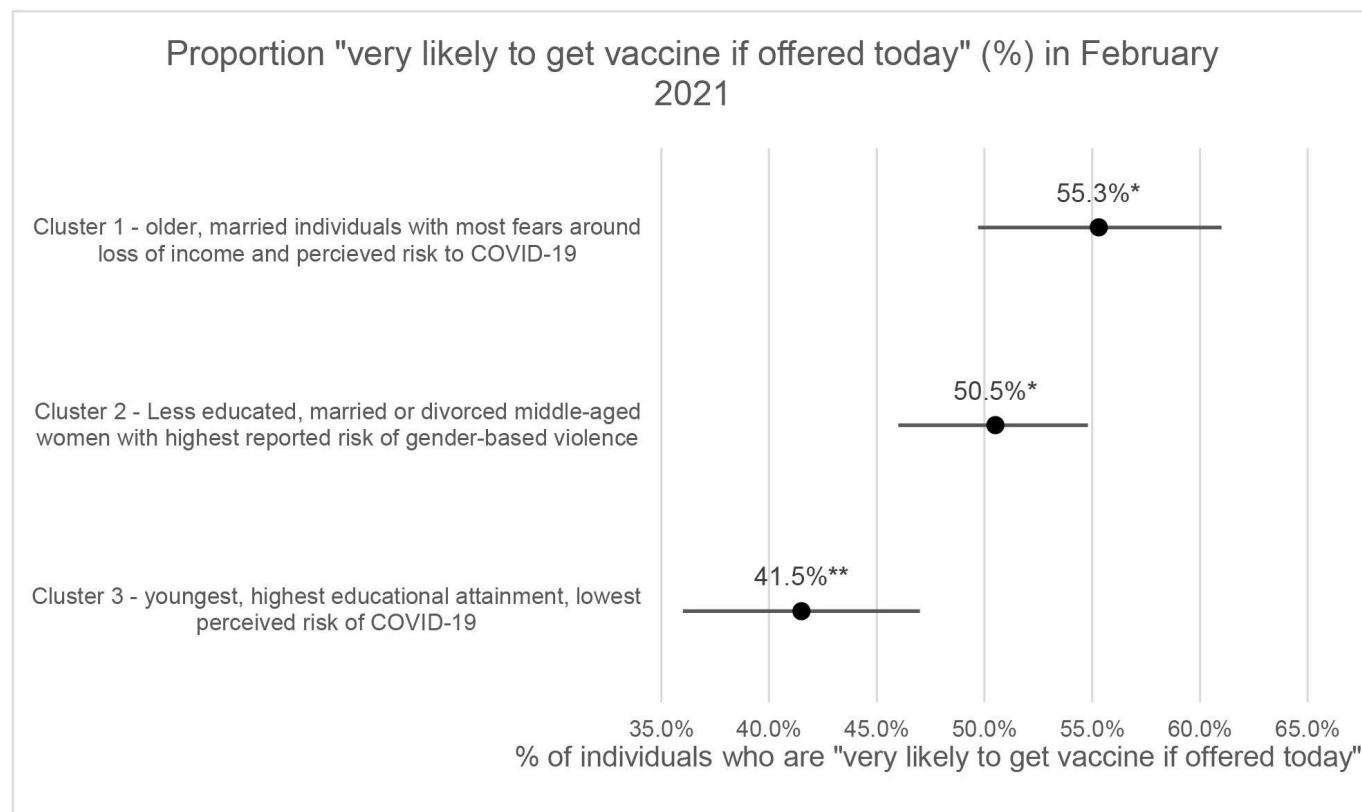


Figure 1 Regression forest analysis plot of vaccine acceptance by cluster, Nairobi, Kenya February 2021 (n=1117). *Cluster 1 and cluster 2 are significantly different than cluster 3, but not each other. **Cluster 3 is significantly lower than cluster 1 and cluster 2 (p<0.01 and p=0.014, respectively).

tensions at home due to the pandemic. Cluster 3 was the youngest group with higher educational attainment, who had a higher average knowledge of COVID-19 symptoms and expressed fewer fears around the economic impacts of lockdowns early in the pandemic. The mean values of each demographic variable per cluster is presented in online supplemental table 2, and clusters are described in online supplemental table 3. The silhouette plots presented in online supplemental figure 1 highlight the three clusters selected that best capture the variation in the dataset.

