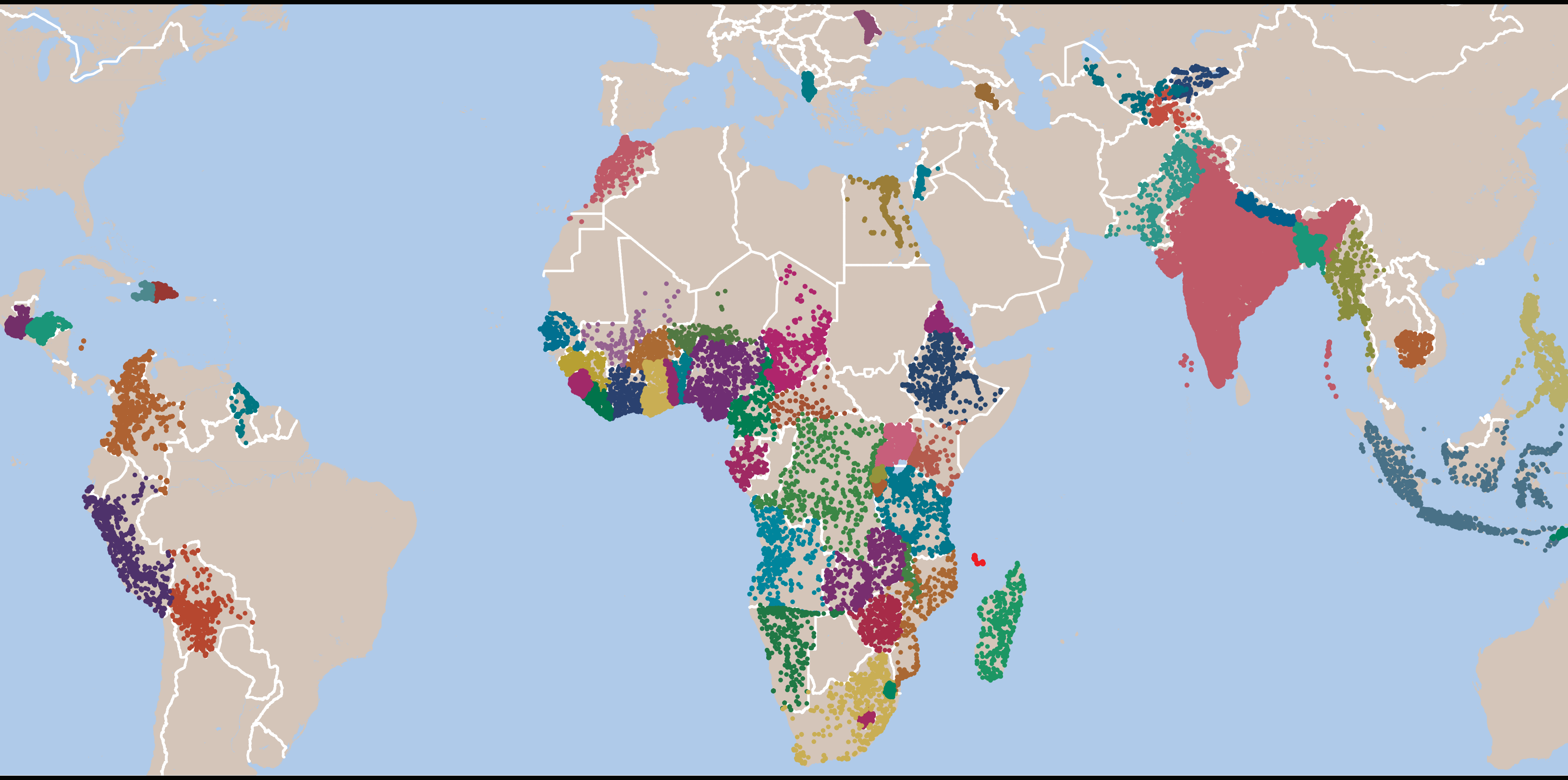
USAID
FROM THE AMERICAN PEOPLE

MODELING COVID-19 RELATED INDICATORS AT THE SECOND SUBNATIONAL ADMINISTRATIVE LEVEL

DHS SPATIAL ANALYSIS REPORTS 22



SEPTEMBER 2022

This publication was produced for review by the United States Agency for International Development (USAID). The report was prepared by Rose E. Donohue, Martha Medina, Bradley Janocha, Benjamin K. Mayala, and Trevor N. Croft.

DHS Spatial Analysis Reports No. 22

Modeling COVID-19 Related Indicators at the Second Subnational Administrative Level

Rose E. Donohue[1]
Martha Medina[1]
Bradley Janocha[1]
Benjamin K. Mayala[1]
Trevor N. Croft[1]

The DHS Program
ICF
Rockville, Maryland, USA

September 2022

[1] The DHS Program, ICF

Corresponding author:
Rose E. Donohue and Benjamin K. Mayala, The DHS Program, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-572-0507; fax: 301-407-6501; email: Rose.Donohue@icf.com and Ben.Mayala@icf.com

Acknowledgments: The authors wish to thank Blake Zachary, Samir Bhatt, and Edson Utazi for their review of this report.

Editor: Diane Stoy
Document Production: Chris Gramer

This study was implemented with support from the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about The DHS Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-407-6500, fax: +1 301-407-6501, email: info@DHSprogram.com, Internet: www.DHSprogram.com.

Recommended citation:

Donohue, Rose E., Martha Medina, Bradley Janocha, Benjamin K. Mayala, and Trevor N. Croft. 2022. *Modeling COVID-19 Related Indicators at the Second Subnational Administrative Level*. DHS Spatial Analysis Reports No. 22. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES v
FIGURES vii
PREFACE ix
ABSTRACT xi
ACRONYMS AND ABBREVIATIONS xiii

1 BACKGROUND AND OBJECTIVES 1
1.1 Background 1
1.2 Objectives 2
1.3 Report Structure 2

2 DHS SURVEYS AND INDICATORS 3

3 METHODS 5
3.1 Overview 5
3.1.1 Classification of data into pre-March 2020 and post-March 2020 groups 5
3.1.2 Evaluation of percentage change in basic handwashing between pre-March 2020 and post-March 2020 samples 5
3.1.3 Evaluation of systematic differences between pre-March 2020 and post-March 2020 samples 6
3.1.4 Evaluation of correlation between districts' COVID-19 risk status and the observed change in access to basic handwashing facilities 6
3.2 Geospatial Covariates 6
3.3 Geostatistical Model 7
3.3.1 Overview of the modeling approach 7
3.3.2 Covariate modeling using stacked generalization 9
3.3.3 Model specification and development 9
3.3.4 Pixel-level model estimates 11
3.3.5 Model estimates at administrative level 2 11
3.3.6 Model validation 11

4 RESULTS 13
4.1 Changes in Basic Handwashing Modeled Admin 2 Estimates from Pre-March 2020 to Post-March 2020 13
4.1.1 Rwanda 13
4.1.2 India 14
4.2 Evaluation of Systematic Differences between Pre-March 2020 and Post-March 2020 Samples 17
4.2.1 Rwanda 17
4.2.2 India 18
4.3 Evaluation of Correlation between COVID-19 Risk Factors and Change in Basic Handwashing Facilities 20
4.3.1 Rwanda 20
4.3.2 India 20
4.4 Model Validation Metrics 21

5 DISCUSSION AND CONCLUSION 23

REFERENCES 25

APPENDIX 31

TABLES

Table 1	Description of DHS indicators used in the study	3
Table 2	Geospatial covariates used to develop the models in this study	7
Table 3	Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before (pre) and after (post) the pandemic onset (March 2020), Rwanda 2019–20 DHS	14
Table 4	Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before (pre) and after (post) the pandemic onset (March 2020), India 2020 DHS	16
Table 5	Results of the evaluation of systematic differences in wealth index (WI) and residence (urban/rural) between the pre-March 2020 and post-March 2020 samples for Admin 2s, Rwanda	17
Table 6	Results of the evaluation of systematic differences in wealth index (WI) and residence (urban/rural) between the pre-March 2020 and post-March 2020 samples, India	19
Table 7	Predictive metrics for each indicator aggregated at admin 2, Rwanda	21
Table 8	Predictive metrics for each indicator aggregated at admin 2, India	21
Appendix Table 1	Modeled Admin 2 estimates of the percentage of overcrowded households data, Rwanda 2019–20 DHS	31
Appendix Table 2	Modeled Admin 2 estimates of the percentage of overcrowded households in districts surveyed both before and during the COVID-19 pandemic, India 2019–21 DHS	32
Appendix Table 3	Modeled Admin 2 estimates of the percentage of households with at least one member age 65+, Rwanda 2019–20 DHS	33
Appendix Table 4	Modeled Admin 2 estimates of the percentage of households with at least one member age 65 or older in districts surveyed both before and during the COVID-19 pandemic, India 2019–21 DHS	34

FIGURES

Figure 1 Geospatial modeling process flowchart 8

Figure 2 Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before and after the pandemic onset (March 2020) using data from the Rwanda 2019–20 DHS: (a) Admin 2 estimates using data from clusters surveyed before the pandemic started; (b) Admin 2 estimates using data from clusters surveyed after the pandemic started; and (c) percentage change from (a) to (b) 13

Figure 3 Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before and after the pandemic onset (March 2020) using data from the India 2019–21 DHS: (a) Admin 2 estimates using data from clusters surveyed before the pandemic started; (b) Admin 2 estimates using data from clusters surveyed after the pandemic started; and (c) percentage change from (a) to (b) 15

Figure 4 Comparison of the percentage change in basic handwashing from pre-March 2020 to post-March 2020 in Rwandan districts to the (a) percentage of overcrowded households and (b) percentage of households with at least one member age 65+ 20

Figure 5 Comparison of the percentage change in basic handwashing from pre-March 2020 to post-March 2020 in Indian districts to the (a) percentage of overcrowded households and (b) percentage of households with at least one member age 65+ 21

Appendix Figure 1 Comparison of in-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Aged 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, Rwanda 35

Appendix Figure 2 Comparison of out-of-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Aged 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, Rwanda 36

Appendix Figure 3 Comparison of in-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Aged 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, India 37

Appendix Figure 4 Comparison of out-of-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Aged 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, India 38

PREFACE

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

The DHS Spatial Analysis Reports supplement the other series of DHS reports that respond to the increasing interest in a spatial perspective on demographic and health data. The principal objectives of all the DHS report series are to provide information for policy formulation at the international level and to examine individual country results in an international context.

The topics in this series are selected by The DHS Program in consultation with the U.S. Agency for International Development. A range of methodologies are used, including geostatistical and multivariate statistical techniques.

It is hoped that the DHS Spatial Analysis Reports series will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

ABSTRACT

The COVID-19 pandemic has been a major burden to global public health in recent years, and has caused substantial morbidity and mortality. At the onset of the pandemic, handwashing was recognized as critically important to reducing the spread of COVID-19. Most countries implemented initiatives to improve access to handwashing during the pandemic. The COVID-19 pandemic also affected fieldwork in the Demographic and Health Surveys Program, and disrupted data collection in the Rwanda 2019–20 DHS and the India 2019–21 DHS. In this report, we use these unique datasets to explore the change in access to a basic handwashing facility using data collected before fieldwork was interrupted by the COVID-19 pandemic (pre) and after fieldwork resumed (post). We use a Bayesian geospatial modeling approach to estimate basic handwashing access at the second subnational administrative level (Admin 2), which is called a district in Rwanda and India. To assess the impact of systematic differences, we also compared the wealth and urban/rural distribution between the pre and post samples. We quantified the risk of COVID-19 transmission and mortality at the Admin 2 level by modeling the percentage of overcrowded households and the percentage of households with at least one member age 65 or older as proxy indicators. We evaluated the association between these risk factors and the percentage change in access to basic handwashing from the pre sample to the post sample at the Admin 2 level.

The study found that basic handwashing access improved in 10 of the 13 districts evaluated in Rwanda, and 47 of the 49 districts evaluated in India. The average percentage increase in the 10 Rwandan districts was 27.7%, and the average percentage increase in the 47 Indian districts was 20.5%. We did not identify any systematic differences in wealth or the proportion of urban clusters between the pre and post samples. We did not identify a significant association between the percentage change in basic handwashing access experienced by districts and the districts' COVID-19 risk, as measured by both the percentage of overcrowded households and the percentage of households with at least one member age 65 or older. Although this study was not designed to assess causative factors for any changes in basic handwashing access, we identified a positive trend in basic handwashing access. Further research and analysis could evaluate this trend and the causative factors underlying any changes in basic handwashing access. The improvement in handwashing has broader implications beyond COVID-19, because handwashing reduces the spread of respiratory and diarrheal diseases. By conducting this study at the policy-relevant Admin 2 level, the COVID-19-related factors modeled in this study could be used by policymakers and program planners to evaluate their COVID-19 response and adapt preventive measures in the future.

ACRONYMS AND ABBREVIATIONS

Admin 1	first subnational administrative level
Admin 2	second subnational administrative level
DHS	Demographic and Health Survey
GBM	gradient boosted trees
GPS	global positioning system
INLA	integrated nested Laplace approximation
LASSO	least absolute shrinkage and selection operator
LST	land surface temperature
MBG	model-based geostatistics
RMSE	root mean square error
SAR	spatial analysis report
SDG	Sustainable Development Goal
SPDE	stochastic partial differential equations
WASH	water, sanitation, and hygiene
WHO	World Health Organization
UNICEF	United Nations Children's Fund

1 BACKGROUND AND OBJECTIVES

1.1 Background

The COVID-19 pandemic has been a major burden to global public health in recent years, and has caused substantial morbidity and mortality. Between January 2020 and December 2021, 5.9 million deaths attributed to COVID-19 were reported globally (Wang et al. 2022). However, this figure is likely to under-represent the actual death toll, particularly in areas that lack strong civil registration and vital statistics systems. Recent modeling studies have produced estimates of 14.9 million excess deaths (Knutson et al. 2022) and 18.2 million excess deaths globally (Wang et al. 2022).

The World Health Organization (WHO) officially declared COVID-19 a global pandemic in March 2020. Shortly after this declaration, technical guidance on water, sanitation, and hygiene (WASH), and waste management in relation to viruses, such as COVID-19, was released. This guidance emphasized handwashing as "extremely important" to preventing COVID-19 transmission (WHO and UNICEF 2020). Water and soap, or alcohol-based hand rub, were recommended as the most effective hand hygiene materials.

While access to a basic handwashing facility, with soap and water available, is of critical importance to slowing the transmission of coronavirus, such access is not universal in many countries. A recent modeling study estimated that 26.1% of the global population lacked access to a handwashing facility with soap and water, and more than 50% lacked access in sub-Saharan Africa (Brauer et al. 2020). Recognizing the critical importance of WASH, most countries implemented initiatives to improve access to WASH services during the pandemic (Giné-Garriga et al. 2021). More than 94% of the 84 countries evaluated by Giné-Garriga et al. (2021) promoted handwashing with soap.

