

Article

Geospatial and Social Factors Influencing Morbidity due to Acute Infection in Pediatric Patients in Matiari, Rural Pakistan.

Elise Corden¹, Saman Hasan Siddiqui², Yash Sharma¹, Muhammad Faraz Raghieb¹, William Adorno III³, Fatima Zulqarnain¹, Lubaina Ehsan¹, Aman Shrivastava¹, Sheraz Ahmed², Fayaz Umrani², Najeeb Rahman², Rafey Ali², Najeeha T. Iqbal², Sean R. Moore¹, Syed Asad Ali^{2,*}, Sana Syed^{1,2,4,5,*}

¹ Department of Pediatrics, School of Medicine, University of Virginia, Charlottesville, 22903 VA, USA,

² Department of Paediatrics and Child Health, Aga Khan University, Karachi, 74800, Pakistan

³ Department of Engineering Systems and Environment, University of Virginia, Charlottesville, 22903, VA, USA

⁴ Department of Public Health Sciences, University of Virginia, Charlottesville, 22903, VA, USA,

⁵ School of Data Science, University of Virginia, Charlottesville, VA, 22903, USA

* Correspondence:

Sana Syed, MD MSCR MSDS

Associate Professor of Pediatrics

Dept. of Pediatrics, Division of Gastroenterology, Hepatology & Nutrition

University of Virginia School of Medicine & UVA Child Health Research Center

MR-4 Bldg, 409 Lane Rd., Charlottesville, VA 22908

Phone: +1 434-297-5064 Email: sana.syed@virginia.edu

Syed Asad Ali

Professor, Department of Paediatrics and Child Health

Associate Dean, Research, Medical College

Aga Khan University

National Stadium Road, Karachi, 74800 Pakistan

Email: asad.ali@aku.edu Phone: +92 21 34930051

Abstract: Infectious disease is the leading cause of mortality in children under five. This study has investigated environmental factors related to the morbidity of acute respiratory infections (ARIs), diarrhea, and growth using geographical information systems (GIS) technology. Anthropometric, address and disease prevalence data were collected through the SEEM study in Matiari, Pakistan. Publicly available map data was used to compile coordinates of healthcare facilities. A Pearson correlation coefficient (r) was used to calculate the correlation between distance from healthcare facilities and participant growth and morbidity. Other continuous variables influencing these outcomes were analyzed using a random forest regression model. In this study of 416 children, we found participants living closer to secondary hospitals had lower prevalence of ARI ($r=0.154$, $p<0.010$) and diarrhea ($r=0.228$, $p<0.001$) as well as participants living closer to Maternal Health Centers (MHCs): ARI ($r=0.185$, $p<0.002$) and diarrhea ($r=0.223$, $p<0.001$) compared to those living near primary facilities. Our random forest model showed distance to have high variable importance in the context of disease prevalence. Our results indicated that participants closer to more basic healthcare facilities reported a higher prevalence of both diarrhea and ARI than those near more urban facilities, highlighting potential public policy gaps in ameliorating rural health.

Keywords: nutrition, pediatrics, geographic information systems, acute respiratory infections, diarrhea, growth

1. Introduction

Globally, the World Health Organization (WHO) reports that infectious diseases, such as pneumonia and malaria, are the leading cause of mortality in children under five. Poor

nutrition has a significant impact on global childhood mortality, with the WHO reporting malnutrition as being implicated in 45% of deaths in children under the age of 5 [1]. Key health indicators in some low-middle income countries (LMICs), such as Pakistan, show high rates of communicable diseases and undernutrition, providing evidence that Pakistan has yet to meet WHO sustainable developmental goals [1, 2]. Undernutrition is also known to contribute to cognitive dysfunction, loss of productivity, and increased morbidity and mortality from infections [3, 4]. Pakistan spends 3.4% of its total gross domestic product (GDP) on healthcare, and resources are often described to be insufficiently funded or poorly allocated [5]. Healthcare in Pakistan is delivered through a three-tiered system. The first tier, primary care centers, include Basic Health Units (BHUs) and Rural Health Centers (RHCs). These facilities focus on preventive, curative, and referral services. RHCs also maintain a small number of inpatient beds. The second tier, secondary care centers, provides both ambulatory and inpatient care. Examples include Tehsil Headquarter Hospitals and District Headquarter Hospitals. The third tier, teaching hospitals, are tertiary care centers that provide specialized inpatient care and receive referrals from other healthcare centers [5].

One reason for the stark health disparities in Pakistan is that many healthcare facilities are located in more densely populated urban areas. Thus, rural populations have disproportionately low access to healthcare. For example, Matiari, a rural town in Pakistan's Sindh province, has been burdened by chronic poverty. In addition, the roadway infrastructure in Matiari is severely underdeveloped, consisting of only 178 kilometers (48 mi) of quality roads within the district's 141,000 hectares (544 sq mi). The closest large city, Hyderabad, is thirty kilometers away along the only national highway in the region [6]. This lack of road infrastructure creates a significant barrier to access to healthcare, and so in Matiari and other remote areas, these transportation-related difficulties and a high disease prevalence combine to affect the health of these rural populations adversely.

This study has investigated the relationship between geospatial factors and the morbidity of acute infections in the pediatric population, specifically acute respiratory infections (ARIs) and diarrhea. Pneumonia is the leading cause of death globally for pediatric patients between 1-59 months of age, accounting for 12.8% of deaths in 2015 [2]. Diarrhea is the third highest cause of mortality in this age group, accounting for 8.6% of deaths [2]. In Pakistan, diarrhea has been estimated to cause 15% of deaths in children under the age of five [7]. We hypothesized that greater travel distance between pediatric patients and healthcare facilities is associated with higher morbidity due to acute infection. We have also explored the role of nutritional status. Literature shows that recurrent acute infections in pediatric patients negatively impact their nutritional status, which places them at higher risk for growth impairment and undernutrition, which, in turn, increases their risk of infection [8]. Both diarrhea and pneumonia have been associated with adverse outcomes in undernourished patients [3, 9, 10].