We then ran the likelihood ratio tests to compare each variable to see if the fit was better with the variable alone (nested model) or with the addition of the cluster indicator (complex model). All of the likelihood ratio tests except for age were significant, revealing that when included in the model, the clusters defined using the K-Means algorithm are a better fit for the data than individual characteristics alone (table 2 presents for outcome of vaccine hesitancy in survey round 5 and table 3 for the outcome of vaccine uptake in round 6).

After completing the likelihood ratio tests and concluding that the clusters offer more information than demographic variables alone, we used regression forest analysis to explore the association between cluster identification and the two vaccine-related outcomes. For vaccine acceptance (2021), cluster 3 was significantly less likely to say they would get the vaccine if it became

available compared with cluster 1 (41.5% vs 55.3%; p<0.01) and compared with cluster 2 (41.5% vs 50.5%; p=0.014) (figure 1). Once the vaccine became available and participants were asked about vaccine uptake in 2022, cluster 1 was significantly more likely to have gotten at least one dose of the vaccine compared with cluster 2 (78.0% vs 69.3%; p<0.01), and more likely than cluster 3 (78.0% vs 66.4%, p<0.01) (figure 2). Additionally, cluster 2 was more likely than cluster 3 to report wanting to get the vaccine (50.5% vs 41.5%; p=0.014) but not more likely to have gotten it (69.3% vs 66.4%; p=0.41). Of the 29% (n=324) in round 6 who have not gotten the vaccine, about half are hesitant (48%) and about half say they are very likely to still get the vaccine (not shown).

DISCUSSION

Our findings suggest that survey respondents from across Nairobi informal settlements fall into three clusters or archetypes each with distinct characteristics that can provide insight into COVID-19 vaccine uptake. Kenya, and our sample specifically, achieved high vaccination coverage (almost three-quarters of respondents). This estimate is in line with a global study that suggested a maximum share of 70% of the total population could be vaccinated, without application of coercive policies or restrictions.³⁶ Our exploratory analyses suggest the cluster indicator adds value to basic models describing

