A Model for the Creation of a Predictive Healthcare Coverage Map in Yemen

Mark Suprenant
Boston University

Anuraag Gopaluni
Harvard University

Meredith Dyson
UNICEF

Fauzia Shafique
UNICEF

Muhammad Zaman (zaman@bu.edu)
Boston University

Research

Keywords: Yemen, Markov Model, Computational modeling, Diarrheal disease, Rotavirus

Posted Date: January 13th, 2020

DOI: https://doi.org/10.21203/rs.2.20673/v1

Version of Record: A version of this preprint was published on August 5th, 2020. See the published version at https://doi.org/10.1186/s13031-020-00300-1.

Abstract

Introduction

The ongoing war in Yemen continues to pose challenges for health care workers in the country. The fighting has destroyed public infrastructure including primary and secondary health care facilities, hindered the movement of people, food, fuel, medical supplies, and information, and restricted access to and availability of social services including safe drinking water and sanitation. This has led to the increase in the spread of diarrheal diseases, including cholera, which, despite the efficacy of zinc and oral rehydration salt solutions to treat the resulting dehydration, remains one of greatest sources of mortality in children under five years old. In contexts such as Yemen, Health Management Information Systems and Surveillance Systems are weak and unreliable to begin with, with conflict and linked disruption of social services these systems are further weakened making monitoring of the situation and evidence-based planning and implementation even more difficult. Without information on the total number of children suffering from these diseases, it is difficult for health officials and aid organizations to make policy level decisions, inform annual and humanitarian response plans, set targets, mobilize resources, order supplies, deploy resources (human and supplies) and monitor based on needs, leading to poor quality decisions. These reasons, coupled with lack of access, security, and financial and human resources make it even more important in conflict settings, than in non-conflict settings, to know where it is best to invest. This manuscript looks at the development of a computational model designed to draw upon available health data and supplement it with additional sources and acceptable assumptions to provide some of the missing data via health access chart to better inform decision making on the above-mentioned policies. This chart is designed to show what percentage of the total estimated sick population is receiving medical assistance without the need for health workers to place themselves in the way of any additional harm.

Methods

A Markov model, which is a probabilistic model that shows how a population moves between different states overtime, was created based on an analysis of Yemen clinical register data from the Ministry of Public Health collected through a third party hired for monitoring purposes covering the period of May through September of 2018. The model was designed with four states for children to transition between over a weekly basis. The probability that a child transitioned from the *Sick* state to the *In-treatment* state during any given week was a time varying function based on the average precipitation recorded monthly for 115 years and the state of the roads and bridges during that week as assessed by the World Food Program. The model examined the number of children treated, incidence rate, mortality rate, treatment efficacy and treatment mortality. Once validated, the model was run for 2019 to provide the weekly estimated coverage of children being treated for diarrheal diseases throughout all of Yemen.

Results

The model was able to recreate the observed trends in treatment on the ground with no significant difference between model output and provided validation data for all metrics. When combined with infrastructure data, the curve of best fit created for the precipitation values depicted a seasonal increase in the number of estimated new diarrheal cases in children under five and a resulting decreasing in the number receiving treatment. This combination has led predictions for the percent coverage to range between an average weekly minimum of 1.73% around the 28th week of the year to a weekly maximum weekly coverage of just over 5% around the new year.

Conclusion

The model created and presented in this manuscript shows a seasonal trend in the spread of diarrheal disease in children under five living in Yemen. Despite the assistance of aid organizations in attending to those in need, during the mid-year rains up to 98% are unable to receive medical aid. The coverage map indicates that community outreach or other types of assistance where aid proactively goes out to those in need should be scaled up during and just prior to these periods. This would serve to offset the decrease in the number receiving treatment by lessening the prohibitive travel burden on families during these times.