We have used geographical information systems (GIS) technology to explore the relationship between health and distance to healthcare. GIS technology comprises geospatial analytical platforms that are "computer-based system(s) used for collecting, editing, visualizing, and analyzing spatially-referenced data [11]." The use of GIS in public health and epidemiology has risen dramatically in the past decade. For example, GIS has been used to map the disease epidemiology such as Chagas disease, dengue fever, and vector-borne arboviruses [11-13]. Recently, GIS was used in vaccine trial development for preventable tropical diseases by gathering the geographical coordinates of the population under surveillance and monitoring disease prevalence in these populations over time [14]. Enabling the visualization of disease trends and disease clustering concerning geographic proximity, GIS provides valuable information on the relationship between disease and geography, clarifying the interplay of so-called "socioecological exposure" and illness [15]. In our study, we present a framework for the analysis of the spatial

distribution of patients in relation to healthcare services in Matiari, Pakistan, to evaluate how a lack of physical access to care impacts the burden of ARI and diarrhea, as well as growth in pediatric patients.

2. Materials and Methods

2.1 Study participant enrollment

Data was collected via the Study of Environmental Enteropathy and Malnutrition (SEEM), a prospective inception cohort study investigating environmental enteropathy in Matiari, Pakistan. Study participants were recruited through community-based surveillance and monitoring of the population. The study enrolled 416 children under the age of 2 in Matiari between 2016 and 2019. Detailed recruitment methods and inclusion and exclusion criteria are detailed elsewhere in the study by Iqbal et al. [16]. Anthropometric data and disease prevalence data were collected by community health workers who visited the children's homes. Subjects were followed for 24 months or until a study participant reached two years of age. Weekly visits included a survey documenting daily symptoms of diarrhea or ARI.

2.2: Defining clinical outcomes: ARI events, diarrhea events, and growth parameters

The threshold for diarrhea or ARI events was established when a child showed signs and symptoms for a minimum of 2 days, followed by a 7-day symptom-free interval. A child was determined to have had a "diarrhea day" when they excreted three or more loose or liquid stools in one day. An ARI day occurred when a subject reported cough or shortness of breath. Prevalence of diarrhea and ARI events was recorded as the number of sick days/observed days multiplied by 365 days. This value was used to represent morbidity. At monthly visits, the community health workers also obtained weight and length measurements, which were used to contrast with the nutritional status and growth of the child. These measurements were used to calculate the height by age Z-score (HAZ), weight for age Z-score (WAZ), and weight for height Z-score (WHZ). The nutritional status of the children was also closely monitored. We want to note that nutritional intervention in the form of education was provided to the parents when the child was at 6 months of age. At 9 months of age, if a child was more than 2 standard deviations below the WHZ, they received nutritional intervention in the form of a supplement at 10 and 11 months. If the study participants were less than 2 standard deviations below the WHZ, they continued with weekly and monthly follow-ups by the community health workers [16].

2.3 Geographical Data Collection of Healthcare facilities and participant residence

A comprehensive list of healthcare facilities in Matiari was compiled by searching historical records and using up-to-date satellite imaging through accessing publicly available map data. We identified a total of 21 healthcare facilities, including 7 Basic Health Units (BHU); 6 Dispensaries; 1 District Headquarter Hospital; 1 Taluka Hospital; 3 Rural Health Centers (RHC); and 3 Maternal Health Centers (MHC). The District Headquarter Hospital and Taluka Hospital were grouped as secondary care hospitals. Spatial coordinates of each healthcare facility were acquired using Google Earth satellite imaging and were recorded as latitude and longitude. In addition, the spatial coordinates of each subject's home were collected upon enrollment in the study. The distance of each subject to the nearest healthcare facility was computed in kilometers (km).

2.4 Calculation of anthropometric data

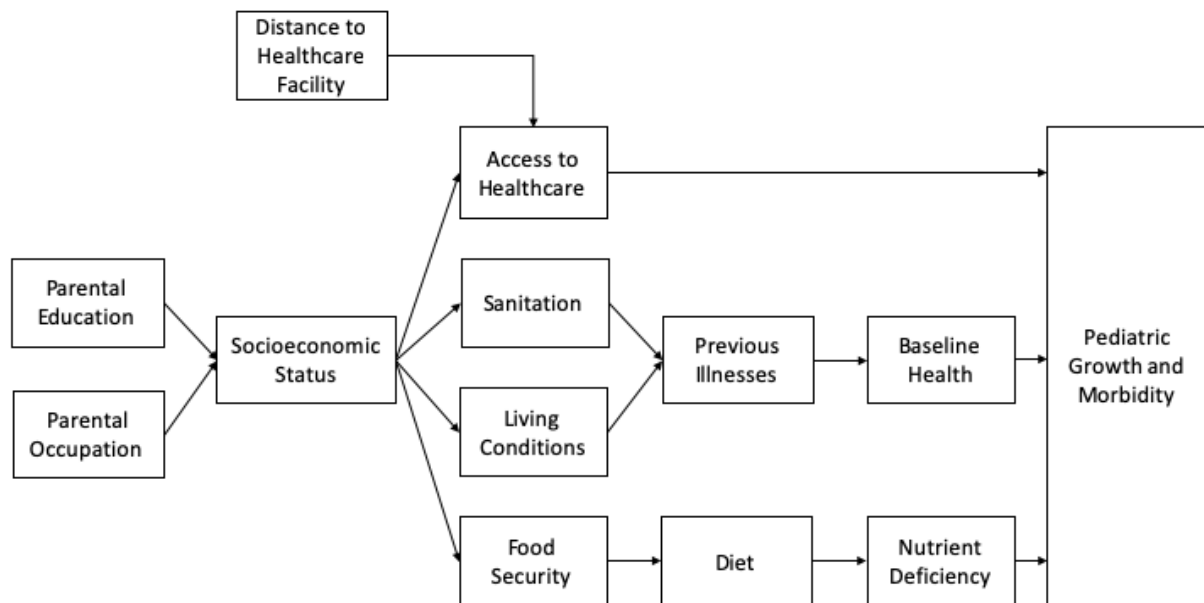
Changes in growth, or deltas, were calculated for the anthropometric variables, measuring the difference from 0-6 months to 24 months of age. All continuous variables,