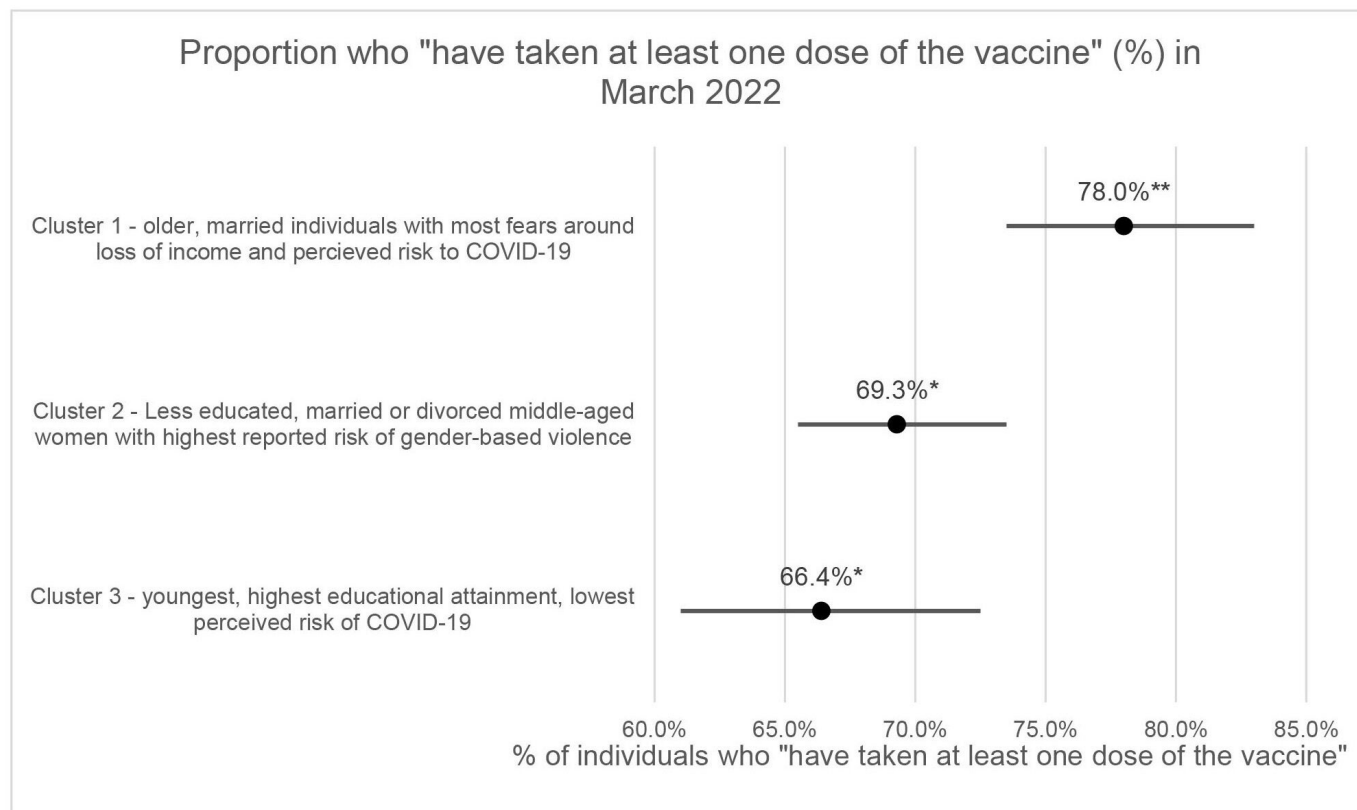


Figure 2 Regression forest analysis plot of vaccine uptake by cluster, Nairobi, Kenya March 2022 (n=1121). *Cluster 2 and cluster 3 are significantly different than cluster 1, but not each other. **Cluster 1 is significantly higher than cluster 2 and cluster 3 ($p < 0.01$ for both).

characteristics associated with vaccine uptake, capturing unmeasured characteristics of participants that are associated with the outcome. The clusters may be useful to identify archetypes of individuals in informal settlements and suggest avenues to explore for communication with subgroups that have different vulnerabilities and risks. Our results suggest some variation between the three groups of respondents in vaccine uptake, information that can be used to better target or improve messaging to increase awareness and adoption of healthy behaviour.^{37–42}

It is concerning to find that primarily younger, more highly educated individuals, with highest knowledge of COVID-19 transmission in cluster 3 are least likely to have gotten the vaccine. They reported being less concerned with COVID-19 infection and the economic impacts, potentially indicating less urgency due to a lack of perceived risk, as initially risks to the elderly were highlighted. A recent study confirms this link, and that lack of perceived risk and low perceived disease severity were leading factors for not getting vaccinated.⁴² Relatedly, those in cluster 3 were less likely to know someone who had tested positive for COVID-19 (17% vs 25% in cluster 2 and 27% in cluster 1) reinforcing their lower perceived risk (Supplementary Table 2) (not shown). It is also likely younger people might be exposed to different information through their higher use of social media. Public health messages tailored to youth⁴³ could highlight vaccine safety, as our participants' main concerns

were about side effects or wanting to wait and see if it is safe. Studies in other settings show young people may be concerned about myths regarding vaccine side effects that affect fertility.⁴⁴ Lastly, it would also be useful to ensure access to vaccines for young people, potentially expanding current outreach to include mobile clinics or other options instead of requiring a visit to a health facility. Nairobi is already employing strategies for vaccine outreach including providing vaccines at social gatherings such as churches or social functions, this may increase uptake.