Introduction

In 2015, after a period of political instability, Yemen descended into an ongoing war between a coalition led by Saudi Arabia in support of the internationally recognized government of Yemen against Houthi led forces that had control of the country's capital and much of the rest of country.[1] This war has disrupted nearly every facet of life for those that remain in the country, including an ever-increasing burden on an already strained health care system in the region's poorest country.[1] With 22.2 million needing aid in some form, the situation is viewed as "the worst humanitarian crisis in the world" by the United Nations.[4] Of these 22.2 million 11.3 are in acute need of humanitarian assistance. According to the United Nations, there are currently 8.4 million people who are food insecure, 16.4 million people in need of health services, and 16 million without access of safe water which has resulted in over 65% of Yemeni families employing adverse coping mechanism to deal with the crisis. [2,3]

Prior to the start of the conflict, it was estimated that about half of Yemeni citizens had access to health care facilities.[5] Due to a combination of the direct effects of the conflict (active fighting, insecurity, population movement and damage to health facilities and other health infrastructure) and the secondary effects of economic crisis and collapsing public systems, just over 60% of the country's medical facilities were fully operational by 2016.[6] In the case of diarrheal diseases, which are common even in the absence of cholera and may result from rotavirus or other common childhood infections, it is well accepted that zinc and oral rehydration salt solution (Zn/ORS) is the best and most cost effective course of treatment as the zinc helps to replenish nutrients to the body while reducing the duration and severity of the disease while the ORS solution prevents death due to dehydration. This combined treatment costs an average of $0.50 per dose.[7,8,9] Diarrheal diseases were common in Yemen even in the absence of conflict. According to the last nationally representative survey of population health status, Yemen National Health and

Demographic Survey (YNHDS) 2013, about 31% of children under-five had had diarrhea during the two weeks before the survey.[10] With an under 5 population estimated at 5.16 million, this could translate into 1.60 million cases of diarrheal diseases in this population at any given time.[11] The causes of diarrhea range from viral, bacterial and parasitic with viral causing the largest number of cases and also the most severe ones.[12] Yemen introduced Rotavirus vaccine in 2012 and in line with WHO guidelines the national strategy calls for use of Zn/ORS for all cases and use of antibiotics when specifically indicated.[12, 13] According to the YNHDS, one-third of children with diarrhea were taken to a health facility or provider for treatment, 60% of these children with diarrhea were treated with ORS or increased fluids, and 19% received no treatment either from a health provider or at home.[10]

Given the destruction of water and sanitation infrastructure leading to poor quality and contaminated water, and poor environmental health conditions, the environment is highly conducive to an increase in diarrheal diseases.[14] For displaced populations the living conditions are even worse leading to an increase in the risk of diarrheal disease. At the same time, the prevailing insecurity and lack of resources for seeking care and the deterioration of health systems and services has likely made access to service even worse that the situation in 2013 as per the YNHDS. However, as this info is out of date due to the rapidly changing situation, the assumptions that were true pre-conflict are now only partially, if at all, applicable. Therefore, the Ministry of Health and all its partners in health including the UN and NGOs aid agencies do not have the required information for policy and strategy level decision making, preparing well informed humanitarian response plans for diarrhea diseases, planning at national and sub national levels for supplies (how much ORS and zinc should be ordered for the whole year, how and where should it be pre-positioned and delivered to districts), and for deployment of human and other resources.

A treatment coverage estimate is thus needed but, due to the conflict, it is impractical for the Ministry of Public Health and Population (MoPHP) and health partners to survey each district due to limited staffing, travel difficulty due to poor road infrastructure and lines of conflict, and safety concerns. While a cadre of Community Health Workers (CHWs) has recently been launched and is currently in scale-up phase, the data from this program is not yet accessible to add to the data from health facilities on the number of children detected and treated with diarrhea. Planning of where to place these community-based service providers has been informed by mapping areas that lack access to a health facility, but not by disease burden due to this lack of data on population in need. Ideally, CHWs and each Health facility report would be submitted on time and compiled at the district level and provide information on cases presenting and treatments provided for the catchment area of each worker or health facility as well as the level of disease burden, care seeking habit and service provisions data. This not being the case, the next best thing is needed– an estimate based on data available and concrete evidence-based assumptions.

To tackle this problem and create a coverage map for children under five years with diarrheal disease, a computational modeling-based solution is proposed. This model combines the health data that is available with other logistic data such as conflict and infrastructure status provided through the Logistics Cluster and World Food Program as well as weather predictions based on historical data from the World

Bank to provide a more accurate picture of the number of people treated, with the goal of estimating burden of disease and the underlying lack of access to care, and locating untreated pockets of children under five suffering from diarrheal diseases. By using the synergies of these data types in a modeling frame work, we can combine the accuracy and reliability of active community outreach approaches without adding to the burden of health workers or exposing them to unnecessary danger, key benefits of passive screening.