including the prevalence of diarrhea, the prevalence of ARI, delta weight, delta HAZ, delta WAZ, and delta WHZ, were expressed as mean (\pm standard deviation; S.D.). The Pearson correlation coefficient (r) was used to calculate the correlation between distance from healthcare facilities and study participant growth and the prevalence of ARI and diarrhea. Significance was tested with a student's t-test on the calculated Pearson's correlation coefficient. A p-value of less than 0.05 was considered statistically significant.

2.5 Random Forest regression modeling of other continuous variables

We recognized that factors other than geographic distance from hospitals influence the growth and morbidity in pediatric patients. Data shows that children in lower socioeconomic groups have a greater incidence of disease morbidity worldwide [17, 18]. Therefore, we increased the scope of our project to investigate other factors influencing the occurrence of acute infections in pediatric patients. To adopt a holistic approach in our investigation of factors contributing to disease burden, we have analyzed the relative contribution of other environmental variables such as household caste and parental demographics to determine their relative impact on the outcomes we established for this project. First, a conceptual causal pathway was created to depict the multi-faceted relationship of the various determinants of pediatric growth and morbidity (Figure 1).

Figure 1. Causal pathways depicting the relationships between various parameters and pediatric growth and morbidity



Next, a random forest regression analysis was used to study these continuous environmental variables that influenced the identified growth or morbidity parameters, including distance to healthcare facilities. Random forest is a commonly used machine learning model that uses many decision trees working together "as an ensemble." The regression model was trained to predict correlated parameters using distance, growth, and other clinically relevant variables such as parental caste, number of household members, and maternal age. Feature importances were extracted from the trained random forest model to rank features with the intent of further exploring these features as potential confounders in the assessment of distance and disease prevalence. For training and testing the model performance, an 80-20 split of data was used respectively. Our random forest

regression model was created using sklearn's Random Forest Classification and Regressor packages and implemented in Python [19].

3. Results

3.1. Descriptive summary of data collection

Our study included 416 children (61% male) with a mean age at enrollment was 4.2 mo (\pm 1.0 mo). The mean distance from healthcare centers was 2.3 km (\pm 1.1 km), and the mean weight difference from 0-6 till 24 months was 3 kg (\pm 0.9 kg). Additional background demographic data are presented in **Table 1**. The prevalence of diarrhea and ARI among the 297 participants was 50% (\pm 34.1) and 54% (\pm 54), respectively.

Table 1. Sociodemographic and Presentation Characteristics of Subjects Enrolled in SEEM Study between 2016-2019 in Matiari, Pakistan (n= 416, 61% Male)

Variable	Mean (\pm SD)
Age at Enrollment (mo)	4.2 (\pm 1.0)
HAZ at Enrollment (mo)	-2.3 (\pm 1.4)
WAZ at Enrollment (mo)	-3.6 (\pm 1.4)
WHZ at Enrollment (mo)	-2.6 (\pm 0.7)
Weight Difference from 0-6 to 24 months (kg)	3.0 (\pm 0.9)
Prevalence of AR per year (*)	54.2 (\pm 53.3)
ARI Episodes (**)	5.0 (\pm 3.6)
Prevalence of Diarrhea per year (***)	50.0 (\pm 34.3)
Diarrhea Episodes (****)	14.8 (\pm 17.7)
Distance from Healthcare Facility	2.3 (\pm 1.1) km 1.4 (\pm 0.7) mi
Distance from Secondary Hospital	6.5 (\pm 3.5) km 4.0 (\pm 2.2) mi
Distance from Maternal Health Center	6.3 (\pm 3.6) km 3.9 (\pm 2.2) mi
Distance from Basic Health Unit	3.6 (\pm 1.4) km 2.2 (\pm 0.9) mi
Distance from Rural Health Center	9.0 (\pm 4.8) km 5.6 (\pm 3.0) mi
Distance from Dispensary	4.0 (\pm 1.8) km 2.5 (\pm 1.1) mi
HAZ = height-for-age Z score	
WAZ = weight-for-age Z score	
WHZ = weight-for-height Z score	
ARI = Acute Respiratory Infection	
ARI days = Subject reporting cough and/or shortness of breath	
* Prevalence of ARI = (number of ARI days / observed days) \times 365	
** ARI episode = ARI for minimum of 2 days with signs followed by a sign free interval of 7 days	
Diarrhea day = Subject excreting three or more loose or liquid stools in one day	
*** Prevalence of Diarrhea = (Number of Diarrhea days / observed days) \times 365	
**** Diarrhea Episode = New episode is defined as not meeting the diarrheal definition for at least 2 days	
1 kilometer = 0.623 miles	