Respondents from cluster 1, mostly men, defined by large households and with less educational attainment, were found to have more economic anxieties due to the pandemic and less knowledge about COVID-19 symptoms and were most likely to have gotten the vaccine. They were also the most likely to believe common myths around COVID-19 but have the highest perceived risk of infection. This may be because this cluster of individuals reported being more likely to need to travel for work (a factor in considering themselves at high risk of infection).⁴⁵ They also may hold jobs that require vaccination. Keeping employment by getting vaccinated may have been deemed worth any potential perceived risk of the vaccine, as this cluster also expressed economic concerns related to the pandemic and were responsible for bringing in income to their large households. This is supported by a recent study that found older adults particularly with

chronic illnesses had the highest vaccination rates, and that this group was responsive to messages to increase vaccination.⁴⁶

Individuals in cluster 2, older women who were married or divorced, seem to carry the highest risk of economic hardship and gender-based violence due to the pandemic,^{37–41} so further investigation to vaccinate and support this group is critical. Cluster 2 comprised older women, with higher risks of food insecurity and gender-based violence due to the pandemic.^{37–41} This group had a lower rate of vaccine uptake in relation to their willingness or interest in getting the vaccine expressed in February 2021. This could point to issues around accessibility of the vaccine, especially for women who may have more familial responsibilities and fewer financial and transportation resources. Government assistance and social support interventions may provide a solution, as well as outreach through churches and other venues, to reach women who are unable to travel to facilities and face other challenges in food and economic insecurity and potential violence risks.

By defining archetypes or groups in the population, we can better inform and target policy to improve the efficacy of public health and social support interventions. These clusters can also be used to inform future modelling and predictive analysis of the data by providing insight into what characteristics and behaviours define subgroups of interest, particularly in a situation with a novel disease such as COVID-19 where a lot is unknown and where no prior information is available to inform messaging or policy. These are major strengths to this statistical approach as it is an efficient way to let the data guide the analysis without potential bias related to the analysts' preconceived beliefs about the population. Some limitations of this approach include possible changes to the clustering of the data when run multiple times due to the use of a random starting point and challenges in interpreting the data when clearly defined subgroups are not present. Another limitation to note was the issue of social desirability bias that possibly arose during the phone interviews. Respondents may have felt compelled to provide socially acceptable responses rather than responses that reflect their true attitudes and beliefs, which may clarify some of the inconsistencies observed in vaccine acceptance and uptake. It is also important to note that the cohort of respondents are not truly representative of the underlying population but rather a subset that have a mobile phone and an adolescent household member that participated in recent survey rounds through AGI-K and NISITU. We conducted a small analysis (not shown) that found no significant differences by age or gender in attrition, but that overrules wealthier participants were slightly less likely to respond, and that participants in Dandora and Kibera slums were slightly more likely to. It is also important to note that vaccine acceptance was recorded before the vaccine was available to the general public, and that there is a gap between the vaccine acceptance and uptake measures during which time perceptions may have shifted.

Overall, respondents in our sample of residents of five informal settlements in Nairobi had higher vaccination rates reported than Nairobi as a whole (nearly 75% compared with the 52% reported for the city⁴⁷) as of March 2022. Of the unvaccinated participants, about half reported interest in receiving the vaccine. This suggests that with additional access and messaging almost all individuals can be vaccinated. We also found that most respondents had received more than one dose, although about 1 in 10 had only received the first dose, suggesting additional outreach is needed to make sure everyone is fully vaccinated. As vaccine immunity wanes and new variants emerge, continued messaging, new vaccinations, and uptake of other non-pharmaceutical interventions to prevent transmission will be critical.^{48–49} Studies to understand how to improve governance to increase vaccination and to determine optimal levels of vaccination are important to inform policy.^{50–52} K-Means clustering may be a useful statistical tool when survey data are available to rapidly understand variation in the population and to highlight different potential approaches to messaging and outreach. This paper summarises our methodology and results to provide a starting point for more investigation into targeted vaccination strategies.