Methods

As the primary function of the model was to determine the number of children in each state of health at a given point in time, the model's core consisted of a basic Markov model. This category of model has been used for various types of probabilistic modeling over the past 20 years, including epidemiology modeling, especially when the model has a defined number of outcomes, or states.[15] The Markov model used for the creation of the coverage map for Yemen was composed of four different states (*Healthy, Sick, In-treatment*, and *Deceased)* for children to move between based on a set of probabilistic transition rates. The model rests on the following assumptions: any sick person that decides to seek treatment will contact a health care provider in community, in a mobile team or at a health facility, and will be treated; to reach the *In-treatment* state one must have previously resided in the *Sick* state; the *Deceased* state is an absorbing state. The health facility data collected through a third party monitoring exercise provided insight on the number of children under five years old seen and treated for diarrheal diseases, the treatment mortality rate of Zn/ORS solutions used in the region, the efficacy of Zn/ORS solutions in treating the disease and the percentage of the population the facilities monitored are estimated to have served. This information was filtered to examine only diarrheal diseases and to eliminate any incomplete reports where the total number of reported children treated for diarrheal diseases did not equal the sum of all the individual children reported to have diarrheal diseases. Coupled with the most recent overall mortality rates in the literature from the WHO and an estimated incidence rate for Yemen found in the literature, the transitions rates were created to allow the model's output to match the previously mentioned data.[11,16] The final transition probabilities were as listed in Table 1. An overview of the terms used, as well as their values and meaning are provided below in Table 2. These are described in more detail later in the manuscript.

Table1: Transition Rate Probability Table

State Transition	Transition Probability
p_{HH} Healthy to Healthy	.7601
p_{HS} Healthy to Sick	.2388
p_{HT} Healthy to Treated	0
p_{HD} Healthy to Deceased	.0011
p_{SH} Sick to Healthy	.6968
p_{SS} Sick to Sick	(1-(YT+6.2712)/9)
p_{ST} Sick to Treated	(YT-.01296834)/9
p_{SD} Sick to Deceased	.00092631
p_{TH} Treated to Healthy	.9902
p_{TS} Treated to Sick	.00961
p_{TT} Treated to Treated	0
p_{TD} Treated to Deceased	.00019
p_{DH} Deceased to Healthy	0
p_{DS} Deceased to Sick	0
p_{DT} Deceased to Treated	0
p_{DD} Deceased to Deceased	1

List of the probabilities of a child transitioning from one state of health to another in the model.

Table 2: Table of Terms

Object Notation	Summary	Value
Susceptible Population	Total under 5 child population	358498
Cycles	Number of weeks the simulation runs for	52
Maxroad	Max value of roads	variable
Roadmin	Minimum value or roads due to washout	equation
Bridge	Percent of open road segments with open bridges	variable
Nonbridgemin	Minimum road value with operating bridges	equation
Bridgecuttoff	Weather threshold where lower quality roads may begin to experience washout	0.5
Scalemax	Parameter determining how long infrastructure stays at max value	variable
Weatherweight	Weighting of weather on transition rate	0.3
Roadweight	Weighting of infrastructure on transition rate	0.7
Y(t)	Weather function	Equation
Road(t)	Road and bridge infrastructure function	Equation
YT(t)	Transition probability equation from Sick to Treated	Equation

Overview of the variables and parameters used throughout the model.

To help account for non-cholera based diarrheal diseases the transition rate from healthy to sick was further subdivided between general sources of the diseases and Rotavirus, the most common source which is noted as the cause of between 35-60 % of enteric diseases in children under five years old.[18] Rotavirus infections confer natural immunity, protecting against 87% of severe diarrhea cases which increases with subsequent infections.[18] Thus, each rotavirus infection after the first is estimated to result in a 0.13^n chance that the child would become sick from the virus alone where n is the number of previous infections a child has contracted.

Along with this information, other types of data outside of traditional health data were incorporated into the model's transition rates, most notably for the probability determining the likelihood that someone will be willing and able to seek out treatment and thus move from the *Sick* state to the *In-treatment* state. As such, this influences the probability that a sick person will remain sick. The state of the infrastructure and the estimated weather conditions for that time period allows for the creation of an equation incorporating the time varying nature of these data types into the probabilistic decision process of the model. Thus, the model is not only probabilistic but also time varying.