The COVID-19 pandemic also affected the fieldwork for surveys conducted by The Demographic and Health Surveys (DHS) Program. Two of the countries in which DHS data collection was ongoing, Rwanda and India, had fieldwork interrupted by the pandemic. As a result, the datasets for these countries include households that were surveyed before the pandemic's onset and during the pandemic. Given the unique datasets caused by the interruption of fieldwork, the analyses in this report focus on these two countries.

Rwanda and India were both affected by the COVID-19 pandemic, albeit to varying degrees. A recent modeling study found that while Rwanda had 1,350 reported COVID-19 deaths, the model estimated 21,900 excess deaths during 2020–2021. India had reported 489,000 COVID-19 deaths, while the model estimated 4,070,000 excess deaths during 2020–2021 (Wang et al. 2022).

Rwanda and India had varying degrees of preventive measures already in place. In the most recent DHS surveys prior to the surveys conducted during the pandemic, only 5.6% of households had a basic handwashing facility in Rwanda (2014–15 DHS). India had a higher degree of preparedness, with 58.7% of households with a basic handwashing facility (2015–16 DHS). However, India remained far from the global target of 100%, which is included in the Sustainable Development Goal (SDG) targets for 2030 (United Nations 2018). When compared to the estimates from the COVID-19-impacted DHS surveys, both Rwanda and India had increased access to basic handwashing facilities from their previous DHS surveys.

1.2 Objectives

Within this context, we were interested in understanding if access to a basic handwashing facility improved by comparing clusters surveyed before the COVID-19 pandemic interrupted fieldwork in March 2020 to those surveyed after fieldwork resumed. In recent years, The DHS Program has been producing estimates of DHS indicators at the second subnational administrative level (Admin 2) (Janocha et al. 2021; Mayala et al. 2019; Mayala et al. 2020). The Admin 2 level estimates are policy-relevant because health decision-making and program implementation are decentralized and often occur at the Admin 2 level. In this report, we applied the methodology employed in earlier reports to estimate COVID-related indicators at the Admin 2 level, which are referred to as districts in both Rwanda and India.

Specifically, we explore the changes in the percentage of households with a basic handwashing facility, including soap and water, in districts surveyed both before and during the COVID-19 pandemic. We produce estimates of the basic handwashing indicator at the Admin 2 level using the model-based geostatistics methodology described in Mayala et al. 2019. By focusing on districts where fieldwork was stopped due to the COVID-19 pandemic and then resumed, we compare the district estimates from households surveyed before the pandemic officially began (March 2020) to households surveyed during the pandemic (after March 2020). We also evaluate if there are any systematic differences in the urban/rural distribution or the asset-based wealth of the clusters surveyed before and during the pandemic. Finally, we model two measures of COVID-19 risk, the percentage of overcrowded households and the percentage of households with at least one member age 65 or older, and we assess if the percentage change observed in basic handwashing was correlated with a district's COVID-19 risk level.

1.3 Report Structure

The remainder of this report describes the analysis conducted to address the objectives, presents results, and discusses the implications. In Section 2, we describe the DHS surveys and indicators we used, and in Section 3 we describe the methodology. We present the results in Section 4 and discuss the implications in Section 5.

2 DHS SURVEYS AND INDICATORS

We used two recent DHS datasets that had fieldwork interrupted by the COVID-19 pandemic: the Rwanda 2019–20 DHS and the India 2019–21 DHS. These two datasets had districts with clusters surveyed before the COVID-19 pandemic started (defined as March 2020) and after the COVID-19 pandemic started. These datasets provide a unique perspective to assess how COVID-19 related indicator estimates changed within the same district over that time.

In this report, we produce Admin 2 estimates for three DHS indicators (Table 1): percentage of households with a basic handwashing facility, the percentage of households with three or more people per sleeping room, and the percentage of households with at least one household member age 65 or older.

Table 1 Description of DHS indicators used in the study

Indicator	Definition
Basic handwashing facility	Percentage of households with a basic handwashing facility, with soap and water available
Overcrowding	Percentage of households where the de jure population divided by the number of sleeping rooms is greater than or equal to 3
Members 65+	Percentage of households with at least one member age 65 or older

The main outcome we modeled was the indicator "percentage of households with a basic handwashing facility." For this indicator, households must meet the following conditions: (1) a handwashing facility must be observed on the premises; (2) the handwashing facility must have water; and (3) the handwashing facility must have soap (WHO and UNICEF 2018). A household that did not meet all three conditions was classified as not having a basic handwashing facility. Although DHS surveys also query if ash and other alternatives are available at the handwashing facility in addition to soap, we chose not to categorize households with ash present as a positive alternative to soap based on the review from Paludan-Müller et al. (2020). This review concluded that the benefits and harms of handwashing with ash compared with soap to reduce the spread of infection were uncertain. Therefore, we did not classify the presence of ash as a 'positive' outcome that would reduce the transmission of coronavirus.

For the overcrowding indicator, we chose a threshold value of three people per sleeping room after reviewing the available definitions. UN-Habitat defines overcrowding as more than three people per habitable room (WHO Housing and Health Guidelines 2018). However, the definition of a habitable room is broader than the sleeping rooms that DHS surveys measure. Other available overcrowding thresholds include 2.5 or more persons per bedrooms from Chile (Bilal et al. 2017), more than 2 persons per bedroom from public housing authorities in the United States (Blake et al. 2007), and more than 2.5 persons per sleeping room in the Disease Outbreak Resilience Index (Koomson et al. 2022).

3 METHODS

3.1 Overview

First, we divided the data into a pre sample and a post sample, following the methodology described in Section 3.1.1. We then evaluated the percentage change in basic handwashing from pre to post by modeling the indicator of basic handwashing for each sample to obtain a district estimate from the pre sample and from the post sample, as described in Section 3.1.2. We then calculated the percentage change from the pre district estimate to the post district estimate. In Section 3.1.3, we describe our evaluation of the differences in wealth or urban/rural status between the pre and post samples. Finally, we modeled two indicators of COVID-19 risk and assessed if the percentage change in handwashing experienced in a district was associated with the risk status of the district, as described in Section 3.1.4.

We employed the same methodology to model the Admin 2 estimates for all three indicators included in this report. The details of this methodology are presented in Section 3.3.

3.1.1 Classification of data into pre-March 2020 and post-March 2020 groups

First, we classified clusters into those surveyed before the start of the COVID-19 pandemic, subsequently referred to as "pre," and those surveyed after the start of the COVID-19 pandemic, subsequently referred to as "post." For Rwanda, pre clusters included those with household interviews that took place between November 2019 and March 2020, while post clusters were interviewed between June and July 2020. For India, pre clusters included those interviewed between June 2019 and March 2020, while post clusters were interviewed between October 2020 and May 2021. Clusters where fieldwork was stopped before completing each household in a cluster were removed from the analysis (n = 13 in Rwanda and n = 5 in India).

3.1.2 Evaluation of percentage change in basic handwashing between pre-March 2020 and post-March 2020 samples

We modeled the indicator percentage of households with a basic handwashing facility separately for the pre and post samples. Since we were interested in evaluating changes in the district-level estimates, we only compared districts with both pre and post clusters. We required at least 3 clusters within a district to be interviewed in both the pre and post samples to be included as a district in the analysis. No districts were dropped in Rwanda, while 27 of the 76 districts were dropped in India. Rwanda had a total of 13 districts with both pre and post data, and India had a total of 49 districts used in the analysis.

We computed the percentage change in the district estimate from the pre sample to the post sample with the following equation:

$$\frac{(post\ \% - pre\ \%)}{pre\ \%}\, x\ 100$$

3.1.3 Evaluation of systematic differences between pre-March 2020 and post-March 2020 samples

We evaluated additional factors to explore any systematic differences that might exist between the pre and post samples that could lead to differences in outcomes. Here, we specifically focused on the distribution of urban and rural clusters, as well as the household wealth index. We explored the distribution of urban and rural clusters, because urban areas generally have greater access to basic handwashing facilities than rural areas (Brauer et al. 2020). We also evaluated if the post sample was wealthier overall because lower socioeconomic status is associated with reduced WASH access and inadequate handwashing practices (Roche et al. 2017; Smith et al. 2021).

To evaluate the urban and rural distribution, we calculated the proportion of urban clusters in the pre sample to the post sample for each district. We conducted a Fisher's exact test to statistically evaluate if the proportion of clusters that were urban in the pre sample differed from the proportion of clusters that were urban in the post sample. To evaluate if households had significantly higher wealth index values in the post samples compared to the pre samples, we conducted a Mann-Whitney U test for each district to test how the wealth index scores among households sampled in pre clusters compared to the wealth index scores among households sampled in post clusters within each district.

3.1.4 Evaluation of correlation between districts' COVID-19 risk status and the observed change in access to basic handwashing facilities

We also modeled Admin 2 estimates for two COVID-19 risk factors using the 2019–20 Rwanda DHS and the 2019–21 India DHS. We modeled the risk of coronavirus transmission using the percentage of overcrowded households as the proxy indicator. We chose this indicator because household overcrowding is associated with an increased risk of coronavirus transmission (Acharya and Porwal 2020; Ghosh et al. 2021). We also modeled the risk of COVID-19 mortality using the percentage of households with at least one member age 65 or older as the proxy indicator. We chose this indicator because there has been a clear association between age and COVID-19 mortality, with older individuals experiencing higher mortality than younger individuals (Grasselli et al. 2020; Lippi et al. 2020). After obtaining the Admin 2 estimates for these indicators, we assessed the association between each risk factor and the percentage change in handwashing for a district that was calculated in Section 3.1.2. We performed a correlation analysis and computed the Spearman's rank correlation coefficient and associated p-value.

3.2 Geospatial Covariates

To model the DHS indicators used in this study, we assembled geospatial covariates datasets, which were obtained from publicly available sources. The geospatial covariates were selected for their potential to predict DHS indicators, and for previously having been shown to correlate with the development of indicators in different settings (Alegana et al. 2015; Gething et al. 2015; Osgood-Zimmerman et al. 2018). Table 2 describes the geospatial covariates.

The covariate data layers used in this analysis were acquired from a variety of data sources, and have different spatial references, projections, extents, and dimensions. Therefore, a spatial processing was required, which involved:

1) re-projecting to the same coordinate reference system (the standard-based World Geodetic System 1984);

2) masking to an extent that encompassed the boundaries of the study area; and

3) resampling to the same (5 x 5 km) spatial resolution used in the modeling.

For the population covariate, we resampled by taking the sum. Covariates that were produced at a 5 × 5 km resolution did not require additional processing. For the other covariates, we resampled using the bilinear interpolation method. The covariates processing was done in R software using the 'raster' and 'shapefiles' packages (R Core Team 2022).

Table 2 Geospatial covariates used to develop the models in this study

Covariates	Spatial resolution	Source
Travel time to nearest settlement >50,000 inhabitants	5 x 5 km	Malaria Atlas Project
Aridity	10 x 10 km	Climatic Research Unit gridded Time Series (CRUTS)
Diurnal temperature range	10 x 10 km	CRUTS
Potential evapotranspiration (PET)	10 x 10 km	CRUTS
Daily maximum temperature	10 x 10 km	CRUTS
Elevation	1 x 1 km	NOAA
Enhanced vegetation index (EVI)	5 x 5 km	NASA
Daytime land surface temperature (LST)	5 x 5 km	NASA
Diurnal difference in LST	5 x 5 km	NASA
Nighttime LST	5 x 5 km	NASA
Population distribution	1 x 1 km	WorldPop

3.3 Geostatistical Model

In this report, we modeled three indicators (described in Table 1) using the modeling approach described in this section.

3.3.1 Overview of the modeling approach

Figure 1 provides a conceptual overview of the geospatial modeling framework used for modeling DHS indicators and the underlying covariates, and for producing the subnational level estimates. The approach involved the following steps:

Step 1 We summarized the individual-level DHS survey data to the finest spatial resolution (latitude and longitude) that represented the location of the survey cluster.

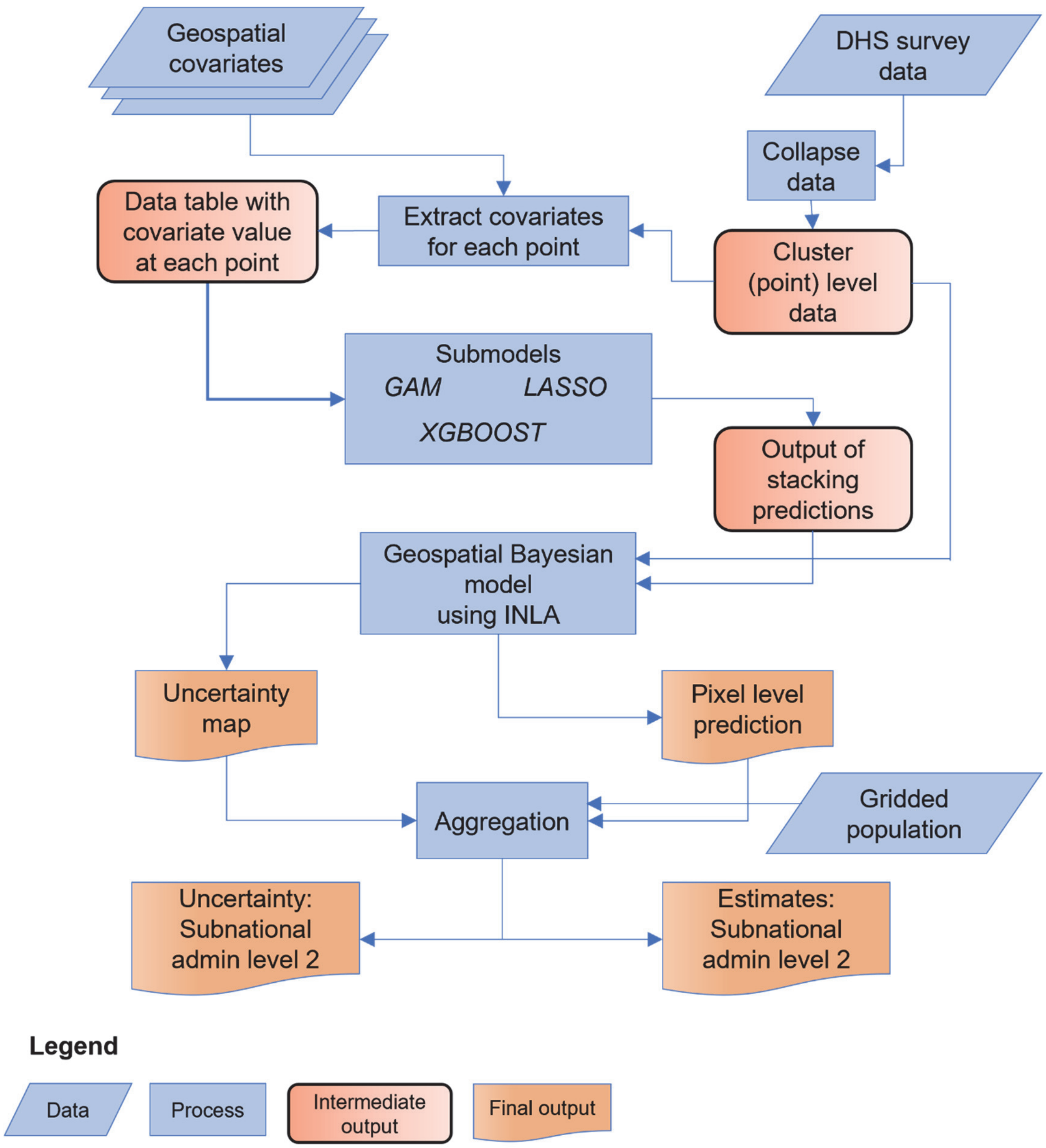
Step 2 The covariates and the cluster (point) level data were imported into the R environment for statistical computing. We then applied the 'raster' package to extract the corresponding covariate pixel values at each survey cluster point.