CONCLUSION

Machine learning techniques, such as K-Means clustering, are useful to investigate the factors that may predict behaviours related to disease prevention and mitigation. By letting the data guide the analysis and identifying naturally occurring subgroups, we identified characteristics associated with vaccine hesitancy and vaccine uptake, useful for informing policies and messages to target different vulnerable groups within a population. Our results highlight that the highest risk individuals (cluster 1) are most likely to get vaccinated, but that younger, more educated respondents (cluster 3) may require additional messaging and persuasion. One group identified (cluster 2) faced many different challenges and barriers to vaccination and in economic security, food security and risk of violence. This group may require more ways to access the vaccine and may require additional access to social support systems. Based on the results of this study, K-Means clustering may be a useful tool to explore to better identify and target vulnerable groups in public health policy at a national and global level. Although this study primarily focused on vaccine acceptance and uptake, these methods can be applied to a wide range of public health behaviours in future use.

Author affiliations

¹Independent Consultant, Atlanta, Georgia, USA

²Social and Behavioral Sciences Research, Population Council, New York, New York, USA

³Harvard University HSPH, Boston, Massachusetts, USA

⁴Health Economics Research Unit, Centre for Geographic Medicine Research Coast, Nairobi, Kenya

⁵Population Council Kenya, Nairobi, Kenya

Twitter Jessie Pinchoff @JessiePinchoff

Acknowledgements The authors acknowledge all of the work done by the Population Council field team to collect these surveys.

Contributors NR conceptualised the project, conducted the data analysis and led development and writing of the manuscript. JP conceptualised the project and supported development and writing of the manuscript. CBB developed and led the data analysis and review of the manuscript. EB and TA supported with conceptualisation of the project, interpretation of results and review of the manuscript. EM, DM and FM supported with data collection, project management, data cleaning and interpretation of results, including review of the manuscript. KA managed the project and data collection, supported with interpretation of results and review of the manuscript. KA is responsible for the overall content as the guarantor.

Funding This study was funded by UK Department for International Development through Innovations for Poverty Action Peace & Recovery COVID-19 rapid response grant (grant MIT0019-X15).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval This study was approved by The Population Council IRB (p936) and AMREFSRC (P803/2020). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. <https://dataverse.harvard.edu/dataset.xhtml?sessionid=438edef13da12805ee8f2a5d7a9d438edef13da12805ee8f2a5d7a9d?persistentId=doi%3A10.7910%2F2FVN%2FV07SU0&version=&q=&fileTypeGroupFacet=&fileAccess=&fileSortField=date>

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Jessie Pinchoff <http://orcid.org/0000-0003-3155-595X>