The state of the weather and the quality of the roads and bridges in Yemen, which are already affected by and dependent on the state of the conflict and so serve as a proxy for the severity of the conflict, have been shown to be related functions. The weather component is assumed to be a sinusoidal function with

a period of one year. This weather function was created as a normalized function of best fit from monthly precipitation data for Yemen from 1901 to 2016 published by the World Bank.[19] The range of probabilities vary between 0, which in the context of the model would indicate a perfect storm during the heaviest rainfall of the year preventing anyone from traveling, to 1, which would result in anyone being able to access roads leading to health services during the driest week of the year. This function is further multiplied by a set constant of 97.84% which is to account for the fact that there are 21.6 airstrikes for every 1000 people according to data published by the Yemen Data Project Organization.[20] Like the rains, this can deter people from traveling and further destroy infrastructure, albeit much more directly.

It is estimated that Yemen has about 50,000 km of roads.[21] As the country does not have a functioning rail system to aid in transportation, these roads are the primary means of transportation for people within the country.[21,22] Despite the importance of these roads, reports from the World Bank estimated that only 28% were all-weather paved before the conflict's start, with only 11% of rural roads being paved.[21] These nonpaved road segments have been reported by USAID to be damaged both by airstrikes and seasonal rains which has led to flooding, hindering travel during the rainy season.[23] The model was constructed on the assumption that the rains predominantly affect the coastal regions to the south and west of the country and so only 50% of the unpaved roads potentially experience wash out from the worst of the rains. The bridges have also been affected by both the weather and the war with air strikes and fighting destroying bridges while many others have their access restricted according to maps published by the World Food Program and the Logistics Cluster.[24] These restricted bridges were noted to have detours that "may not be accessible during the rainy season".[24] Thus, when the rains begin the weather function will start to decrease. When the function drops below "bridgecuttoff", a tunable threshold, a percentage of the roads corresponding to passable bridge detours that were previously opened will close. This percentage was designed to scale with the weather and vary from its max and minimum values due to the conflict.

A road can have multiple different conditions depending on which section of the road was examined. Because of this, the roads shown on the maps published by the Logistic Cluster and World Food Program were subdivided into individual road segments. A road segment was defined as a unique stretch of road that connects between marked communities or another road segment. As the road function was dependent on the weather it too was cyclical. While its maximum value occurred when there were no weather based restrictions on travel, this value ultimately was determined from the extent of the conflict as described by Equation 1. These values were based on how extensive the conflict was during the period examined while the weather provided more regular variation determining how useable these open roads were.

$$Roadmax=1-(\%\ closed\ road\ due\ to\ conflict)-(\%\ road\ with\ impassable\ bridge) \quad (1)$$

The minimum was calculated as the percent of road segments left functional after the assumed max wash-out minus the percent of road segments that have a weather dependent detour as shown in Equation 2 below.

$$Roadmin=.64\times Roadmax-bridge \quad (2)$$

With these extrema determined, the road function is as noted below in Equation 3.

$$Road[(t)=] [min\{][([Y[(t) \times(scalemax-nonbridgemin)+nonbridgemin]-bridgedown[(t)),]] MaxRoad\}] \quad (3)$$

Bridgedown is defined as Equation 4 below.

$$\text{Bridgedown}(t) = \begin{cases} \left(1 - \frac{1}{bridgecutoff}\right) \times bridge \times \text{Y}(t), & \text{Y}(t) \leq bridgecutoff \\ 0, & \text{Y}(t) > bridgecutoff \end{cases} \quad (4)$$

The overall transition probability function was then created as a weighted average of the weather function, Y, and the infrastructure and conflict function, Road, as seen in Equation 5.

$$\text{YT}(t) = weatherweight \times \text{Y}(t) + roadweight \times \text{Road}(t) \quad (5)$$

With the weather and road components combined to form the transition rate function for the probability that a person will travel to seek treatment and thus change from the Sick state to the Treated state, it became possible to monitor the percent coverage. It was assumed that each person that made it to a treatment center (*Sick* to *In-treatment* transition states) was treated with Zn /ORS while medical supplies were available. As this transition occurred only after a person became sick, the percent treatment coverage was calculated as:

$$\text{Coverage}(t) = \frac{Treated(t+1)}{Sick(t)} \quad (6)$$

Since those infected during week t would only be transitioning into the *In-treatment* state during the following week, the numerator's week number must be one ahead of the denominator for the calculation.

Results

Table 3: Model validation against published 2016 WHO Life Tables for Yemen.