3.2 Prevalence of ARI and diarrhea correlates with study participant distance from healthcare facilities.

Table 2 reports the relationship between growth parameters and the prevalence of acute infections and the correlation between morbidity and distance from healthcare facilities. Our results indicate a positive relationship between ARI and distance, suggesting that participants located closer to secondary hospitals have a lower prevalence of both ARI ($r=0.154$, $p<0.010$) and diarrhea ($r=0.228$, $p<0.001$). Our results revealed a similar relationship between disease burden and proximity to MHCs. Participants living closer to MHCs had a statistically significant decrease in the prevalence of ARIs ($r=0.185$, $p<0.002$) and diarrhea ($r=0.223$, $p<0.001$). Conversely, subjects that lived near BHUs and RHCs demonstrated a higher burden of disease. The prevalence of ARI was increased in patients living in closer proximity to BHUs ($r=-0.126$, $p=0.034$) and RHCs ($r=-0.212$, $p<0.001$). Additionally, proximity to RHCs was also associated with an increased prevalence of diarrhea ($r=-0.258$, $p<0.001$). There was no statistically significant relationship to report between the distance from any type of healthcare facility and the delta weight, delta HAZ, delta WAZ, or delta WHZ scores (**Table 2**). Proximity to dispensaries also did not play a statistically significant role in daily morbidity or growth.

Table 2. Correlation of growth parameters and prevalence of acute infections with distance from healthcare facilities

Variable	Secondary Hospital		Maternal Health Center		Basic Health Unit		Rural Health Center		Dispensary	
	Correlation Coefficient	p-value	Correlation Coefficient	p-value	Correlation Coefficient	p-value	Correlation Coefficient	p-value	Correlation Coefficient	p-value
ARI Episodes	0.057	0.334	0.067	0.258	-0.081	0.176	-0.154	0.009	0.022	0.711
Prevalence of ARI	0.154	0.009	0.185	0.002	-0.126	0.034	-0.212	<0.001	-0.044	0.466
Diarrhea Episodes	0.151	0.010	0.163	0.006	-0.079	0.183	-0.260	<0.001	-0.057	0.335
Prevalence of Diarrhea	0.228	<0.001	0.223	<0.001	-0.044	0.459	-0.212	<0.001	-0.005	0.940
Delta Weight	-0.039	0.515	-0.049	0.416	-0.108	0.070	0.066	0.271	-0.026	0.662
Delta HAZ	0.067	0.260	-0.014	0.816	-0.104	0.080	0.099	0.097	-0.009	0.885
Delta WAZ	0.008	0.893	-0.021	0.723	-0.087	0.144	0.074	0.215	-0.046	0.442
Delta WHZ	-0.044	0.464	-0.024	0.687	-0.041	0.497	0.030	0.617	-0.042	0.477

Results of student's t-test on Pearson's correlation coefficient

P < 0.05 considered significant

ARI = Acute Respiratory Infection

ARI day = Subject reporting cough and/or shortness of breath

Prevalence of ARI = (number of ARI days / observed days) * 365

Diarrhea day = Subject excreting three or more loose or liquid stools in one day

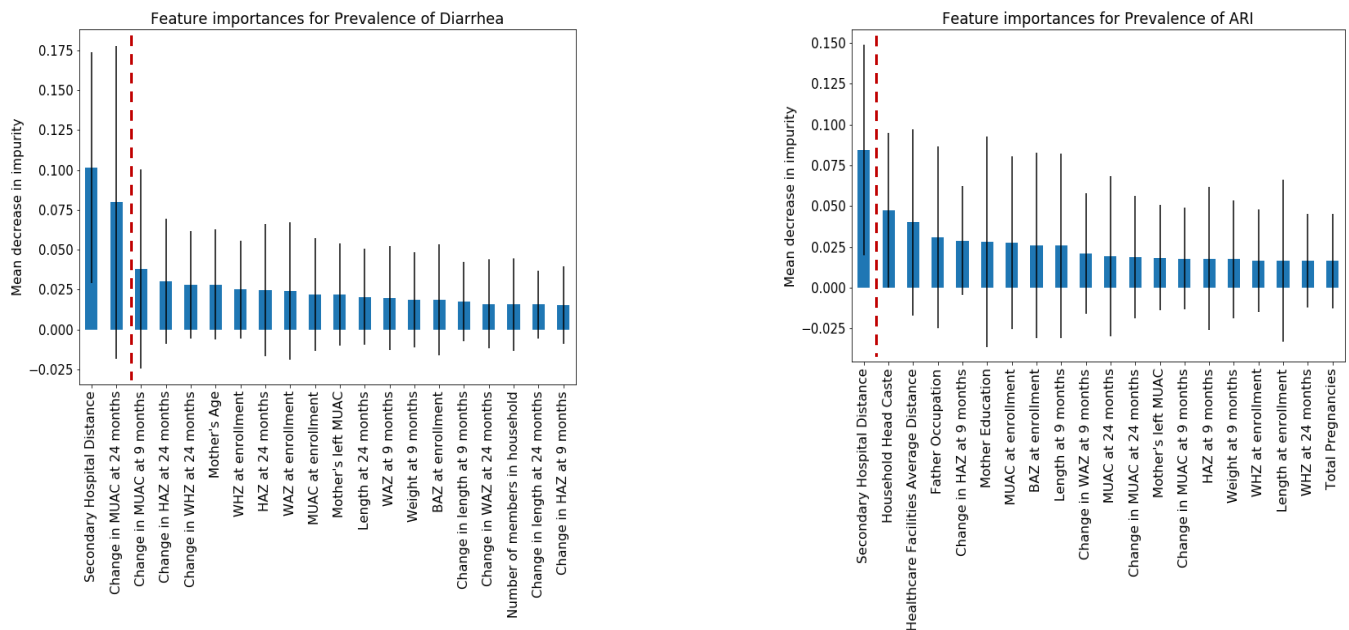
Prevalence of Diarrhea = (number of Diarrhea days / observed days) * 365