Timothy Abuya <http://orcid.org/0000-0001-8815-8299>

Karen Austrian <http://orcid.org/0000-0001-5464-7908>

REFERENCES

- Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta Biomedica*, 2020.
- Corburn J, Vlahov D, Mberu B, *et al.* Slum health: arresting COVID-19 and improving well-being in urban informal settlements. *J Urban Health* 2020;97:348–57.
- Mukumbang FC, Ambe AN, Adebiyi BO. Unspoken inequality: how COVID-19 has exacerbated existing Vulnerabilities of asylum-seekers, refugees, and Undocumented migrants in South Africa. *Int J Equity Health* 2020;19:141.
- Stanturf JA, Goodrick SL, Warren ML, *et al.* Social vulnerability and Ebola virus disease in rural Liberia. *PLoS One* 2015;10:e0137208.
- Ahmed SAKS, Ajisola M, Azeem K, *et al.* Impact of the societal response to COVID-19 on access to Healthcare for non-COVID-19 health issues in slum communities of Bangladesh, Kenya, Nigeria and Pakistan: results of pre-COVID and COVID-19 Lockdown Stakeholder engagements. *BMJ Glob Health* 2020;5:e003042.
- Shah H. COVID-19 recovery: science isn't enough to save us. *Nature* 2021;591:503.
- Rouw A, Wexler A, Kates J, *et al.* Tracking global COVID-19 vaccine equity. *Kaiser Family Foundation* 2021.
- Lazarus JV, Abdool Karim SS, van Selin L, *et al.* COVID-19 vaccine wastage in the midst of vaccine inequity: causes, types and practical steps. *BMJ Glob Health* 2022;7:e009010.
- Harman S, Erfani P, Goronga T, *et al.* Global vaccine equity demands Reparative justice — not charity. *BMJ Glob Health* 2021;6:e006504.
- Khosla R, Gruskin S. Equity without human rights: a false COVID-19 narrative. *BMJ Glob Health* 2021;6:e006720.
- World Health Organization. n.d. WHO Coronavirus (COVID-19) dashboard.
- Kenya Ministry of Health. Health Ministry calls on Kenyans to go for COVID-19 jab. n.d. Available: <https://www.kenyanews.go.ke/health-ministry-calls-on-kenyans-to-go-for-covid-19-jab>
- Adepoju P. Kenya mandates COVID-19 vaccines for civil servants as Africa's vaccine Rollout gathers speed. *Health Policy Watch Independent Global Health Reporting* 2021.
- Wasike A. Kenyan teachers given 7 days to get COVID vaccine or face punishment. *Andalou Agency* 2021.
- Fick M. Kenya's COVID-19 vaccine mandate draws praise and criticism. *Reuters* 2021.
- Human Rights Watch. Kenya: vaccine requirements violate rights. 2021.
- Ackah BBB, Woo M, Stallwood L, *et al.* COVID-19 vaccine hesitancy in africa: a scoping review. *Glob Health Res Policy* 2022;7.
- Sallam M, Al-Sanafi M, Sallam M. A global map of covid-19 vaccine acceptance rates per country: an updated concise narrative review. *JMDH* 2022;Volume 15:21–45.
- Lazarus JV, Wyka K, White TM, *et al.* Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. *Nat Commun* 2022;13:3801.
- Ackah M, Ameyaw L, Gazali Salifu M, *et al.* COVID-19 vaccine acceptance among health care workers in Africa: A systematic review and meta-analysis. *PLoS One* 2022;17:e0268711.
- Detoc M, Bruel S, Frappe P, *et al.* Intention to participate in a COVID-19 vaccine clinical trial and to get vaccinated against COVID-19 in France during the pandemic. *Vaccine* 2020;38:7002–6.
- Orangi S, Pinchoff J, Mwanga D, *et al.* Assessing the level and determinants of COVID-19 vaccine confidence in Kenya. *Vaccines (Basel)* 2021;9:936.
- Bono SA, Faria de Moura Villela E, Siau CS, *et al.* Factors affecting COVID-19 vaccine acceptance: an international survey among Low- and middle-income countries. *Vaccines* 2021;9:515.
- Caserotti M, Girardi P, Rubatelli E, *et al.* Associations of COVID-19 risk perception with vaccine hesitancy over time for Italian residents. *Social Science & Medicine* 2021;272:113688.
- Ackah BBB, Woo M, Ukah UV, *et al.* COVID-19 vaccine hesitancy in africa: a scoping review. *In Review [Preprint]* 2021.
- Osuri J, Muinga E, Carter J, *et al.* COVID-19 vaccine hesitancy: vaccination intention and attitudes of community health volunteers in Kenya. *PLOS Glob Public Health* 2022;2:e0000233.
- Nasimiyu C, Ngere I, Dawa J, *et al.* Near-complete SARS-Cov-2 Seroprevalence among rural and urban Kenyans despite significant vaccine hesitancy and refusal. *Vaccines* 2023;11:68.
- Lazarus JV, Wyka K, White TM, *et al.* A survey of COVID-19 vaccine acceptance across 23 countries in 2022. *Nat Med* 2023;29:366–75.
- COVID 19 vaccine perceptions: A 15 country study. 2021.
- Quaife M, van Zandvoort K, Gimma A, *et al.* The impact of COVID-19 control measures on social contacts and transmission in Kenyan informal settlements. *BMC Med* 2020;18:316.
- Hudson A, Montelpare WJ. Predictors of vaccine hesitancy: implications for COVID-19 public health Messaging. *Int J Environ Res Public Health* 2021;18:8054.
- Abbasi J. Widespread misinformation about infertility continues to create COVID-19 vaccine hesitancy. *JAMA* 2022;327:1013.
- Ainslie D, Ogwuru C, Sinclair R. Coronavirus and vaccine hesitancy, great Britain. 2021.
- Blandi L, Sabbatucci M, Dallagiacomma G, *et al.* Digital information approach through social media among Gen Z and Millennials: the global scenario during the COVID-19 pandemic. *Vaccines (Basel)* 2022;10:1822.
- Ahlqvist E, Storm P, Kärjämäki A, *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–9.
- Coccia M. Improving preparedness for next Pandemics: Max level of COVID-19 Vaccinations without social impositions to design effective health policy and avoid flawed democracies. *Environ Res* 2022;213:S0013-9351(22)00893-3.
- Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. *The Lancet* 2020;395:846–8.