Step 3 The point level data (from Step 2) and their associated covariates were used in the stacked generalization ensemble model (described in Section 3.3.2). The prediction surfaces generated from the stacked ensemble models were then used as covariates to calibrate the final geospatial

Bayesian model. The outputs of the final model are pixel-level mean estimates with associated uncertainty at the 5 x 5 km resolution.

Step 4 We aggregated the prediction output from the final model (Step 3) to the second subnational administrative level (Admin 2) level.

Figure 1 Geospatial modeling process flowchart



* Modified from Mayala et al. 2019.

3.3.2 Covariate modeling using stacked generalization

In many applications, the generic geospatial modeling is sufficient to produce a highly predictive model. However, when modeling outcomes in which the underlying process is linked to the covariates and demographic parameters through complex non-linearities and interactions, a simple linear mean of the form βX can be insufficient. We therefore use a framework formed from a body of theory known as "stacked generalization" to pre-process the covariates through a set of highly predictive machine learning methods (Breiman 1996; Wolpert 1992). Stacked generalization is a general ensemble modeling approach that combines multiple model algorithmic methods to produce a meta-model that has equal or better predictive performance relative to a single modeling approach.

We employed this approach to capture the potential complex interactions and non-linear effects among the geospatial covariates. The approach improves the predictive accuracy of the geostatistical models, as compared to prediction with any single method (Bhatt et al. 2017). Numerous studies have implemented the stacking approach to derive continuous estimated surfaces of indicators of interest from DHS household surveys. These include mapping of household overcrowding (Chipeta et al. 2022), HIV prevalence (Dwyer-Lindgren et al. 2019), vaccine coverage (Mayala et al. 2019; Mosser et al. 2019), exclusive breastfeeding (Bhattacharjee et al. 2019), child growth failure (Osgood-Zimmerman et al. 2018), education attainment (Graetz et al. 2018), and childhood diarrheal diseases (Reiner et al. 2018).

Our choice of algorithmic methods included (1) GAM: generalized additive model (Wood 2017); (2) LASSO: least absolute shrinkage and selection operator regression (Zou and Hastie 2005); and (3) XGBOOST: gradient boosting (Friedman 2001). We fitted the three algorithmic methods (submodels) to each set of the selected DHS indicator survey data by using the geospatial covariates (described in Table 2) as exploratory predictors. The submodels were implemented in R statistical software for the computing environment by using packages 'caret', 'mgcv', 'xgboost', and 'glmnet' (R Core Team 2022). Covariate selection was performed automatically in the stacked generalization framework, which removed the effect of covariates with little predictive contribution.

To make better predictions and avoid overfitting, each submodel was fit by using five-fold cross-validation, which generated the out-of-sample predictions that were included as exploratory geospatial covariates when fitting the geostatistical model. In addition, each submodel was fit with a full dataset, which produced the in-sample predictions that were then used as covariates when generating predictions from the full geospatial Bayesian model. A logit transformation of the predictions placed the out-of-sample and in-sample predictions on the same scale as the linear predictor in the geostatistical model. This process has been described in detail by Bhatt et al. (2017) and Dwyer-Lindgren et al. (2019).

3.3.3 Model specification and development

As described in the previous section, the stacked generalization ensemble modeling approach allows for non-linear relationships and interactions between the geospatial covariates to better predict the DHS indicators. Since the approach does not explicitly account for spatial patterns in the data, we used the Bayesian geostatistical modeling framework in our analysis to account for the spatial dependence.

For each indicator of interest, we modeled Y_i, the number of 'positive' individuals among those sampled at cluster location $s_i, i = 1, \ldots n$, using a binomial spatial regression with a logit link function (Banerjee, Carlin, and Gelfand 2014; Diggle and Giorgi 2019). If N_i is the total number of individuals sampled at cluster s_i, the model can be written as:

$$Y_i \sim Binomial(N_i, p_i)$$

$$logit(p_i) = \beta_0 + \beta \mathrm{X}_i + \omega_i + \varepsilon_i$$

$$\omega_i \sim GP(0, \Sigma)$$

Where:

- β_0 denotes the intercept,
- p_i is the probability, representing the underlying prevalence at cluster s_i,
- $X_i = \left(X_{i1,}X_{i2}, \dots X_{im}\right)$ is the vector of logit-transformed covariates for location s_i obtained from the submodels (GAM, LASSO, and XGBOOST) generated from the stacked generalization modeling (as described in Section 3.3.2),
- $\beta = (\beta_1, \beta_2, \dots \beta_m)$ vector of regression coefficients on the submodels represent their respective predictive weighting and are constrained to the sum of one (Bhatt et al. 2017),
- ω_i is a correlated spatial error term, accounting for spatial autocorrelation between data points, and
- $\varepsilon_i \sim N\left(0, \sigma_{nug}^2\right)$ is an independent error term known as the nugget effect.

The spatial error term ω_i is modeled as a Gaussian process with a zero-mean and spatially structured covariance matrix $\sum$. The spatial covariance $\sum$ was modeled using a stationary and isotropic Matérn function (Banerjee, Carlin, and Gelfand 2014), given by:

$$\Sigma(s_i, s_j) = \frac{\sigma^2}{\Gamma(\lambda) 2^{\lambda-1}} \left(\kappa d(s_i, s_j)^{\lambda} K_\lambda \left(\kappa d(s_i, s_j) \right) \right)$$

Where $d(s_i, s_j)$ is the distance between the two locations and σ^2 is the spatial process variance. The term K_λ denotes the modified Bessel function of second kind and order λ, which measures the degree of smoothness. Conversely, κ is a scaling parameter related to the range r, which is the distance at which the spatial correlation becomes almost null (smaller than 10%), and the definition for the range is given in equation below. See example by Lindgren (2011) for a detailed description.

$$r = \frac{\sqrt{8\lambda}}{\kappa}$$

The Bayesian geostatistical model analysis was implemented through a stochastic partial differential equations (SPDE) approach in the integrated nested Laplace approximation (INLA) algorithm as applied in the R-INLA package (Rue, Martino, and Chopin 2009). This algorithm provides an effective estimation and spatial prediction strategy for spatial data by specifying a spatial data process, as well as a spatial covariance function depending on the locations and time points at which outcome and covariate data are collected (Rue, Martino, and Chopin 2009). The INLA approach offers the advantage of accurate and fast results as compared to the Markov Chain Monte Carlo algorithms, which have problems of convergence and dense covariate matrices that increase the computational time. Thus, for large datasets, spatial and spatiotemporal estimation could require several days of computing time (Blangiardo and Cameletti 2015; Cameletti et al. 2012; Rue, Martino, and Chopin 2009).

The SPDE approach allows us to define a grid on spatial data by creating a constrained refined Delaunay triangulation (usually called mesh) over the study region. The mesh needs to cover the region of study and an outer extension to avoid boundary effects, which would increase the variance near the boundary. To fit a mode with this approach, observations are treated as initial vertices for the triangulation. Further vertices are then added or removed to satisfy triangulation quality constraints defined by three parameters: (1) mesh offset, (2) maximum edge, and (3) cutoff (Blangiardo and Cameletti 2015; Cameletti et al. 2012; Rue, Martino, and Chopin 2009).

We specified a cutoff value to avoid building too many small triangles around the clustered data locations. An offset value defined how far the mesh should be extended in the inner part (within areas where predictions are required) and the outer part (outside the area where predictions are required). The maximum edge value specified the maximum allowed edge length of the triangle in the inner domain and the outer extension. The inner maximum edge value was small enough to allow the triangulation to support functions with small enough features, and typically smaller than the spatial correlation range of the model (Lindgren, Rue, and Lindström 2011).

As opposed to the regular grid, this approach is denser in regions where there are more observations and consequently generates more information. Another advantage is that this approach saves computing time because prediction locations are typically much lower in number than those in a regular grid.

3.3.4 Pixel-level model estimates

The prediction surfaces generated from the submodels (described in Section 3.3.2) were used as input covariates in the geostatistical models implemented in INLA. The final estimates (and uncertainty) for each indicator were generated by taking $k = 1, \ldots 1000$ samples from the posterior predictive distribution. Pixel level estimates that covered the modeling country were produced at a high spatial resolution of 5 x 5 km.

3.3.5 Model estimates at administrative level 2

In addition to the 5 x 5 km pixel level estimates, we overlaid the prediction prevalence surfaces (from Section 3.3.4) with the total population layer for each indicator we modeled. We then constructed estimates of each indicator at the Admin 2, or district level, by calculating population-weighted averages of prevalence for all grid cells within a given administrative boundary. The procedure was performed for each of the 1,000 posterior predictive samples with final point estimates derived from the mean of these draws and uncertainty intervals from the 2.5 and 97.5 percentiles.

3.3.6 Model validation

For each of the indicator model outputs, a validation procedure was implemented, and a set of performance statistics was calculated. This proceeded with an out-of-sample cross-validation consisting of a five-fold hold-out procedure, with the predicted values at the locations of the hold-out data compared to their observed values. This procedure was repeated five times without replacement, so that every data point was held out once across the five validation runs. Validation statistics were then computed as measures of the predictive accuracy of the fitted models. This includes mean absolute error (MAE: measure of total variation in the errors), the correlation, root mean squared error (RMSE: a measure of the total variance), and 95% coverage of our predictive intervals (the proportion of observed data that fall within our predicted 95% uncertainty intervals). Each predictive metric was calculated by first simulating predictive draws using a binomial distribution. The predictive metric of interest was then calculated as a sample-size-weighted mean over the second administrative levels (Chipeta et al. 2022; Mayala et al. 2019; Mosser et al. 2019). Finally, to complement the out-of-sample predictive validity metrics, we also calculated in-sample predictive validity metrics using the same process but matching each data point to predictions from a model fitted with all data.

4 RESULTS

4.1 Changes in Basic Handwashing Modeled Admin 2 Estimates from Pre-March 2020 to Post-March 2020

4.1.1 Rwanda

For Rwanda's 13 districts that were surveyed both before and after the start of the COVID-19 pandemic, 10 showed an increase in basic handwashing from before the start of the COVID-19 pandemic to after (Figure 2). The average percentage change for these 10 districts was 27.7%. Five of the six districts in the East Province showed an increase in basic handwashing, while all three districts of the Kigali City Province showed an increase, and one of the three districts in the Southern Province showed an increase (Table 3). Since the districts in North or West Provinces were not surveyed before and after the start of the COVID-19 pandemic, there were no data available to compare any potential change.

Although 10 of the 13 districts showed an increase in basic handwashing, the uncertainty intervals for these estimates are fairly wide (Table 3).

Figure 2 Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before and after the pandemic onset (March 2020) using data from the Rwanda 2019–20 DHS: (a) Admin 2 estimates using data from clusters surveyed before the pandemic started; (b) Admin 2 estimates using data from clusters surveyed after the pandemic started; and (c) percentage change from (a) to (b)

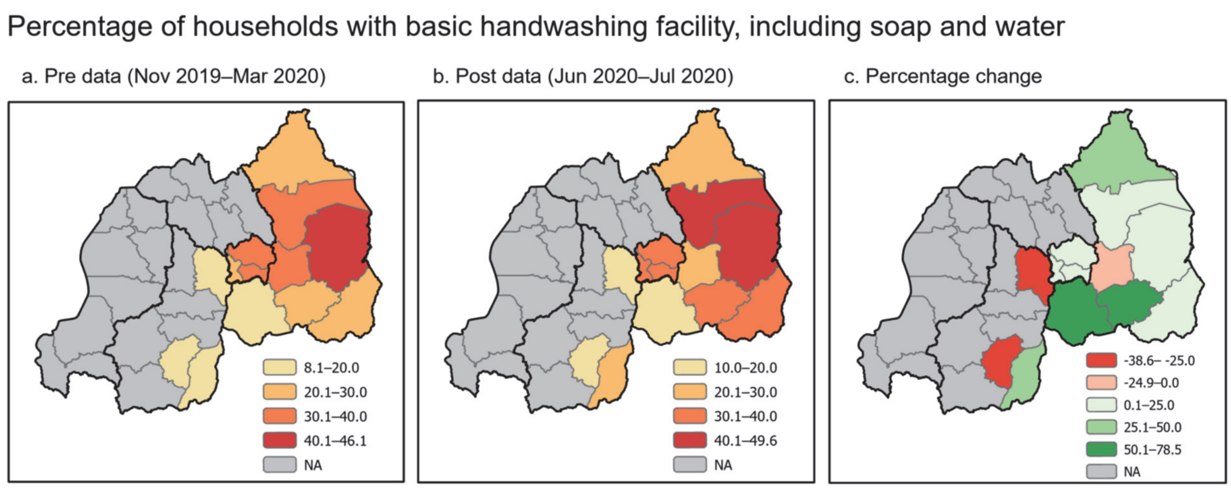


Table 3 **Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before (pre) and after (post) the pandemic onset (March 2020), Rwanda 2019–20 DHS**

Admin 1	Admin 2	Pre %	Pre Lower UI	Pre Upper UI	Post %	Post Lower UI	Post Upper UI	Raw Difference	% Change
East	Bugesera	8.1	4.3	13.8	14.4	8.2	22.8	6.4	78.5
East	Gatsibo	39.3	27.0	52.1	46.4	33.5	58.1	7.1	17.9
East	Kayonza	46.1	31.5	60.9	49.6	37.5	61.6	3.5	7.6
East	Kirehe	29.3	18.6	41.4	32.6	22.4	44.2	3.3	11.4
East	Ngoma	25.0	15.6	36.2	39.1	27.7	51.2	14.1	56.5
East	Nyagatare	20.2	11.2	32.1	27.6	15.8	42.0	7.4	36.7
East	Rwamagana	30.2	18.6	43.7	25.6	16.0	37.5	-4.6	-15.2
Kigali City	Gasabo	31.7	18.1	48.0	36.8	23.8	50.9	5.1	16.0
Kigali City	Kicukiro	32.7	16.7	51.8	35.7	18.0	54.9	3.0	9.2
Kigali City	Nyarugenge	28.6	11	52.2	30.7	12.7	55.5	2.2	7.6
South	Gisagara	17.5	9.2	28.6	23.7	14.1	35.6	6.2	35.6
South	Huye	15.5	7.8	26.5	10.0	4.9	17.3	-5.5	-35.4
South	Kamonyi	18.8	11.0	28.6	11.6	6.0	19.8	-7.3	-38.6

Note: Pre % = Percentage estimate using pre data. Post % = Percentage estimate using post data. UI = Uncertainty interval; Raw Difference = (post % - pre %); % change = Percentage change from pre to post.