HAZ = height-for-age Z score; WAZ = weight-for-age Z score; WHZ = weight-for-height Z score

Delta indicates the change over time from subject enrollment to study completion at 24 months of age

3.3 Distance has the highest variable importance in ARI and diarrhea outcomes.

In our random forest analysis of variables deemed contributory to increased prevalence of diarrhea and ARI or growth parameters, we found that distance was the features were found to have the highest variable importance in relation to the prevalence of both ARIs and diarrhea. It was also observed that the change in mean upper arm circumference (MUAC) at 24 months was found to be a feature of importance in relation to diarrhea, but not in ARI. Other features, such as household caste, father occupation, and the WHZ and HAZ scores at various age points, did not appear as important on our model (**Figure 2**). The mean decrease in impurity is a feature importance metric describing the improvement in predictability observed due to each variable. Based on the importance score and ranking, distance-based features were most important for both ARI and diarrhea. All the other features (**Figure 2**, after the red dashed line) showed high standard deviations, which in some instances was a negative value and, in most instances, crossed zero. These high standard deviations, combined with low feature importance, suggest that these features had limited to no contribution to the predictability of the prevalence of ARI or diarrhea.



A. Results of the random forest analysis for diarrhea

B. Results of the random forest analysis for acute respiratory infection (ARI).

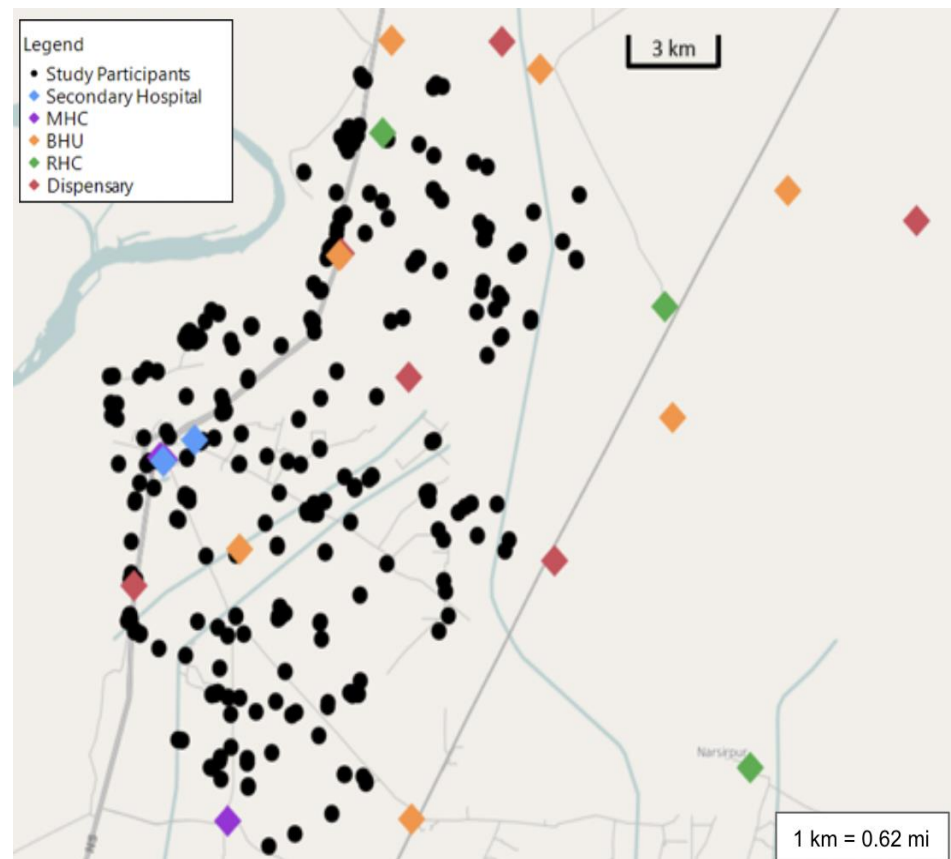
Figure 2. Feature Importance Graphs for ARI (A) and diarrhea (B). Note: The mean decrease in impurity is a feature importance metric describing the improvement in predictability observed due to each variable. Based on the importance score and ranking, distance-based features were most important in both ARI and diarrhea, followed by the change in the child's MUAC for diarrhea. All the other features (after the red dashed line) showed high standard deviations, which in some instances was a negative value and, in most instances, crossed zero. These high standard deviations, combined with low feature importance, suggest that these features had limited to no contribution to the predictability of the prevalence of ARI or diarrhea. MUAC=Mid upper arm circumference; HAZ=Height for Age Z-score; BAZ=Body Mass Index for Age Z-score; WHZ=Weight for Height Z-score; WAZ=Weight for Age Z-score

4. Discussion

In this study, we present a framework for the use of geospatial data to explore the relationship between a patient's geographic proximity to healthcare facilities and the prevalence of ARI and diarrhea as well as growth. We have also presented a framework to minimize the confounding effects of a myriad of other variables, such as maternal and paternal sociodemographic characteristics, by using a random forest regression model to show that distance from healthcare facilities has high relative importance in increased morbidity from ARIs and diarrhea.