- 38 John N, Casey SE, Carino G, *et al.* Lessons never learned: crisis and gender-based violence. *Developing World Bioeth* 2020;20:65–8. 10.1111/dewb.12261 Available: <https://onlinelibrary.wiley.com/toc/14718847/20/2>
- 39 Gould C. Gender-based violence during Lockdown - looking for answers [analysis] market watch. n.d. Available: <https://www.marketwatch.com/press-release/gender-based-violence-during-lockdown---looking-for-answers-analysis-2020-05-12>
- 40 Azcona G, Bhatt A, Robert N. COVID-19 exposes the harsh realities of gender inequality in slums. n.d. Available: <https://data.unwomen.org/features/covid-19-exposes-harsh-realities-gender-inequality-slums>
- 41 Pinchoff J, Austrian K, Rajshekhar N, *et al.* Gendered economic, social and health effects of the COVID-19 pandemic and mitigation policies in Kenya: evidence from a prospective cohort survey in Nairobi informal settlements. *BMJ Open* 2021;11:e042749.
- 42 Davis TP, Yimam AK, Kalam MA, *et al.* Behavioural determinants of COVID-19-vaccine acceptance in rural areas of six lower-and middle-income countries. *Vaccines (Basel)* 2022;10:214.
- 43 Limaye RJ, Balgobin K, Michel A, *et al.* What message appeal and messenger are most persuasive for COVID-19 vaccine uptake: results from a 5-country survey in India. *PLoS One* 2022;17:e0274966.
- 44 Diaz P, Reddy P, Ramasahayam R, *et al.* COVID-19 vaccine hesitancy linked to increased Internet search queries for side effects on fertility potential in the initial Rollout phase following emergency use authorization. *Andrologia* 2021;53:e14156.
- 45 Pinchoff J, Kraus-Perrotta C, Austrian K, *et al.* Mobility patterns during COVID-19 travel restrictions in Nairobi urban informal settlements: who is leaving home and why. *J Urban Health* 2021;98:211–21.
- 46 Yego J, Korom R, Eriksson E, *et al.* A comparison of strategies to improve uptake of COVID-19 vaccine among high-risk adults in Nairobi, Kenya in 2022. *Vaccines (Basel)* 2023;11:209.
- 47 Report of the OAG, as at 31 March, 2022, of COVID-19 vaccine roll out for Nairobi city.
- 48 Coccia M. Sources, diffusion and prediction in COVID-19 pandemic: lessons learned to face next health emergency. *AIMS Public Health* 2023;10:145–68.
- 49 Dobrev Z, Gimma A, Rohan H, *et al.* Characterising social contacts under COVID-19 control measures in Africa. *BMC Med* 2022;20:344.
- 50 Coccia M. COVID-19 vaccination is not a sufficient public policy to face crisis management of next pandemic threats. *Public Organiz Rev* 2022.
- 51 Coccia M. Optimal levels of vaccination to reduce COVID-19 infected individuals and deaths: A global analysis. *Environ Res* 2022;204:S0013-9351(21)01615-7.
- 52 Benati I, Coccia M. Global analysis of timely COVID-19 Vaccinations: improving governance to reinforce response policies for pandemic crises. *IJHG* 2022;27:240–53.