4.1.2 India

For India, 47 of the 49 districts showed an increase in basic handwashing from pre to post, with an average percentage increase of 20.5%. The 49 districts included in this analysis only represented 12 of India's states and union territories, the largest administrative divisions (Admin 1) in India. The maps of the India results shown in Figure 3 only include these 12 states to better visualize the results.

The positive percentage changes in the 47 districts varied from 0.1% to 56.7% (Table 4), while the negative percentage changes for the two districts were -2.2% and -4.0% (Table 4).

Figure 3 **Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before and after the pandemic onset (March 2020), India 2019–21 DHS: (a) Admin 2 estimates using data from clusters surveyed before the pandemic started; (b) Admin 2 estimates using data from clusters surveyed after the pandemic started; and (c) percentage change from (a) to (b)**

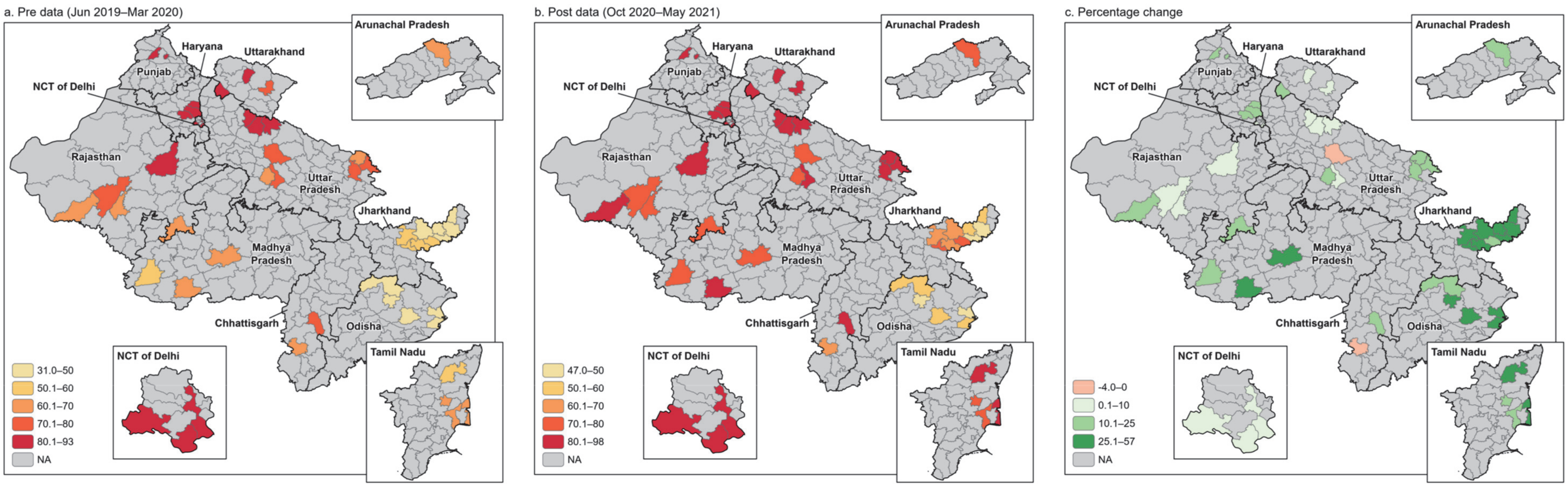


Table 4 Modeled Admin 2 estimates of the percentage of households with basic handwashing facility, including soap and water, in districts surveyed before (pre) and after (post) the pandemic onset (March 2020), India 2019–21 DHS

Admin 1	Admin 2	Pre %	Pre Lower UI	Pre Upper UI	Post %	Post Lower UI	Post Upper UI	Raw Difference	% Change
Arunachal Pradesh	Upper Siang	65.7	58.1	73.0	76.5	68.1	84.0	10.8	16.4
Chhattisgarh	Narayanpur	63.5	54.0	72.1	60.9	51.1	69.3	-2.5	-4.0
Chhattisgarh	Dhamtari	73.5	67.2	79.1	81.4	75.7	85.9	7.9	10.8
Haryana	Panipat	83.6	76.8	88.8	93.8	89.8	96.4	10.2	12.1
Haryana	Rohtak	82.0	74.7	87.6	91.9	88.1	94.8	9.9	12.1
Haryana	Sonipat	83.4	78.6	87.4	93.9	91.2	95.9	10.5	12.5
Jharkhand	Bokaro	52.4	44.4	59.6	68.5	57.2	77.6	16.1	30.7
Jharkhand	Ramgarh	53.0	44.5	61.3	70.3	60.4	79.0	17.3	32.6
Jharkhand	Jamtara	31.7	25.4	38.4	47.5	37.1	58.4	15.8	50.0
Jharkhand	Hazaribagh	50.4	43.8	57.0	67.8	61.0	74.5	17.5	34.7
Jharkhand	Dumka	34.1	28.0	40.3	48.6	38.9	58.3	14.5	42.6
Jharkhand	Godda	41.6	34.5	48.3	58.1	48.3	67.3	16.5	39.5
Jharkhand	Giridih	45.7	39.2	51.6	62.6	53.2	71.9	16.9	37.1
Jharkhand	Dhanbad	58.0	49.4	66.0	70.3	58.9	79.9	12.3	21.2
Jharkhand	Deoghar	40.7	34.3	47.5	54.3	43.1	64.7	13.7	33.6
Madhya Pradesh	Dhar	58.7	51.6	65.8	72.3	65.2	78.9	13.6	23.1
Madhya Pradesh	Khandwa	62.5	53.8	70.7	82.3	76.3	87.3	19.7	31.6
Madhya Pradesh	Raisen	63.5	56.8	70.5	79.7	74.2	84.5	16.2	25.4
NCT of Delhi	South West	91.8	85.3	96.0	94.5	88.7	97.8	2.7	2.9
NCT of Delhi	Central	90.9	79.8	97.2	97.1	91.4	99.3	6.2	6.8
NCT of Delhi	South	89.7	79.9	95.4	95.4	90.0	98.3	5.7	6.3
NCT of Delhi	South East	92.5	83.6	97.6	97.4	92.6	99.3	4.9	5.3
Odisha	Debagarh	31.0	24.9	37.4	48.6	36.3	61.3	17.6	56.7
Odisha	Kendrapara	43.7	37.0	50.5	58.6	46.0	70.3	14.9	34.1
Odisha	Bhadrak	39.4	33.3	45.9	49.5	37.0	62.3	10.1	25.7
Odisha	Dhenkanal	36.7	30.4	42.7	52.4	41.8	63.9	15.8	43.1
Odisha	Sundargarh	47.6	39.4	55.2	59.3	50.8	66.8	11.7	24.5
Punjab	Kapurthala	80.9	74.0	86.2	90.4	86.5	93.3	9.5	11.8
Rajasthan	Jalor	66.4	60.6	71.8	80.7	73.3	86.4	14.3	21.6
Rajasthan	Jhalawar	63.8	57.6	69.3	74.9	68.0	81.2	11.1	17.4
Rajasthan	Jaipur	82.3	78.1	86.0	88.1	82.9	92.1	5.7	7.0
Rajasthan	Pali	78.8	74.8	82.6	78.9	71.6	85.3	0.1	0.1
Rajasthan	Rajsamand	69.9	62.3	76.5	74.4	67.3	80.7	4.4	6.3
Tamil Nadu	Nagapattinam	64.7	58.3	70.4	82.4	75.5	88.4	17.7	27.4
Tamil Nadu	Perambalur	60.1	52.2	67.7	74.2	65.1	81.8	14.1	23.5
Tamil Nadu	Thanjavur	64.2	58.1	70.8	78.8	72.1	84.5	14.5	22.6
Tamil Nadu	Tiruvannamalai	58.8	51.5	65.9	83.4	77.7	87.8	24.6	41.9
Uttar Pradesh	Kanpur Nagar	79.1	73.1	84.1	83.6	75.9	89.7	4.4	5.6
Uttar Pradesh	Gorakhpur	71.2	65.1	77.0	84.2	78.8	88.6	12.9	18.1
Uttar Pradesh	Kushinagar	70.1	63.6	76.4	82.8	77.1	87.7	12.7	18.1
Uttar Pradesh	Pilibhit	82.9	77.7	87.4	87.6	83.6	91.3	4.8	5.8
Uttar Pradesh	Rampur	80.4	73.5	86.3	84.3	78.8	88.4	3.9	4.9
Uttar Pradesh	Hardoi	75.4	69.8	79.9	73.7	64.8	81.3	-1.7	-2.2
Uttar Pradesh	Kanpur Dehat	67.8	60.9	74.1	76.1	69.3	82.2	8.3	12.2
Uttar Pradesh	Maharajganj	68.7	62.2	74.6	81.3	75.0	87.3	12.6	18.3
Uttar Pradesh	Bareilly	85.2	80.3	88.9	88.2	83.5	91.9	3.0	3.6
Uttarakhand	Hardwar	81.2	75.4	86.0	90.6	86.0	94.2	9.4	11.6
Uttarakhand	Bageshwar	76.3	68.9	82.5	80.9	74.5	86.4	4.6	6.0
Uttarakhand	Rudraprayag	80.5	73.0	86.4	85.9	80.6	90.5	5.4	6.6

Note: Pre % = Percentage estimate using pre data. Post % = Percentage estimate using post data. UI = Uncertainty interval; Raw Difference = (post % - pre %); % change = Percentage change from pre to post.

4.2 Evaluation of Systematic Differences between Pre-March 2020 and Post-March 2020 Samples

As described in Section 3.1.3, we evaluated the potential for systematic differences in additional factors that included the urban and rural distribution and a wealth metric, between the pre-March 2020 sample and the post-March 2020 sample.

4.2.1 Rwanda

For the urban and rural distribution comparison in Rwanda, the percentage of interviewed clusters that were urban in the pre-March 2020 sample and the post-March 2020 sample were 27.6% and 31.6%, respectively. No districts exhibited a significant difference in the proportion of urban clusters between the pre and post samples (Table 5).

We also evaluated if the clusters surveyed post-March 2020 were systematically wealthier, as measured by the asset-based wealth index variable, which might help explain any improvement in outcomes. For Rwanda, there was no significant difference in seven of the 13 districts. The wealth index value was significantly higher in the post sample for four of the 13 districts, and the wealth index value was significantly higher in the pre sample for two of the 13 districts (Table 5). When comparing the wealth index value changes to the changes observed in basic handwashing in these districts, there was no clear pattern. Among the three districts that observed a decrease in basic handwashing from pre to post, all three had significantly higher wealth index values in the post-sample. Among the 10 districts that observed an increase in basic handwashing from pre- to post-COVID, two had significantly higher wealth index values in the pre-sample, one had a significantly higher wealth index value in the post-sample, and seven had no significant difference. Therefore, only one district with an increase in basic handwashing from pre to post also showed a significant increase in wealth index value from pre to post.

Table 5 Results of the evaluation of systematic differences in wealth index (WI) and residence (urban/rural) between the pre-March 2020 and post-March 2020 samples for Admin 2s, Rwanda

Admin 1	Admin 2	% Change	Pre WI	Post WI	p-value	WI Change	Pre % Urban	Post % Urban	p-value
East	Bugesera	78.5	2.81	2.95	0.301		0.0	20.0	0.485
East	Gatsibo	17.9	3.06	2.94	0.422		0.0	20.0	0.524
East	Kayonza	7.6	2.48	2.87	0.012*	⇧	0.0	20.0	1.000
East	Kirehe	11.4	2.91	2.59	0.018*	⇩	28.6	0.0	0.154
East	Ngoma	56.5	3.47	2.90	<0.001*	⇩	20.0	0.0	0.333
East	Nyagatare	36.7	2.84	2.92	0.546		14.3	22.2	1.000
East	Rwamagana	-15.2	2.88	3.23	0.012*	⇧	0.0	20.0	0.524
Kigali City	Gasabo	16.0	4.17	4.22	0.737		66.7	76.9	1.000
Kigali City	Kicukiro	9.2	4.59	4.66	0.402		62.5	100.0	0.069
Kigali City	Nyarugenge	7.6	4.30	4.44	0.321		75.0	75.0	1.000
South	Gisagara	35.6	2.32	2.14	0.398		20.0	10.0	1.000
South	Huye	-35.4	2.26	2.66	0.030*	⇧	0.0	16.7	1.000
South	Kamonyi	-38.6	2.54	3.05	<0.001*	⇧	16.7	10.0	1.000

Note: * statistically significant at p<0.05. % change = Percentage change in basic handwashing from pre to post sample (see Table 3). Pre WI = Average wealth index value for the pre-March 2020 sample. Post WI = Average wealth index value for the post-March 2020 sample. WI change = direction of change from pre to post samples. Pre % urban = percentage of urban clusters in the pre sample. Post % urban = percentage of urban clusters in the post sample.

4.2.2 India

The percentage of urban clusters in India surveyed was slightly larger in the post-March 2020 sample. Urban clusters comprised 25.2% of the clusters surveyed in the pre-March 2020 sample and 31.8% in the post-March 2020 sample. Only four of the 49 districts exhibited a significant difference ($p<0.05$) in the proportion of urban clusters between the pre-March 2020 and post-March 2020 samples (Table 5).

For India, there was no significant difference in wealth index when comparing the pre and post samples for 23 of the 49 districts. Of the 26 districts that did show a significant difference, nine districts had significantly lower wealth index values in the post sample, while 17 districts had significantly higher wealth index values in the post sample. For the two districts that showed a decrease in basic handwashing, one had a significantly higher wealth index value in the post sample, while the other had a significantly lower wealth index value in the post sample. For the 47 districts that showed an increase in basic handwashing, eight had a significantly lower wealth index value in the post sample, 16 had a significantly higher wealth index value in the post sample, and 23 showed no significant difference.