Our results indicated that distance to various tiers of healthcare facilities was associated with different disease-related outcomes. Study participants located closer to RHCs and BHUs reported a higher prevalence of both diarrhea and ARI as compared to secondary and tertiary care facilities. A review of the map of healthcare facilities reveals that RHCs and BHUs are more likely to be found in more rural areas of Matiari, providing a socio-ecologic explanation for the increased disease prevalence in these areas (Figure 3). Thus, while proximity to healthcare facilities may have some role in child health, our data suggest that living in more rural areas may also negatively impact childhood morbidity. Conversely, children living closer to secondary hospitals and MHCs had a lower prevalence of diarrhea and ARI. Notably, the secondary hospitals and MHCs are located in more urban areas and along main roads, thus making them more accessible to subjects (Figure 3). Our random forest analyses provide further evidence that distance between subjects and healthcare facilities is related and has an impact on pediatric morbidity (Figure 2).

Figure 2. Map of healthcare facilities and study participants in Matiari, Pakistan.



MHC: Maternal Health Center

BHU: Basic Health Unit

RHC: Rural Health Center

Data was collected via the Study of Environmental Enteropathy and Malnutrition (SEEM), a prospective inception cohort study that investigated environmental enteropathy in Matiari, Pakistan, maps.google.com, and reliefweb.int.

Although 70% of Pakistan's population lives in rural regions, the vast majority of healthcare facilities and 85% of doctors are located in more populated areas (16). This discrepancy places a fiscal and infrastructure burden on rural provinces such as Sindh. There are three doctors for every 10,000 residents in the Sindh province, while Pakistan has seven doctors for every 10,000 residents [20]. The urban versus rural health care imbalance can be further demonstrated by the fact that 23% of Sindh's healthcare facilities, which hold 40% of the beds, are located in the six districts of urban Karachi, the largest city in Pakistan [5]. The scarcity of healthcare facilities in rural Pakistan often necessitates extensive travel for patients to seek care. Physical distance has been recognized as a significant barrier to accessing healthcare, as previous studies have concluded that distance has the potential to negatively impact the overall health outcomes of patients [21]. While extended travel is challenging for any ill patient, it is challenging for patients who live in remote areas. These areas may lack adequate road systems and challenging topography [22]. The road network infrastructure in Matiari is severely underdeveloped, and most individuals travel on foot or by rickshaw, making travel with young children difficult, dangerous, and expensive [23]. Puett et al. noted that Sindh weather conditions and seasonality could pose significant challenges, and patients traveling on foot are at risk of exposure to the elements. They reported that frequent flooding during the rainy season and higher temperatures in the summer months could dissuade families from traveling to seek care [22]. On average, the subjects enrolled in our study lived 2.25 km away from the closest healthcare facility, making travel on foot even more difficult. In addition, once families transport the child, they arrive at healthcare facilities that serve a larger geographic area consisting of a sizeable patient population. Therefore, parents

who have overcome the burden of travel are likely to have long wait times. This places an additional burden on patients and family members, as they must sacrifice a significant amount of time to travel to and wait to receive care [23].

Our random forest analyses support the hypothesis that distance to healthcare facilities is associated with childhood morbidity. Furthermore, using our random forest analysis, we aimed to clarify the complex interplay between various socioeconomic factors and the health of this population of pediatric patients. The relationship between socioeconomic status and health has been extensively described in the literature, and it has been established that lower status has a negative impact on the physical health of children [17, 18, 23]. The main factors contributing to socioeconomic status include household income, education, and parental occupation, and research has shown that parental socioeconomic status is intrinsically connected to their children's outcomes [24]. Furthermore, a global view reveals significant gaps in childhood mortality between high and low-middle income countries, a trend that extends to inter-country populations with wealth disparity. This disparity has been attributed to lower disease resistance in impoverished children due to poor nutrition, hazardous living conditions, and reduced access to healthcare [25]. Unfortunately, the relationship between health and socioeconomic status represents a negative feedback loop, where frequent or severe illnesses may harm one's social position [18]. However, in our random forest analysis, the relative importance of these features was low, and their standard deviations were high, indicating that based on our data, these social features had limited association with the prevalence of diarrhea and ARI. (Figure 3).

In Pakistan, urban facilities, such as district headquarters and tehsil hospitals employ numerous staff members and are frequently visited. In contrast, facilities in remote regions are more susceptible to severe damage or unreported abandonment due to lower utilization and lack of staff. Even within the staffed and functioning facilities, the services and supplies required to provide comprehensive care are often extremely limited [26]. Inpatient facilities, usually with limited beds, are often burdened with overcrowding, while outpatient clinics consistently have long wait times [26]. These shortcomings impact pediatric patients especially hard as children were among the most vulnerable patient populations. Given that only 42% of facilities offer pediatric services, pediatric growth monitoring is not widely available. On average, 52% of facilities can assess and treat ARI and diarrhea in children under 5, and only 36% can assess and manage childhood nutrition [10, 26]. We have provided data that suggest that distance to health care facilities may be another factor adding to the burden of child health in Pakistan.

To our knowledge, this is the first study to examine the relationship between proximity to healthcare centers and childhood morbidity and growth in the Sindh region of Pakistan. We have outlined a framework for the use of GIS technology to explore these variables in other similar populations. This work highlights the need for future efforts to improve access to healthcare facilities for rural patient populations. Additionally, this study prompted the establishment of an up-to-date and complete list of healthcare facilities in Matiari, Pakistan, which will aid in future research efforts in this area. However, our study was also faced with several limitations. Although the original data was collected via SEEM, a prospective inception cohort study, our data analysis was performed retrospectively. Thus, we would like to have collected data points of interest if real-time analysis had taken place. There was no data collected regarding which healthcare centers the study participants received their care. Therefore, we should note that our analysis assumes that subjects sought care at the government-funded healthcare facilities closest to their place of residence and not private healthcare institutions. While this assumption is likely to sound given that Matiari is a rural area and most transportation takes place by foot, it cannot be confirmed. Additionally, it was assumed that subjects did not move residence for the duration of the study. Furthermore, healthcare facilities were grouped based on their type, which assumes that the care provided at each type of

healthcare facility in one group is roughly equal. Another critical component of rural health care infrastructure in Pakistan are community-based lady health workers (LHWs). LHWs provide limited preventive and curative maternal and child health services, including childhood immunizations and essential treatment of diarrhea and ARI. The eligibility criteria for recruitment of LHWs is minimum 8th grade, followed by a 15-month training program. In Matiari, more than 500 LHWs work to cover 60% of the area, again covering more populated areas. We have not included the density of LHWs and the populations they serve in our model [27]. Other potential confounding factors to be addressed in future studies include cultural and social factors such as healthcare literacy, willingness to seek care, and beliefs in alternative medicine.