2016					
WHO Global Health Observatory Data Repository				Mathematical Model	
Indicator	Age Group	Male	Female	Age Group	Number of Children
Cumulative probability of dying	<1-4 years	0.05745	0.05152	<5 Years	.05804
Cumulative number of people dying	<1-4 years	5710.947	5146.47	<5 Years	5804

The WHO section on the table's left shows aggregated values for children under five years old, providing an overall mortality probability as well as the number of children dying throughout the year based on a hypothetical birthrate of 100,000. The Mathematical Model section in the two rightmost columns examines the aggregated values from the model using the same hypothetical population size.

As a means of calibrating the model and to preliminarily assess its accuracy, the model was validated against historical data from 2018 as well as the most recent published information on the under five years old mortality estimates for Yemen provided by the World Health Organization.[18] To better validate the risk of infection for a child, the incidence rate for Yemen published by El Bcheraoui et al. was compared against the model's incidence rate as a means of determining the probability of general diarrheal infections in the country.[25] Running the model for one year from January through December produces overall values for the mortality rate slightly above WHO estimates and an incidence rate just below a prior estimated value from the literature. Discrepancies between the model and the data can in part be attributed to uncertainties in the data due to the highly fluid nature of the situation on the ground and the difficulties in collecting accurate data in the region. Data used has been selected as the most complete and verifiable. Table 3 compares the average general under five mortality of the model run seven times against the published probability of mortality for children under five in Yemen for 2016 while the following figure depicts the comparison for the incidence rate for this same time period.

As an additional source of information to validate the model against, primary data from 463 health facilities were used. This data showed the number of children under five years old suffering from diarrheal diseases treated at a primary health care facility monitored by UNICEF. These numbers were recorded daily and reported monthly in a clinical data registry. This data is beneficial as it provides information about both the total number of children treated for diarrheal diseases as well as the disease's course in terms of the child's health improving, deteriorating or dying. However, we note that this data does not capture the complete picture as sometimes the number of children treated during a time period does not equal the sum of those children whose illnesses improved, worsened or resulted in death over the same period. In this case, only data where these values matched were used for model validation. Having the model match these data points indicated that the model's output was matching the conditions observed in the health facilities, providing more assurances of the model's validity.

Compared against the health facility data for the five-month period ranging from May to September 2018 the model could recapitulate the observed mortality rate of children that were treated for diarrheal diseases averaged across the 463 facilities reporting complete data. The under-five population was estimated at 5.16 million children.[26] Half of this population was unable to reach places of care before the war started.[11] Of these facilities, the percentage that are estimated to be open and completely operating assuming that any partially operation facilities average to about half of one completely operational facility was estimated at about 64% while only 53% of the facilities in the data set had complete data on the number of diarrheal cases they saw for children under five years old, which came from a list consisting of about 41% of all the primary care facilities monitored by UNICEF. Based on this, the susceptible population that the data is most likely resulting from is 5,156,951 x 50% x 64.26% x 53% x 40.82% = 358498 children and so this value was used for the validation results. A side by side comparison of the model and the metrics used for validation is provided in Figure 3. From this comparison, it was observed that the model's tracking of the number of children treated was also in agreement with the health facility data collected through a third party and provided by UNICEF.

After comparing the model's output against the data, the model was viewed as sufficiently validated. Predictive estimates for the percent coverage were then able to be made with an increased degree of confidence. The percent coverage was estimated on a weekly basis by comparing the values of the total number of sick children according to the model against the newly validated number of treated children at each time-point. The first weekly coverage assessment was during the same time as the 2018 validation data. These predictions yielded the following treatment percent coverage and trends shown in Figure 4 for a population matching the disease prevalence of 31% from the last National survey conducted in 2013.

While useful to retrospectively assess treatment efficacy, the model is more powerful when used to predict future trends and coverage amounts so that practical policy decisions can be implemented. This aspect is the most important. By knowing how successful an intervention will be in advance and when there will be an increase in demand, the national and local health authorities, with the support of aid organizations, can select the policies that maximize the number of children treated. Considering this, Figure 5 shows the predictions made for 2019. These novel, post-validation results show that the weeks with the lowest percent coverage coincided with the middle of the year during the rainy season. For 2019, it is expected that the average country-wide weekly percent coverage will be below 2% for about 8 weeks, with a peak in coverage of about 5.1-5.2% of the population of children under five years old with diarrheal diseases receiving treatment near the beginning and end of the year. As shown in Figure 5, this decrease comes from the fact that despite the decrease in the overall sick population, the decrease in the number treatment occurs more quickly, resulting in the observed change in coverage. Thus, future efforts would have the largest room for improvement during this time as the number of children that need treatment remains high and the least number of children are being treated.