Table 6 **Results of the evaluation of systematic differences in wealth index (WI) and residence (urban/rural) between the pre-March 2020 and post-March 2020 samples, India**

Admin 1	Admin 2	% Change	Pre WI	Post WI	p-value	WI Change	Pre % Urban	Post % Urban	p-value
Arunachal Pradesh	Upper Siang	16.4	2.23	2.80	<0.001*	⇧	21.6	0.0	0.316
Chhattisgarh	Narayanpur	-4.0	1.51	1.69	0.006*	⇧	9.1	19.4	0.654
Chhattisgarh	Dhamtari	10.8	2.73	2.85	0.173		16.0	23.5	0.694
Haryana	Panipat	12.1	3.99	4.15	0.059		39.1	52.6	0.535
Haryana	Rohtak	12.1	3.79	4.32	<0.001*	⇧	0.0	56.7	0.001*
Haryana	Sonipat	12.5	4.17	4.22	0.428		23.1	43.8	0.187
Jharkhand	Bokaro	30.7	2.82	2.26	0.003*	⇩	48.8	25.0	0.611
Jharkhand	Ramgarh	32.6	2.45	2.77	0.004*	⇧	42.1	57.1	0.682
Jharkhand	Jamtara	50.0	1.73	1.45	0.046*	⇩	10.8	0.0	1.000
Jharkhand	Hazaribagh	34.7	2.27	2.65	<0.001*	⇧	11.8	27.3	0.337
Jharkhand	Dumka	42.6	1.57	1.74	0.033*	⇧	6.3	8.3	1.000
Jharkhand	Godda	39.5	1.60	1.97	<0.001*	⇧	0.0	11.8	0.137
Jharkhand	Giridih	37.1	2.08	2.16	0.324		7.5	20.0	0.387
Jharkhand	Dhanbad	21.2	2.79	2.62	0.230		56.8	62.5	1.000
Jharkhand	Deoghar	33.6	1.82	1.72	0.785		17.1	20.0	1.000
Madhya Pradesh	Dhar	23.1	2.45	1.82	0.001*	⇩	12.5	0.0	1.000
Madhya Pradesh	Khandwa (East Nimar)	31.6	2.67	2.30	0.003*	⇩	23.1	0.0	0.087
Madhya Pradesh	Raisen	25.4	2.39	2.13	0.070		22.2	5.3	0.234
NCT of Delhi	South West	2.9	4.65	4.78	0.079		88.6	100.0	1.000
NCT of Delhi	Central	6.8	4.32	4.82	<0.001*	⇧	97.1	100.0	1.000
NCT of Delhi	South	6.3	4.56	4.78	<0.001*	⇧	100.0	100.0	1.000
NCT of Delhi	South East	5.3	4.55	4.52	0.985		100.0	100.0	1.000
Odisha	Debagarh	56.7	1.92	1.52	0.007*	⇩	7.9	0.0	1.000
Odisha	Kendrapara	34.1	2.25	2.46	0.118		5.3	0.0	1.000
Odisha	Bhadrak	25.7	2.07	1.99	0.221		8.3	40.0	0.104
Odisha	Dhenkanal	43.1	2.20	2.27	0.963		8.8	14.3	0.542
Odisha	Sundargarh	24.5	2.44	2.48	0.470		39.3	28.6	0.734
Punjab	Kapurthala	11.8	4.35	4.57	<0.001*	⇧	9.5	65.0	0.000*
Rajasthan	Jalor	21.6	2.81	3.09	0.002*	⇧	9.4	7.7	1.000
Rajasthan	Jhalawar	17.4	2.65	2.15	<0.001*	⇩	16.1	14.3	1.000
Rajasthan	Jaipur	7.0	3.66	3.90	0.007*	⇧	51.7	56.3	1.000
Rajasthan	Pali	0.1	3.58	3.77	0.063		20.0	40.0	0.306
Rajasthan	Rajsamand	6.3	3.05	3.08	0.611		16.1	14.3	1.000
Tamil Nadu	Nagapattinam	27.4	2.72	3.96	<0.001*	⇧	20.5	33.3	0.525
Tamil Nadu	Perambalur	23.5	2.97	3.18	0.005*	⇧	20.7	7.7	0.405
Tamil Nadu	Thanjavur	22.6	3.15	3.31	0.187		30.6	66.7	0.164
Tamil Nadu	Tiruvannamalai	41.9	2.88	2.99	0.230		10.5	26.1	0.258
Uttar Pradesh	Kanpur Nagar	5.6	3.52	3.38	0.381		68.3	50.0	0.591
Uttar Pradesh	Gorakhpur	18.1	2.72	2.49	0.003*	⇩	24.1	6.3	0.226
Uttar Pradesh	Kushinagar	18.1	2.52	2.25	<0.001*	⇩	4.0	5.0	1.000
Uttar Pradesh	Pilibhit	5.8	2.20	2.82	<0.001*	⇧	0.0	27.6	0.037*
Uttar Pradesh	Rampur	4.9	2.96	3.35	<0.001*	⇧	0.0	37.9	0.009*
Uttar Pradesh	Hardoi	-2.2	1.87	1.73	0.047*	⇩	13.2	14.3	1.000
Uttar Pradesh	Kanpur Dehat	12.2	2.16	2.21	0.710		3.6	17.6	0.144
Uttar Pradesh	Maharajganj	18.3	2.41	2.39	0.471		3.8	5.3	1.000
Uttar Pradesh	Bareilly	3.6	3.22	3.08	0.127		41.2	32.1	0.749
Uttarakhand	Hardwar	11.6	3.69	3.85	0.883		36.6	33.3	1.000
Uttarakhand	Bageshwar	6.0	2.64	2.84	0.016*	⇧	0.0	10.5	0.173
Uttarakhand	Rudraprayag	6.6	2.93	2.87	0.250		0.0	5.4	1.000

Note: * statistically significant at p<0.05. % change = Percentage change in basic handwashing from pre to post sample (see Table 3).Pre WI = Average wealth index value for the pre-March 2020 sample. Post WI = Average wealth index value for the post-March 2020 sample. WI change = direction of change from pre to post samples. Pre % urban = percentage of urban clusters in the pre sample. Post % urban = percentage of urban clusters in the post sample.

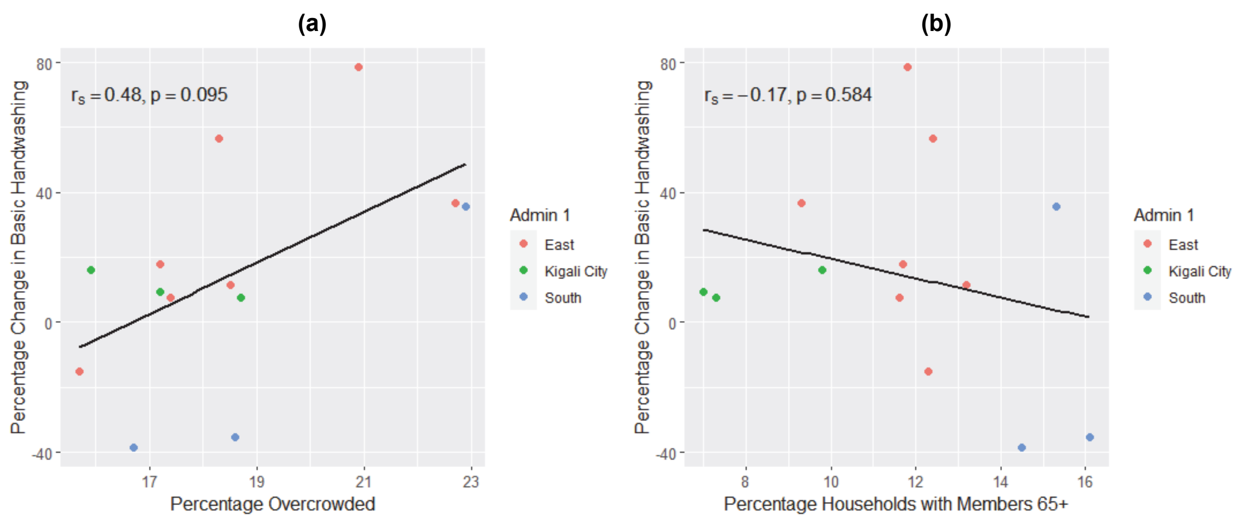
4.3 Evaluation of Correlation between COVID-19 Risk Factors and Change in Basic Handwashing Facilities

The modeled Admin 2 estimates and uncertainty intervals of the two COVID-19 risk factors (percentage of overcrowded households as a proxy for coronavirus transmission and the percentage of households with at least one member age 65 or older as a proxy for COVID-19 mortality) can be found in Appendix Tables 1 and 2.

4.3.1 Rwanda

For Rwanda, the correlation between the percentage change in basic handwashing and the risk of coronavirus transmission, as measured by the percentage overcrowded, was not statistically significant (Spearman's $r_s = 0.48$, $p = 0.095$). The correlation between the percentage change in basic handwashing and the risk of COVID-19 mortality, as measured by the percentage of households with at least one member age 65+, was not statistically significant (Spearman's $r_s = -0.17$, $p = 0.584$).

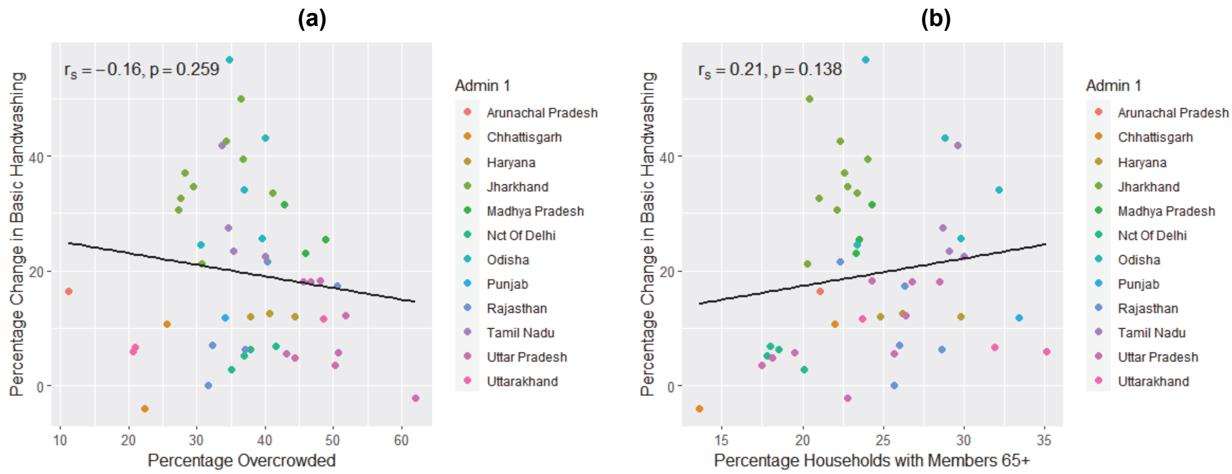
Figure 4 Comparison of the percentage change in basic handwashing from pre-March 2020 to post-March 2020 in Rwandan districts to the (a) percentage of overcrowded households and (b) percentage of households with at least one member age 65+



4.3.2 India

For India, the correlation between the percentage change in basic handwashing and the risk of coronavirus transmission, as measured by the percentage overcrowded, was not statistically significant (Spearman's $r_s = -0.16$, $p = 0.259$). The correlation between the percentage change in basic handwashing and the risk of COVID-19 mortality, as measured by the percentage of households with at least one member age 65 or older, was also not statistically significant (Spearman's $r_s = 0.21$, $p = 0.138$).

Figure 5 **Comparison of the percentage change in basic handwashing from pre-March 2020 to post-March 2020 in Indian districts to the (a) percentage of overcrowded households and (b) percentage of households with at least one member 65+ years old**



4.4 Model Validation Metrics

As described in Section 3.3.6, model validation was performed by calculating mean absolute error (MAE); variance (RMSE); and 95% data coverage within prediction intervals; and the correlation between observed data and predictions. Results indicate that the models were correlated to the input data with values ranging from 0.858 to 0.976 for Rwanda. For India, the correlation between the model and observed data ranged from 0.938 to 0.988. Overall, each indicator performed consistently well for both countries with lower MAE and RMSE, and higher correlation values (Table 7 and Table 8).

Table 7 **Predictive metrics for each indicator aggregated at Admin 2, Rwanda**

Indicator		MAE	RMSE	95% Coverage	Correlation
Basic handwashing (pre)	In-sample	0.0483	0.0603	0.9829	0.9762
	Out-of-sample	0.0624	0.0791	0.9686	0.9522
Basic handwashing (post)	In-sample	0.0619	0.0736	0.9852	0.9278
	Out-of-sample	0.0707	0.0888	0.9409	0.9179
Overcrowding	In-sample	0.0183	0.0250	0.9799	0.8577
	Out-of-sample	0.0191	0.0242	0.9658	0.8903
Age 65+	In-sample	0.0135	0.0168	0.9839	0.9165
	Out-of-sample	0.0158	0.0188	0.9781	0.9161

Table 8 **Predictive metrics for each indicator aggregated at Admin 2, India**

Indicator		MAE	RMSE	95% Coverage	Correlation
Basic handwashing (pre)	In-sample	0.0262	0.0332	0.9833	0.9796
	Out-of-sample	0.0250	0.0374	0.9731	0.9703
Basic handwashing (post)	In-sample	0.0232	0.0305	0.9878	0.9850
	Out-of-sample	0.0228	0.0338	0.9805	0.9795
Overcrowding	In-sample	0.0166	0.0215	0.9733	0.9882
	Out-of-sample	0.0191	0.0266	0.9627	0.9810
Age 65+	In-sample	0.0121	0.0157	0.9738	0.9691
	Out-of-sample	0.0138	0.0214	0.9659	0.9382

We also conducted a comparison of the modeled and direct estimates. For Rwanda, Appendix Figures 1 and 2 show the Admin 2 estimates for each indicator produced by the models in our analysis and the equivalent estimates from the observed DHS survey data. The results indicate a high correlation between MBG and DHS estimates for the indicators modeled. Similarly, the results for India also indicate a high correlation between MBG and DHS estimates for all the indicators we modeled (Appendix Figures 3 and 4).