5. Conclusions

Many factors contribute to the growth and health of pediatric patients, including proximity and access to quality healthcare and timely nutritional and medical interventions. Increased distance from such facilities was found to contribute to the increased morbidity of ARI and diarrhea in children in Matiari, Pakistan. Given the imbalance of urban and rural healthcare facilities in the Pakistani healthcare system, we conclude that it is exceedingly difficult for patients living in remote areas to access quality healthcare. This study highlights the need for comprehensive healthcare reform in developing countries, specifically in remote areas with a high burden of disease morbidity. Further exploration of variables associated with poor health outcomes must be carried out to improve our understanding of the pitfalls of accessing quality healthcare and the future directions for public health experts, policymakers, and stakeholders to improve the health in these rural populations.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1

Supplementary Table 1. Comprehensive list of Healthcare Facilities in Matiari, Pakistan				
Union Council	Health Facility Name	Health Facility Type	Latitude	Longitude
Bao Khan Pathan	BHU Bao Khan Pathan	BHU	25.607036	68.598257
Faqir Nooh Hothiani	BHU Faqir Nooh Hothiani	BHU	25.705407	68.513972
Oderolal Village	BHU Tahir Hinngoro	BHU	25.666246	68.632816
Oderolal Village	BHU Oderolal Village	BHU	25.697929	68.558528
Sekhat	BHU Sekhat	BHU	25.649272	68.498243
Shah Alam Shahwari	BHU Laloo Mangwano	BHU	25.572692	68.468336
Tajpur	BHU Allah Dinosand	BHU	25.50227	68.519977
Bao Khan Pathan	G.D. Shahmeer Jiwari	Dispensary	25.61755	68.519116
Bao Khan Pathan	DCD Palijani	Dispensary	25.569672	68.562855
Faqir Nooh Hothiani	G.D. Sher Muhammad Thorha	Dispensary	25.705117	68.54707
Oderolal Village	G.D. Dhaloo Khaskeli	Dispensary	25.658382	68.671428
Sekhat	G.D. Juma Shaok	Dispensary	25.64985	68.498859
Shah Alam Shah	G.D. Soomro Chad	Dispensary	25.563194	68.436686
Matiari	Matiari Health Services	Private hospital	25.601161	68.454878
Matiari	Taluka Hospital Matiari	Taluka Hospital	25.595665	68.445469
Matiari	Taluka Hospital Saeedabad	Taluka Hospital	25.959014	68.375756
Matiari	Hala District Hospital	District Headquarter Hospital	25.818910	68.417196
Nasarpur*	RHC Nasarpur	RHC	25.515687	68.621544
Oderolal Station	RHC Oderolal Station	RHC	25.636002	68.595958
Sekhat	RHC Khyber	RHC	25.68128	68.511284
Matiari	Ali Medical Center & Maternity Home	MHC	25.59605	68.445869
Matiari	Noor Jahan Clinic & Maternity Home	MHC	25.596633	68.444591
Matiari	Haq Bahod Clinic & Maternity Home	MHC	25.501785	68.464783

Union council is the smallest administrative unit in in Pakistan.

*Nasarpur is not located in Matiari district but is in close proximity of the union council.

BHU = Basic Health Unit RHC = Rural Health Center MHC = Maternal Health Center

Author Contributions: Conceptualization, Elise Corden, Saman Siddiqui, Yash Sharma, Aman Shrivastava, Sean Moore, Syed Asad Ali and Sana Syed; Data curation, Yash Sharma, Muhammad Raghbir, William Adorno III, Lubaina Ehsan, Aman Shrivastava, Fayaz Umrani, Najeeb Rahman, Rafey Ali and Najeeha Iqbal; Formal analysis, Saman Siddiqui, Yash Sharma, William Adorno III, Lubaina Ehsan, Sheraz Ahmed, Fayaz Umrani, Najeeb Rahman and Sana Syed; Funding acquisition, Syed Asad Ali and Sana Syed; Investigation, Saman Siddiqui, Muhammad Raghbir, Lubaina Ehsan, Aman Shrivastava, Sheraz Ahmed, Fayaz Umrani and Najeeha Iqbal; Methodology, Elise Corden, Saman Siddiqui, William Adorno III, Lubaina Ehsan, Aman Shrivastava, Sheraz Ahmed, Najeeb

Rahman and Sana Syed; Project administration, Syed Asad Ali and Sana Syed; Resources, Najeeb Rahman, Najeeha Iqbal and Syed Asad Ali; Software, William Adorno III; Supervision, Sean Moore, Syed Asad Ali and Sana Syed; Validation, William Adorno III, Fatima Zulqarnain and Sana Syed; Visualization, Sana Syed; Writing – original draft, Elise Corden, Saman Siddiqui, Yash Sharma, Muhammad Raghieb, Fatima Zulqarnain and Lubaina Ehsan; Writing – review & editing, Fatima Zulqarnain and Sheraz Ahmed.