Discussion

As previously noted, the model is not without its own limitations. The assumptions that everyone who makes it to a facility receiving treatment while medical supplies last is one aspect that may cause the model to slightly diverge from the realities on the ground. Due to the ever-changing conditions on the ground, acquiring data in a timely manner that meets the necessary criteria to be integrated into the model is extremely difficult. While this difficulty is one of the initial motivators of the model's creation, it also serves as another key limitation in the model's predictive power. By only looking at the data where follow ups occurred to paint the most complete picture, as well as the other assumptions made throughout, it is unsurprising that the model had discrepancies. These are most noticeable in the comparison between the oldest data set, 2013 YNHDS document which stated that in the two weeks prior to the survey it was discovered that 33% of the sick population went for advice or treatment from a health facility or provider and of that group 60% were treated with Zn ORS for a total of about 20% of this sick population receiving appropriate treatment over a two week interval.[10] The model however underestimates this number and shows that between 3.61% to 10.65% of the sick population received the appropriate treatment of Zn/ORS over a two-week period. Despite this discrepancy, and the fact that this

survey depicted conditions before the conflict began, the model still can provide useful and usable information.

The extent of the conflict's effect can be seen in a comparison between the same week across different years. As the weather is assumed to follow the same pattern between years, the only between the same week during two different years is the state of the conflict. As fighting continues and the infrastructure is further damaged or otherwise inaccessible, the values of the road maximum, minimum, and bridge values are decreased. This results in a decrease in probability that a sick child makes it to a primary care facility and thus ultimately transitions from the *Sick* state to the *In-treatment* state, as seen in the week 28 value for 2018 of 2.49% in Figure 4 decreasing by about 31% over the following year to 1.73% as shown in Figure 5. In these cases, the extent of treatment coverage and the value of the transition function, YT(t), serves as an indirect measurement of the conflict's worsening effect on the health of children in Yemen.

The most noticeable trends are the seasonal nature of those seeking treatment and the decaying exponential behavior of the overall number sick. The weekly number of children treated closely follows the sinusoidal nature of the roads and weather which makes sense as these were the main factors influencing this transition probability though it is not exact as the population continues to decrease due to a combination of the absorbing nature of the *Deceased* state coupled with the shrinking *Sick* population that directly feeds this state. While the overall sick population decreases in a manner similar to an exponential decay curve, it begins to abruptly slow the rate of decrease around the 20th-25th week, around the same time that the *In-treatment* population begins to increase. The transition point where the slope of the *Sick* graph is lowest indicates that the number of weekly diarrheal cases has increased to nearly match the number of children leaving the *Sick* state during the time that the average rainfall historically is at and near its peak. Despite this increase in the number of new diarrheal cases, the number of treated does not increase proportionally, leading to a period of decreased coverage. Beyond this point, the model depicts movement in opposite directions for the number being treated and the burden of disease. This anti-parallel movement appears to make intuitive sense; if more people are treated there should be fewer sick children as they will leave the sick state to transition into the treated state. This trend can also be explained by the fact that diarrhea is often water-borne and thus seasonal with rains, while the ability to access treatment would decline as the road networks deteriorate at the same time. Conversely, during dry seasons, fewer diarrheal disease cases would be expected because transmission factors are reduced, and those that are sick would be better able to access treatment as the road network is more passable. Although there is a greater numerical change in the number of children residing in the *Sick* state compared to the *In-treatment* state, the percent change in the *In-treatment* state seems to be more drastic. This could indicate that although more children may be helped by slowing the infection rate, improving access to care via minimizing the effects of the weather and travel restrictions could lead to a greater percent change for the number receiving treatment.