5 DISCUSSION AND CONCLUSION

In this report, we estimated the percentage of households with a basic handwashing facility, including soap and water, in districts with households surveyed before and after the start of the COVID-19-pandemic, defined here as March 2020. We found that most districts in Rwanda (10/13) and India (47/49) had higher estimates using the post-March 2020 data compared to the pre-March 2020 data. It should be noted however, that most of the uncertainty intervals were wide and mostly overlapping. This was likely to be driven by the smaller sample sizes after dividing the data into pre and post samples. While we cannot conclude that there was a definitive improvement in most of these districts, it is notable that almost all districts observed a higher estimate in the post sample compared to the pre sample.

From our exploration of systematic differences in wealth or urban/rural distribution between the pre and post samples that may be driving any improvements in outcomes, we did not find any consistent differences. For Rwanda, only one district with an increase in basic handwashing from the pre to the post samples had a significantly higher wealth index value. Thus, it is unlikely that differences in wealth drove the general improvements in basic handwashing. It should be noted that the pandemic may have affected household-level asset ownership, which might affect this comparison. The percentage of urban clusters was slightly higher in the post sample than the pre sample, which may contribute to the increase in basic handwashing because urban areas generally have greater access to basic handwashing facilities than rural areas (Brauer et al. 2020). However, none of the districts showed a significantly higher proportion of urban clusters in the post sample compared to the pre sample.

For India, the results were similarly inconsistent. While it is possible that higher levels of wealth contributed to the increase in basic handwashing observed in 16 districts, there are also 31 districts that improved in basic handwashing, but saw no increase or had a significant decrease in wealth index (Table 6). The percentage of urban clusters was also slightly higher in the post sample, while only four of the 49 districts exhibited a significantly higher proportion of urban clusters in the post sample compared to the pre sample. Although the greater proportion of urban clusters surveyed in these districts could explain the increases in basic handwashing, differences in the urban and rural distribution between the pre and post samples do not appear to be a consistent driver in improved basic handwashing access in India.

There was not a statistically significant correlation between Rwandan or Indian districts' risk of coronavirus transmission and mortality, as measured by overcrowding and households with members age 65 or older, and the percentage change in basic handwashing facility access (Figure 4 and Figure 5). While none of these indicators exhibited a statistically significant correlation, it should be noted that these indicators did not quantify perceived risk. Perceived risk is an important determinant in the adoption of protective health behaviors (Aduh et al. 2021). Studies have identified associations between perceived COVID-19 infection risk and the uptake of protective behaviors (de Bruin et al. 2020; Dryhurst et al. 2020). Further studies could determine the relationship between perceived and actual risk in these contexts, and how these factors affected any potential health behavioral changes.

In addition to the factors evaluated here, a multitude of other factors could explain any observed increase between the pre and post sample. The percentage of households with a basic handwashing facility, with soap and water available, increased nationally in Rwanda and India when compared to the previous DHS surveys. Rwanda saw an overall increase from 5.6% to 24.5% when comparing the 2014–15 and 2019–20 DHS surveys. India experienced an increase from 58.7% during the 2015–16 DHS to 69.6% during the 2019–21 DHS. This trend appears to have continued on a smaller scale when comparing district estimates with clusters surveyed prior to March 2020 to district estimates with clusters surveyed after March 2020 (Tables 3 and 4). The factors that have driven this increasing trend over time, prior to the COVID-19 pandemic, could be driving the improvement between pre and post samples. This might be more of a factor for India, where fieldwork was conducted over a longer period, from June 2019 through May 2021, as compared to Rwanda, where fieldwork was conducted between November 2019 and July 2020. Whether this change was accelerated by the COVID-19 pandemic or other factors, the improvement in availability of basic handwashing facilities has wide-reaching benefits.

Beyond its importance to reducing coronavirus transmission, handwashing reduces the spread of a variety of respiratory and diarrheal diseases (Aiello et al. 2008; Rabie and Curtis 2006; Wolf et al. 2022). One recent study estimated that 370,000 deaths from acute respiratory infections in 2016 were attributed to inadequate handwashing, while 165,000 of the total diarrheal deaths in 2016 were attributable to inadequate hygiene behaviors, including handwashing (Prüss-Ustün et al. 2019). While having access to a basic handwashing facility does not measure utilization of such facilities, having a basic handwashing facility in the household is important to ensure that routine handwashing is a convenient and available option. This is reflected in SDG 6.2, which includes access to adequate and equitable sanitation and hygiene for all and the end of open defecation by 2030 (United Nations 2018). One of the key indicators for this target is the percentage of the population with access to a basic handwashing facility with soap and water.

Although this study identified a general increase in handwashing from the pre-March 2020 to post-March 2020 samples for Rwanda and India, more research is needed to evaluate the effect of the COVID-19 pandemic on WASH and broader health behaviors. This study was not designed to evaluate causative factors that underlie any observed change in access to basic handwashing, and was limited by the available data. Further exploration of perceived risk compared to actual risk and an evaluation of the country and district-specific interventions would be helpful. Finally, although this study evaluated changes at the Admin 2 level, additional analyses at different spatial scales would be helpful to better understand the relationship between access to basic handwashing and COVID-19 risk factors.

This study also had modeling limitations that are consistent with those previously described in the geospatial modeling of DHS indicators at the Admin 2 level (Bhatt et al 2017; Mayala et al 2019). Limitations include the use of covariates that may not be ideal proxies for the modeled outcome variable. Furthermore, it was not computationally feasible to propagate uncertainty from the submodels in stacking through the geostatistical model. In addition, splitting the datasets into the pre and post samples may have affected the spatial coverage of the modeled data and the robustness of the geostatistical analyses.

This study, while limited by its design, provides a broad overview of district-level estimates of COVID-related indicators. By evaluating indicators at the policy-relevant Admin 2 level, these data can be used by district-level program planners to better understand the COVID risk of mortality and transmission in their districts, as well as the change in basic handwashing access observed within the timeframe of their district's DHS survey fieldwork.

REFERENCES

Acharya, R., and A. Porwal. 2020. “A Vulnerability Index for the Management of and Response to the COVID-19 Epidemic in India: an Ecological Study.” *The Lancet Global Health* 8 (9): e1142-e1151. https://doi.org/10.1016/S2214-109X(20)30300-4

Aduh, U., M. O. Folayan, A. Afe, A. A. Onyeaghala, I. O. Ajayi, M. Coker, et al. 2021. “Risk Perception, Public Health Interventions, and Covid-19 Pandemic Control in Sub-Saharan Africa.” *Journal of Public Health in Africa.*

Aiello, A. E., R. M. Coulborn, V. Perez, and E. L. Larson. 2008. “Effect of Hand Hygiene on Infectious Disease Risk in the Community Setting: A Meta-Analysis.” *American Journal of Public Health* 98 (8): 1372–1381. https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2007.124610

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. 2nd edition. Boca Raton, FL, USA: CRC Press /Chapman and Hall. https://doi.org/10.1111/biom.12290

Bhatt, S., E. Cameron, S. R. Flaxman, D. J. Weiss, D. L. Smith, and P. W. Gething. 2017. “Improved Prediction Accuracy for Disease Risk Mapping using Gaussian Process Stacked Generalization.” *Journal of the Royal Society Interface* 14 (134): 20170520. https://doi.org/10.1098/rsif.2017.0520

Bhattacharjee, N. V., L. E. Schaeffer, L. B. Marczak, J. M. Ross, S. J. Swartz, J. Albright, W. M. Gardner, et al. 2019. “Mapping Exclusive Breastfeeding in Africa between 2000 and 2017.” *Nature Medicine* 25 (8): 1205–1212. https://doi.org/10.1038/s41591-019-0525-0

Bilal, U., T. Alfaro, and A. Vives. 2021. “COVID-19 and the Worsening of Health Inequities in Santiago, Chile.” *International Journal of Epidemiology* 50 (3): 1038–1040. https://doi.org/10.1093/ije/dyab007

Blake, K. S., R. L. Kellerson, and A. Simic. 2007. *Measuring Overcrowding in Housing*. Washington, DC, USA: US Department of Housing and Urban Development. https://www.huduser.gov/publications/pdf/measuring_overcrowding_in_hsg.pdf

Blangiardo, M., and M. Cameletti. 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. West Sussex, UK: John Wiley & Sons, Ltd. https://www.wiley.com/en-us/Spatial+and+Spatio+temporal+Bayesian+Models+with+R+INLA-p-9781118326558

Brauer, M., J. T. Zhao, F. B. Bennitt, and J. D. Stanaway. 2020. “Global Access to Handwashing: Implications for COVID-19 Control in Low-Income Countries.” *Environmental Health Perspectives* 128 (5): 057005. https://doi.org/10.1289/EHP7200

Breiman, L. 1996. “Stacked Regressions.” *Machine Learning* 24 (1): 49–64. https://doi.org/10.1007/BF00117832

Cameletti, M., F. Lindgren, D. Simpson, and H. Rue. 2012. “Spatio-Temporal Modeling of Particulate Matter Concentration through the SPDE Approach.” AStA Advances in Statistical Analysis 97 (2): 109–131. https://doi.org/10.1007/s10182-012-0196-3

Chipeta, M. G., E. P. Kumaran, A. J. Browne, B. H. Hamadani, G. Haines-Woodhouse, B. Sartorius, R. C. Reiner, et al. 2022. “Mapping Local Variation in Household Overcrowding across Africa from 2000 to 2018: A Modelling Study.” *The Lancet Planetary Health* 6 (8): e670-e681. https://doi.org/10.1016/S2542-5196(22)00149-8

de Bruin, W. B. and D. Bennett. 2020. “Relationships between Initial COVID-19 Risk Perceptions and Protective Health Behaviors: A National Survey.” *American Journal of Preventive Medicine* 59 (2): 157–167. https://doi.org/10.1016/j.amepre.2020.05.001

Diggle, P. J., and E. Giorgi. 2019. *Model-Based Geostatistics for Global Public Health: Methods and Applications*. New York, NY, USA: CRC Press/Chapman and Hall. https://doi.org/10.1201/9781315188492

Dryhurst, S., C. R. Schneider, J. Kerr, A. L. Freeman, G. Recchia, A. M. Van Der Bles, et al. 2020. “Risk Perceptions of COVID-19 Around the World.” *Journal of Risk Research* 23 (7–8): 994–1006. https://doi.org/10.1080/13669877.2020.1758193

Dwyer-Lindgren, L., M. A. Cork, A. Sligar, K. M. Steuben, K. F. Wilson, N. R. Provost, B. K. Mayala, et al. 2019. “Mapping HIV Prevalence in Sub-Saharan Africa between 2000 and 2017.” *Nature* 570 (7760): 189–193. https://doi.org/10.1038/s41586-019-1200-9

Friedman, J. H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* 29 (5): 1189–1232. https://doi.org/10.1214/aos/1013203451

Giné-Garriga R., A. Delepiere, R. Ward, J. Alvarez-Sala, I. Alvarez-Murillo, V. Mariezcurrena, H. Göransson Sandberg, P. Saikia, P. Avello, K. Thakar, E. Ibrahim, A. Nouvellon, O. El Hattab, G. Hutton, and A. Jiménez. 2021. “COVID-19 Water, Sanitation, and Hygiene Response: Review of Measures and Initiatives Adopted by Governments, Regulators, Utilities, and Other Stakeholders in 84 Countries.” *Science of the Total Environment* 795: 148789. https://doi.org/10.1016/j.scitotenv.2021.148789

Ghosh, A. K., S. Venkatraman, O. Soroka, E. Reshetnyak, M. Rajan, A. An, et al. 2021. “Association between Overcrowded Households, Multigenerational Households, and COVID-19: A Cohort Study.” *Public Health* 198: 273–279. https://doi.org/10.1016/j.puhe.2021.07.039

Graetz, N., J. Friedman, A. Osgood-Zimmerman, R. Burstein, M. H. Biehl, C. Shields, J. F. Mosser, et al. 2018. “Mapping Local Variation in Educational Attainment across Africa.” *Nature* 555 (7694): 48–53. https://doi.org/10.1038/nature25761

Grasselli, G., M. Greco, A. Zanella, G. Albano, M. Antonelli, G. Bellani, et al. 2020. “Risk Factors Associated with Mortality among Patients with COVID-19 in Intensive Care Units in Lombardy, Italy.” *JAMA Internal Medicine* 180 (10): 1345–1355. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2768601

Janocha, B., R. E. Donohue, T. D. Fish, B. K. Mayala, and T. N. Croft. 2021. *Guidance and Recommendations for the Use of Indicator Estimates at Subnational Administrative Level 2.* DHS Spatial Analysis Reports No. 20. Rockville, Maryland, USA: ICF. https://dhsprogram.com/pubs/pdf/SAR20/SAR20.pdf

Knutson, V., S. Aleshin-Guendel, A. Karlinsky, W. Msemburi, and J. Wakefield. 2022. "Estimating Global and Country-Specific Excess Mortality During the COVID-19 Pandemic." arXiv preprint arXiv: 2205.09081. https://doi.org/10.48550/arXiv.2205.09081

Koomson, I., M. Okumu, and D. Ansong. 2022. "Introducing the Disease Outbreak Resilience Index (DORI) using the Demographic and Health Surveys Data from sub-Saharan Africa." *Social Indicators Research* 162: 1149–1175. https://doi.org/10.1007/s11205-022-02881-1

Lindgren, F., H. Rue, and J. Lindström. 2011. "An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (4): 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lippi, G., C. Mattiuzzi, F. Sanchis-Gomar, and B. M. Henry. 2020. "Clinical and demographic characteristics of patients dying from COVID-19 in Italy vs China." *Journal of Medical Virology* 92 (10): 1759. https://doi.org/10.1002/jmv.25860

Mayala, B. K., S. Bhatt, and P. Gething. 2020. *Predicting HIV/AIDS at Subnational Levels using DHS Covariates Related to HIV*. DHS Spatial Analysis Reports No. 18. Rockville, Maryland, USA: ICF. https://dhsprogram.com/pubs/pdf/SAR18/SAR18.pdf

Mayala, B. K., T. Dontamsetti, T. D. Fish, and T. N. Croft. 2019. *Interpolation of DHS Survey Data at Subnational Administrative Level 2*. DHS Spatial Analysis Reports No. 17. Rockville, Maryland, USA: ICF. https://dhsprogram.com/pubs/pdf/SAR17/SAR17.pdf

Osgood-Zimmerman, A., A. I. Millear, R. W. Stubbs, C. Shields, B. V. Pickering, L. Earl, N. Graetz, et al. 2018. "Mapping Child Growth Failure in Africa between 2000 and 2015." *Nature* 555 (7694): 41–47. https://doi.org/10.1038/nature25760

Paludan-Müller A.S., K. Boesen, I. Klerings, K.J. Jørgensen, and K. Munkholm. 2020. "Hand Cleaning with Ash for Reducing the Spread of Viral and Bacterial Infections: A Rapid Review*." Cochrane Database of Systematic Reviews* 4 (4): CD013597. https://doi.org/10.1002/14651858.CD013597