Funding: Research reported in this manuscript was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number K23DK117061- 01A1 (Syed) Bill & Melinda Gates Foundation under award numbers OPP1066203 (Ali)

Institutional Review Board Statement: Ethical approval was obtained from the Aga Khan University Hospital Ethical Review Committee (3836-Ped-ERC-15), Cincinnati Children’s Hospital Medical Center (CCHMC, Study ID 2016–0387), Washington University in St Louis Human Research Protection Office (IRB ID 201111065), and the University of Virginia Institutional Review Board (UVa-IRB, Study ID 19856). All participants were enrolled after they signed an informed consent. After complete disclosure, a signed informed consent was obtained from each participant’s parent or legal guardian. The consent was obtained, preferably, where the participant resided at the time of enrollment. If the parent(s)/guardian agreed to participate in the study, the consent form was signed, or an impression of their thumb must be provided. The investigator and a witness also signed the form.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study

Data Availability Statement: This study used data collected as part of the Study of Environmental Enteropathy and Malnutrition (SEEM.)

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Children: improving survival and well-being. 2020 [August 24, 2021]; Available from: <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality>.
2. Liu, L., et al., Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet*, 2016. 388(10063): p. 3027-3035.
3. Black, R.E., et al., Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet*, 2008. 371(9608): p. 243-60.
4. Zaidi, S.I., N; Riaz, A, Primary health care systems (PRIMASYS): comprehensive case study from Pakistan, W.H. Organization, Editor. 2017: Geneva
5. Hassan, A.M., Khalid ; Allah Bukhsh, Hudebia, Healthcare System Of Pakistan. *International Journal of Advanced Research and Publications*, 2017. 1(4).
6. USAID, PAKISTAN EMERGENCY SITUATIONAL ANALYSIS: District Matiari. 2014.
7. Quadri, F., et al., Health care use patterns for diarrhea in children in low-income periurban communities of Karachi, Pakistan. *Am J Trop Med Hyg*, 2013. 89(1 Suppl): p. 49-55.
8. de Onis, M. and F. Branca, Childhood stunting: a global perspective. *Matern Child Nutr*, 2016. 12 Suppl 1: p. 12-26.
9. Kossman, J., et al., Undernutrition in relation to childhood infections: a prospective study in the Sudan. *Eur J Clin Nutr*, 2000. 54(6): p. 463-72.
10. Olofin, I., et al., Associations of suboptimal growth with all-cause and cause-specific mortality in children under five years: a pooled analysis of ten prospective studies. *PLoS One*, 2013. 8(5): p. e64636.
11. Shaw, N.T. and S.K. McGuire, Understanding the use of geographical information systems (GISs) in health informatics research: a review. *BMJ Health & Care Informatics*, 2017. 24(2): p. 228-233.

12. Duncombe, J., et al., Geographical information systems for dengue surveillance. *Am J Trop Med Hyg*, 2012. 86(5): p. 753-5.
13. Diptyanusa, A., L. Lazuardi, and R.H. Jatmiko, Implementation of geographical information systems for the study of diseases caused by vector-borne arboviruses in Southeast Asia: A review based on the publication record. *Geospat Health*, 2020. 15(1).
14. Khan, O.A., et al., Geographical information systems and tropical medicine. *Ann Trop Med Parasitol*, 2010. 104(4): p. 303-18.
15. Ali, M., et al., Use of satellite imagery in constructing a household GIS database for health studies in Karachi, Pakistan. *International journal of health geographics*, 2004. 3(1): p. 20-20.
16. Iqbal, N.T., et al., Study of Environmental Enteropathy and Malnutrition (SEEM) in Pakistan: protocols for biopsy based biomarker discovery and validation. *BMC Pediatrics*, 2019. 19(1): p. 247.
17. Anderson, N.B. and C.A. Armstead, Toward understanding the association of socioeconomic status and health: a new challenge for the biopsychosocial approach. *Psychosom Med*, 1995. 57(3): p. 213-25.
18. Poulain, T., M. Vogel, and W. Kiess, Review on the role of socioeconomic status in child health and development. *Curr Opin Pediatr*, 2020. 32(2): p. 308-314.
19. Pedregosa, F., et al., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2012. 12.
20. Zaidi, S.A. and S.A. Sahibzada, Issues in Pakistan's Health Sector [with Comments]. *The Pakistan Development Review*, 1986. 25(4): p. 671-682.
21. Douthit, N., et al., Exposing some important barriers to health care access in the rural USA. *Public Health*, 2015. 129(6): p. 611-620.
22. Puett, C. and S. Guerrero, Barriers to access for severe acute malnutrition treatment services in Pakistan and Ethiopia: a comparative qualitative analysis. *Public Health Nutr*, 2015. 18(10): p. 1873-82.
23. Adler, N.E. and K. Newman, Socioeconomic disparities in health: pathways and policies. *Health Aff (Millwood)*, 2002. 21(2): p. 60-76.
24. Vukojevic, M., et al., Parental Socioeconomic Status as a Predictor of Physical and Mental Health Outcomes in Children - Literature Review. *Acta Clin Croat*, 2017. 56(4): p. 742-748.
25. Victora, C.G., et al., Applying an equity lens to child health and mortality: more of the same is not enough. *Lancet*, 2003. 362(9379): p. 233-41.
26. Saleh, R., Secondary Level Health Facilities Assessment in Sindh 2011, World Health Organization: Islamabad, Pakistan.
27. Salam, R.A., et al., Potential for task-sharing to Lady Health Workers for identification and emergency management of pre-eclampsia at community level in Pakistan. *Reprod Health*, 2016. 13(Suppl 2): p. 107.