Prüss-Ustün, A., J. Wolf, J. Bartram, T. Clasen, O. Cumming, M. C. Freeman, B. Gordon, et al. 2019. "Burden of Disease from Inadequate Water, Sanitation and Hygiene for Selected Adverse Health Outcomes: An Updated Analysis with a Focus on Low- and Middle-Income Countries." *International Journal of Hygiene and Environmental Health* 222 (5): 765–777. https://doi.org/10.1016/j.ijheh.2019.05.004

R Core Team 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rabie, T., and V. Curtis. 2006. "Handwashing and Risk of Respiratory Infections: a Quantitative Systematic Review." *Tropical Medicine & International Health* 11 (3): 258–267. https://doi.org/10.1111/j.1365-3156.2006.01568.x

Reiner, R. C., Jr., N. Graetz, D. C. Casey, C. Troeger, G. M. Garcia, J. F. Mosser, A. Deshpande, et al. 2018. “Variation in Childhood Diarrheal Morbidity and Mortality in Africa, 2000–2015.” *The New England Journal of Medicine* 379 (12): 1128–1138. https://www.nejm.org/doi/10.1056/NEJMoa1716766

Roche, R., R. Bain, and O. Cumming. 2017. “A Long Way to Go–Estimates of Combined Water, Sanitation and Hygiene Coverage for 25 sub-Saharan African Countries.” *PLoS One* 12 (3): e0171783. https://doi.org/10.1371/journal.pone.0173702

Rue, H., S. Martino, and N. Chopin. 2009. “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 (2): 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Smith, L., L. Butler, M. A. Tully, L. Jacob, Y. Barnett, G. F. López-Sánchez, et al. 2021. “Hand-washing Practices among Adolescents Aged 12–15 Years from 80 Countries.” *International Journal of Environmental Research and Public Health* 18 (1): 1-15. https://doi.org/10.3390/ijerph18010138

United Nations. 2018. *The 2030 Agenda and the Sustainable Development Goals: An Opportunity for Latin America and the Caribbean* (LC/G.2681-P/Rev.3). United Nations: Santiago, Chile. 2018. https://www.cepal.org/sites/default/files/events/files/2030_agenda_and_the_sdgs_an_opportunity_for_latin_america_and_the_caribbean.pdf

Wang, H., K. R. Paulson, S. A. Pease, S. Watson, H. Comfort, P. Zheng, P., A. Y. Aravkin, et al. 2022. “Estimating Excess Mortality Due to the COVID-19 Pandemic: a Systematic Analysis of COVID-19-Related Mortality, 2020–21.” *The Lancet* 399 (10334): 1513–1536.

Wolf, J., S. Hubbard, M. Brauer, A. Ambelu, B. F. Arnold, R. Bain, et al. (2022). “Effectiveness of Interventions to Improve Drinking Water, Sanitation, and Handwashing with Soap on Risk of Diarrheal Disease in Children in Low-Income and Middle-Income Settings: a Systematic Review and Meta-Analysis.” *The Lancet* 400 (10345): 48–59. https://doi.org/10.1016/S0140-6736(22)00937-0

Wolpert, D. H. 1992. “Stacked Generalization.” *Neural Networks* 5 (2): 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

World Health Organization (WHO) and the United Nations Children’s Fund (UNICEF) Joint Programme (JMP) for Water Supply, Sanitation, and Hygiene. 2018. *JMP Methodology: 2017 Update & SDG Baselines*. https://washdata.org/sites/default/files/documents/reports/2018-04/JMP-2017-update-methodology.pdf

World Health Organization (WHO) and United Nations Children’s Fund (UNICEF). 2020. “Water, Sanitation, Hygiene, and Waste Management for SARS-Cov-2, The Virus that Causes COVID-19: Interim Guidance, 29 July 2020.” Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/333560

WHO. 2018. *Housing and Health Guidelines.* Geneva: World Health Organization. https://www.ncbi.nlm.nih.gov/books/NBK535289/

Zou, H. and T. Hastie. 2005. “Regularization and Variable Selection via the Elastic Net.” *Journal of the Royal Statistical Society Statistical Methodology Series B* 67 (2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

APPENDIX

Appendix Table 1 **Modeled Admin 2 estimates of the percentage of overcrowded households data, Rwanda 2019–20 DHS**

Admin 1	Admin 2	Estimate (%)	Lower UI (%)	Upper UI (%)	UI Range (%)	Basic Hand-washing % Change
East	Bugesera	20.9	18.0	23.8	5.7	78.5
East	Gatsibo	17.2	14.9	19.6	4.7	17.9
East	Kayonza	17.4	14.8	20.2	5.4	7.6
East	Kirehe	18.5	16.1	21.1	5.0	11.4
East	Ngoma	18.3	15.6	21.2	5.6	56.5
East	Nyagatare	22.7	20.0	25.6	5.6	36.7
East	Rwamagana	15.7	13.2	18.6	5.4	-15.2
Kigali City	Gasabo	15.9	12.5	20.1	7.5	16.0
Kigali City	Kicukiro	17.2	12.2	23.5	11.3	9.2
Kigali City	Nyarugenge	18.7	12.6	27.0	14.5	7.6
North	Burera	15.2	12.7	17.8	5.1	NA
North	Gakenke	13.6	11.3	16.1	4.8	NA
North	Gicumbi	13.7	11.4	16.3	4.9	NA
North	Musanze	15.5	12.6	18.6	6.0	NA
North	Rulindo	13.4	11.1	15.8	4.8	NA
South	Gisagara	22.9	19.3	26.9	7.5	35.6
South	Huye	18.6	15.6	21.9	6.4	-35.4
South	Kamonyi	16.7	14.1	19.3	5.2	-38.6
South	Muhanga	14.4	11.9	17.1	5.2	NA
South	Nyamagabe	20.9	17.9	24.2	6.2	NA
South	Nyanza	21.0	18.0	24.5	6.5	NA
South	Nyaruguru	24.3	21.1	27.9	6.9	NA
South	Ruhango	19.8	16.6	22.8	6.2	NA
West	Karongi	19.8	17.0	22.6	5.7	NA
West	Ngororero	16.1	13.6	18.8	5.2	NA
West	Nyabihu	19.0	15.8	22.8	7.0	NA
West	Nyamasheke	18.7	15.9	21.5	5.6	NA
West	Rubavu	22.4	17.9	27.5	9.6	NA
West	Rusizi	16.4	13.5	19.8	6.3	NA
West	Rutsiro	23.2	19.7	26.8	7.2	NA

Note: Estimate % = Percentage estimate. UI = 95% Uncertainty interval; UI Range = Upper UI – Lower UI; % change = Percentage change from pre to post. NA = Not applicable.

Appendix Table 2 **Modeled Admin 2 estimates of the percentage of overcrowded households in districts surveyed both before and during the COVID-19 pandemic, India 2019–21 DHS**

Admin 1	Admin 2	Estimate (%)	Lower UI (%)	Upper UI (%)	UI Range (%)	Basic Hand-washing % Change
Arunachal Pradesh	Upper Siang	11.2	9.3	13.2	3.9	16.4
Chhattisgarh	Dhamtari	25.6	22.4	29.0	6.5	10.8
Chhattisgarh	Narayanpur	22.3	18.6	26.0	7.4	-4.0
Haryana	Panipat	44.3	38.7	50.1	11.4	12.1
Haryana	Rohtak	37.8	33.3	42.8	9.5	12.1
Haryana	Sonipat	40.7	36.8	44.9	8.1	12.5
Jharkhand	Bokaro	27.3	23.7	31.4	7.7	30.7
Jharkhand	Deoghar	41.1	36.9	45.6	8.6	33.6
Jharkhand	Dhanbad	30.7	26.3	35.5	9.2	21.2
Jharkhand	Dumka	34.3	30.6	38.2	7.6	42.6
Jharkhand	Giridih	28.3	25.4	31.6	6.3	37.1
Jharkhand	Godda	36.7	32.5	40.7	8.2	39.5
Jharkhand	Hazaribagh	29.5	26.4	32.9	6.4	34.7
Jharkhand	Jamtara	36.4	32.5	40.5	8.0	50.0
Jharkhand	Ramgarh	27.7	23.9	32.0	8.1	32.6
Madhya Pradesh	Dhar	46.0	42.0	49.8	7.9	23.1
Madhya Pradesh	Khandwa (East Nimar)	42.8	37.9	47.9	10.1	31.6
Madhya Pradesh	Raisen	48.9	45.0	53.0	8.0	25.4
NCT of Delhi	Central	41.6	30.3	53.9	23.6	6.8
NCT of Delhi	South	37.9	29.5	46.2	16.7	6.3
NCT of Delhi	South East	37.0	27.5	48.1	20.6	5.3
NCT of Delhi	South West	35.1	27.1	43.5	16.4	2.9
Odisha	Bhadrak	39.5	35.4	43.7	8.3	25.7
Odisha	Debagarh	34.8	30.8	38.8	8.0	56.7
Odisha	Dhenkanal	40.1	36.0	44.1	8.1	43.1
Odisha	Kendrapara	36.9	32.5	41.2	8.7	34.1
Odisha	Sundargarh	30.6	27.0	35.1	8.1	24.5
Punjab	Kapurthala	34.2	30.6	37.8	7.2	11.8
Rajasthan	Jaipur	32.3	28.4	36.3	7.9	7.0
Rajasthan	Jalor	40.4	36.8	44.1	7.3	21.6
Rajasthan	Jhalawar	50.5	47.0	53.9	7.0	17.4
Rajasthan	Pali	31.7	28.2	35.2	7.0	0.1
Rajasthan	Rajsamand	37.1	33.2	41.0	7.8	6.3
Tamil Nadu	Nagapattinam	34.6	31.0	38.4	7.3	27.4
Tamil Nadu	Perambalur	35.4	31.1	40.0	8.9	23.5
Tamil Nadu	Thanjavur	40.0	36.3	43.6	7.4	22.6
Tamil Nadu	Tiruvannamalai	33.6	30.1	37.3	7.2	41.9
Uttar Pradesh	Bareilly	50.2	45.8	54.5	8.7	3.6
Uttar Pradesh	Gorakhpur	46.7	43.0	50.6	4.9	18.1
Uttar Pradesh	Hardoi	62.0	58.4	65.5	4.1	-2.2
Uttar Pradesh	Kanpur Dehat	51.8	47.6	55.7	4.7	12.2
Uttar Pradesh	Kanpur Nagar	43.1	37.9	48.6	5.9	5.6
Uttar Pradesh	Kushinagar	45.6	41.7	49.7	4.7	18.1
Uttar Pradesh	Maharajganj	48.1	43.9	52.2	4.9	18.3
Uttar Pradesh	Pilibhit	50.7	46.7	54.5	4.1	5.8
Uttar Pradesh	Rampur	44.4	40.4	49.1	4.1	4.9
Uttarakhand	Bageshwar	20.7	17.5	24.1	6.0	6.0
Uttarakhand	Hardwar	48.6	44.4	53.5	5.0	11.6
Uttarakhand	Rudraprayag	21.0	17.8	24.2	6.4	6.6

Note: Estimate % = Percentage estimate. UI = 95% Uncertainty interval; UI Range = Upper UI – Lower UI; % change = Percentage change from pre to post.

Appendix Table 3 **Modeled Admin 2 estimates of the percentage of households with at least one member age 65+, Rwanda 2019–20 DHS**

Admin 1	Admin 2	Estimate (%)	Lower UI (%)	Upper UI (%)	UI Range (%)	Basic Hand-washing % Change
East	Bugesera	11.8	10.7	12.9	2.2	78.5
East	Gatsibo	11.7	10.5	12.9	2.4	17.9
East	Kayonza	11.6	10.4	12.9	2.5	7.6
East	Kirehe	13.2	11.9	14.5	2.6	11.4
East	Ngoma	12.4	11.2	13.7	2.6	56.5
East	Nyagatare	9.3	8.3	10.4	2.1	36.7
East	Rwamagana	12.3	11.0	13.8	2.8	-15.2
Kigali City	Gasabo	9.8	8.2	11.4	3.2	16.0
Kigali City	Kicukiro	7.0	5.3	9.2	3.9	9.2
Kigali City	Nyarugenge	7.3	5.3	9.8	4.5	7.6
North	Burera	14.1	12.6	16.0	3.4	NA
North	Gakenke	15.9	14.3	17.6	3.4	NA
North	Gicumbi	14.7	13.2	16.6	3.4	NA
North	Musanze	12.2	10.5	14.0	3.6	NA
North	Rulindo	15.7	14.0	17.5	3.6	NA
South	Gisagara	15.3	13.7	17.2	3.6	35.6
South	Huye	16.1	14.3	18.2	3.9	-35.4
South	Kamonyi	14.5	12.9	16.1	3.2	-38.6
South	Muhanga	16.0	14.1	18.0	3.8	NA
South	Nyamagabe	16.6	14.8	18.5	3.7	NA
South	Nyanza	14.9	13.2	16.7	3.5	NA
South	Nyaruguru	17.1	15.3	19.0	3.7	NA
South	Ruhango	14.8	13.2	16.8	3.5	NA
West	Karongi	15.6	14.0	17.3	3.3	NA
West	Ngororero	14.6	13.0	16.3	3.3	NA
West	Nyabihu	13.1	11.4	15.0	3.6	NA
West	Nyamasheke	17.7	16.0	19.5	3.6	NA
West	Rubavu	12.6	10.6	14.9	4.2	NA
West	Rusizi	18.4	16.2	20.7	4.6	NA
West	Rutsiro	13.1	11.7	14.8	3.2	NA

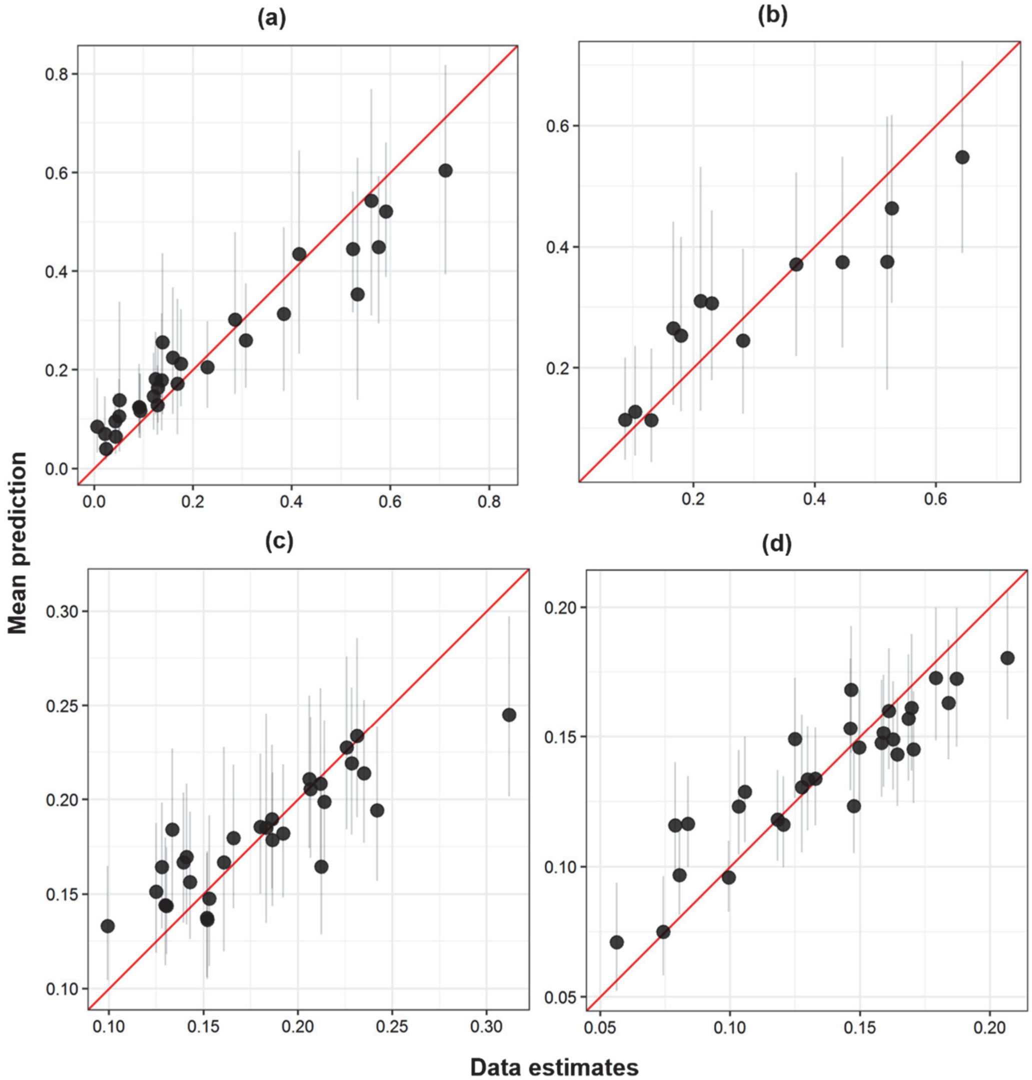
Note: Estimate % = Percentage estimate. UI = 95% Uncertainty interval; UI Range = Upper UI – Lower UI; % change = Percentage change from pre to post. NA = Not applicable.

Appendix Table 4 **Modeled Admin 2 estimates of the percentage of households with at least one member age 65 or older in districts surveyed both before and during the COVID-19 pandemic, India 2019–21 DHS**

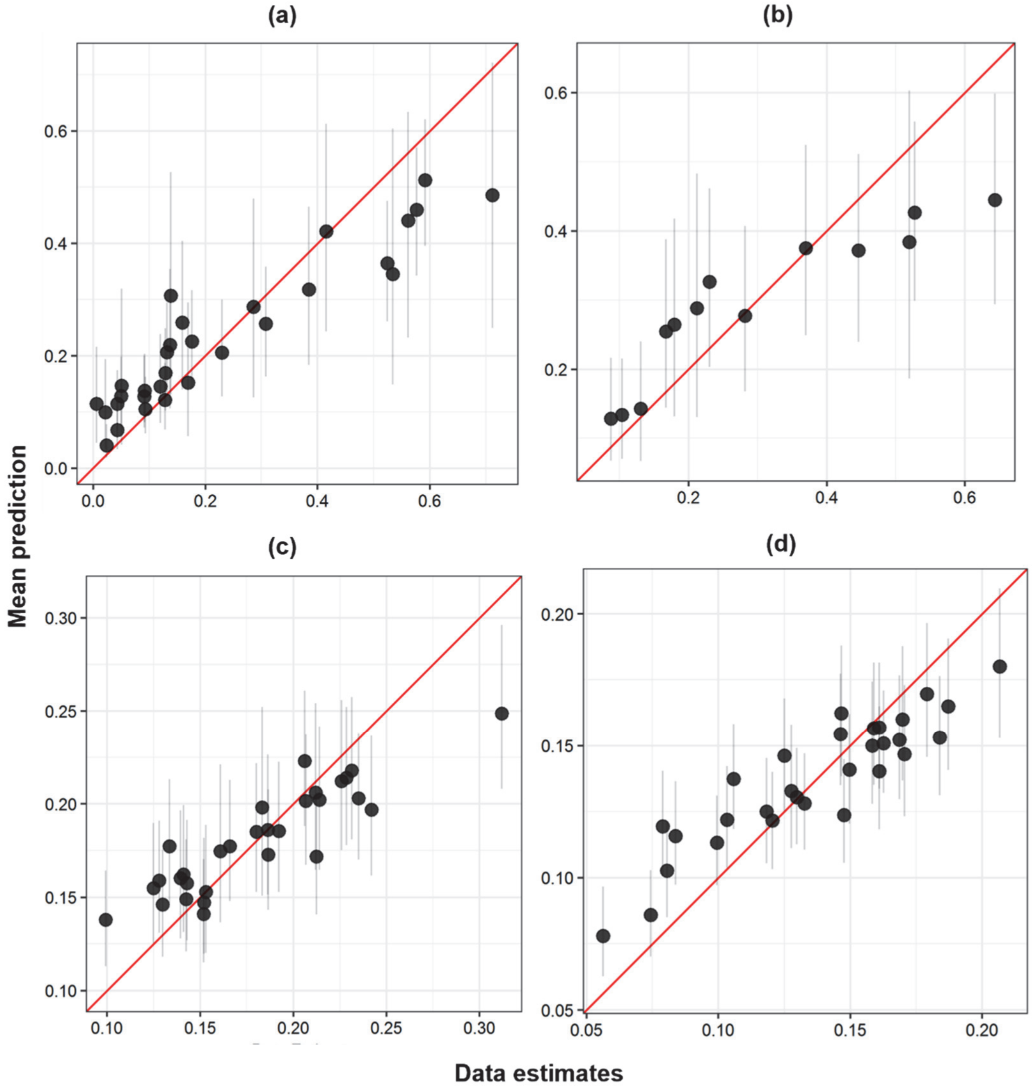
Admin 1	Admin 2	Estimate (%)	Lower UI (%)	Upper UI (%)	UI Range (%)	Basic Hand-washing % Change
Arunachal Pradesh	Upper Siang	21.1	18.8	23.8	5.0	16.4
Chhattisgarh	Dhamtari	22.0	19.9	24.3	4.5	10.8
Chhattisgarh	Narayanpur	13.6	11.6	15.9	4.3	-4.0
Haryana	Panipat	24.8	22.0	28.1	6.1	12.1
Haryana	Rohtak	29.8	27.1	32.8	5.8	12.1
Haryana	Sonipat	26.2	23.9	28.5	4.6	12.5
Jharkhand	Bokaro	22.1	19.6	24.6	5.1	30.7
Jharkhand	Deoghar	23.4	21.0	25.8	4.8	33.6
Jharkhand	Dhanbad	20.3	17.7	23.1	5.4	21.2
Jharkhand	Dumka	22.3	20.3	24.4	4.1	42.6
Jharkhand	Giridih	22.6	20.5	24.8	4.3	37.1
Jharkhand	Godda	24.0	21.6	26.4	4.8	39.5
Jharkhand	Hazaribagh	22.8	20.8	24.9	4.1	34.7
Jharkhand	Jamtara	20.4	18.4	22.6	4.2	50.0
Jharkhand	Ramgarh	21.0	18.7	23.5	4.8	32.6
Madhya Pradesh	Dhar	23.3	21.0	25.6	4.6	23.1
Madhya Pradesh	Khandwa (East Nimar)	24.3	21.5	27.2	5.7	31.6
Madhya Pradesh	Raisen	23.5	21.3	25.9	4.6	25.4
NCT of Delhi	Central	18.0	13.2	23.7	10.5	6.8
NCT of Delhi	South	18.5	15.0	22.6	7.6	6.3
NCT of Delhi	South East	17.8	13.9	22.7	8.7	5.3
NCT of Delhi	South West	20.1	16.6	24.5	7.9	2.9
Odisha	Bhadrak	29.8	27.0	32.6	5.7	25.7
Odisha	Debagarh	23.9	21.4	26.6	5.2	56.7
Odisha	Dhenkanal	28.8	26.4	31.3	4.9	43.1
Odisha	Kendrapara	32.2	29.3	35.0	5.7	34.1
Odisha	Sundargarh	23.4	21.0	25.9	5.0	24.5
Punjab	Kapurthala	33.4	30.6	36.1	5.5	11.8
Rajasthan	Jaipur	26.0	23.7	28.6	4.9	7.0
Rajasthan	Jalor	22.3	20.2	24.6	4.4	21.6
Rajasthan	Jhalawar	26.3	24.2	28.6	4.4	17.4
Rajasthan	Pali	25.7	23.6	28.0	4.4	0.1
Rajasthan	Rajsamand	28.6	26.1	31.3	5.2	6.3
Tamil Nadu	Nagapattinam	28.7	26.2	31.1	5.0	27.4
Tamil Nadu	Perambalur	29.1	26.3	32.0	5.7	23.5
Tamil Nadu	Thanjavur	30.0	27.7	32.3	4.6	22.6
Tamil Nadu	Tiruvannamalai	29.6	26.9	32.3	5.4	41.9
Uttar Pradesh	Bareilly	17.5	15.6	19.7	4.1	3.6
Uttar Pradesh	Gorakhpur	28.5	26.2	31.1	4.9	18.1
Uttar Pradesh	Hardoi	22.8	20.8	24.9	4.1	-2.2
Uttar Pradesh	Kanpur Dehat	26.4	24.1	28.8	4.7	12.2
Uttar Pradesh	Kanpur Nagar	25.7	23.0	28.8	5.9	5.6
Uttar Pradesh	Kushinagar	26.8	24.6	29.2	4.7	18.1
Uttar Pradesh	Maharajganj	24.3	21.9	26.8	4.9	18.3
Uttar Pradesh	Pilibhit	19.5	17.5	21.6	4.1	5.8
Uttar Pradesh	Rampur	18.1	16.3	20.4	4.1	4.9
Uttarakhand	Bageshwar	35.1	32.1	38.2	6.0	6.0
Uttarakhand	Hardwar	23.7	21.2	26.2	5.0	11.6
Uttarakhand	Rudraprayag	31.9	28.8	35.2	6.4	6.6

Note: Estimate % = Percentage estimate. UI = 95% Uncertainty interval; UI Range = Upper UI – Lower UI; % change = Percentage change from pre to post.

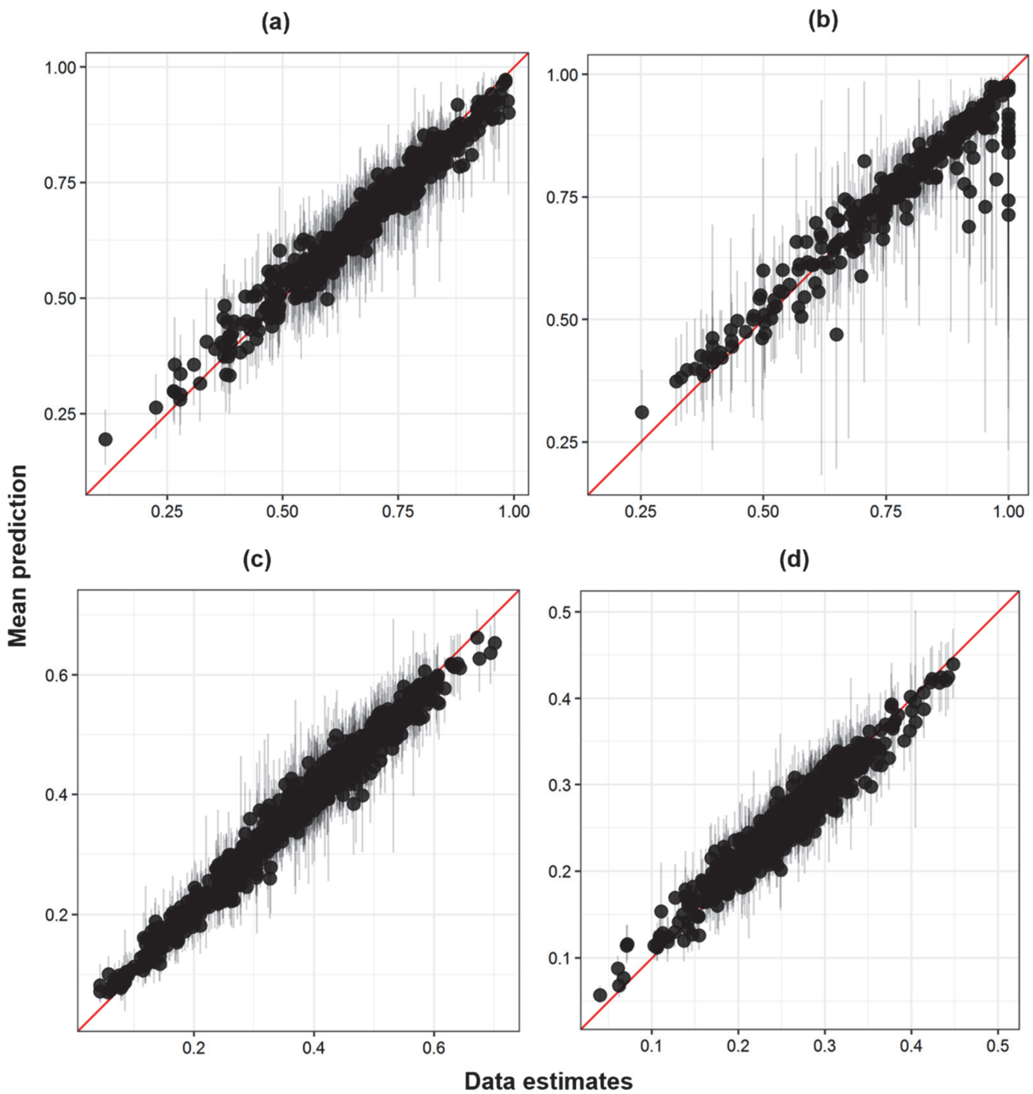
Appendix Figure 1 **Comparison of in-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Age 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, Rwanda**



Appendix Figure 2 **Comparison of out-of-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Age 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, Rwanda**



Appendix Figure 3 **Comparison of in-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Age 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, India**



Appendix Figure 4 **Comparison of out-of-sample predictions for (a) Basic handwashing (pre), (b) Basic handwashing (post), (c) Overcrowding, and (d) Age 65+, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level, India**