It is also worth noting that despite the years of war, the overall disease burden represented by the number of children in the Sick state seems to have decreased. Although the YNHDS value of 31% was used in both simulations as the initial prevalence of the disease, the number sick rapidly dropped at the start both

times, seeming to indicate that the actual value may be closer to about half as much, a testament to the work done by all health actors in the region that have improved conditions here such as the tripling in number of people reached by water, sanitation and hygiene (WASH) efforts between the months of May and August.[25] As the model predicts, this will be the period when there is an increase in the amount of contaminated surface water due to the rain so these preventative efforts are especially important during this time and this trend should continue. This serves to decrease the burden of disease and the ultimate population in need as shown in the model. While critical for this work to continue, especially in order to achieve a continued reduction in the prevalence of these diseases, these burden decreasing approaches do not directly address those that are already sick and in need of treatment which is an important aspect of the percent coverage. As this population is often smaller than the population in need, an increase in the number receiving treatment will have a larger effect on the percent coverage compared to an equal size change of those in need.

Based on this set of observations, it seems that the best way to increase coverage would be to increase treatment efforts during the peak periods of rain around the middle of the year. This might best be accomplished by focusing on community outreach efforts during and just prior to this time since the disease burden is expected to increase while at the same time a smaller proportion of children are reaching facilities for treatment compared to other parts of the year. This could allow for an increase in medication to be delivered in a pre-emptive fashion to outreach programs, helping to keep the number and severity of cases low as people would not have to endure the poor roads and weather to reach medication. Having a defined period where outreach is increased serves to be more cost effective compared to increasing these efforts during the same period in the dry season when they may have less of an impact or year-round where the operating cost would be increased. Ultimately, a dual approach that leads to an increase in outreach and community-based workers especially during the periods of decreased travel abilities and an increase in fixed facilities when travel conditions improve, should be utilized. Given the full year prediction for 2019 produces similar trends, this also seems to support the fact that an additional treatment approach should be utilized to improve the treatment coverage of Zn/ORS for diarrheal diseases in children under five. Without this change, potentially up to 98% of children will not be able to receive the treatment that they so desperately require.

To reach these conclusions, a variety of different metrics were examined. These factors were the number of children treated, the incidence rate, the efficacy of zinc and oral rehydration salt solution as a means of treatment, the mortality rate of the treatment and the overall mortality rate of the under five-year-old age group in the country. The number of children treated was calculated as the number of people that transitioned from the *Sick* state to the *In-treatment* state. This was the primary output the model tried to best match, as the third-party health facility data directly measured this value and other parameters were partially related to this one. Incidence rate was used as an indirect estimation for the infection rate. As published by El Bcheraoui et al., the incidence rate is a measurement of the number of new episodes of diarrhea over the total number of person-years examined.[11] This provides then the average number of times a child would become sick with a diarrheal disease over the course of a year. The treatment efficacy

was calculated as the number that became healthy again after treatment divided by the total number being treated while treatment mortality was measured as the number that died the following week after treatment divided by the total number being treated that week respectively. Lastly, the overall mortality rate was the total number that had died over the course of the simulation divided by the original population size.

When the values found in both the data provided and the literature for these metrics were compared against the model's output, there was no statistically significant difference found. Although care was taken to examine information from different sources to limit errors, the primary data metrics comes from one source and is not without limitations, which as mentioned early could be a source of discrepancy. Along with the previously mentioned fact that 47% of the clinical registry data was deemed incomplete and not used due to mismatching total values, the number of primary data sources was limited due to the conflict itself. Furthermore, each metric was limited in the information it could provide. For example, the incidence rate and overall mortality rate published is for 2016. Changes in the incidence rate between 2016 and 2019 are likely to have been influenced by a variety of factors. Continuation and increased intensity of the conflict over the years may have led to an increase in both values, although increased efforts by health sector partners including UNICEF to expand functionality and accessibility of health services could have improved treatment rates. Due to the scarcity of reporting from the region due to the conflict, more recent published data has not been found.[11]

Monitoring the effects of Zn/ORS on the course of the disease is also difficult to assess, calling into question the exact efficacy of the treatment. Although it is well established that this is the best course of treatment for those that suffer from diarrheal diseases, its exact efficacy in the region is unknown. Given that this treatment for the majority of diarrheal cases does not require patients to remain in a health facility under the care and observation of a health worker, it is unknown the extent to which children adhere to the global childhood diarrhea treatment guideline of 10 to 14 days of 20 mg doses of zinc for children between the ages of 6 and 59 months or 10 mg doses for the same duration of time for children under 6 months old.[27] It is also not recorded if these children received some other type of treatment from an alternative source during this time. Both factors could affect treatment recovery and mortality rates in the data.

Although this type of model has been used for epidemiology predictions in the past and the model presented does produce similar values to those seen in the literature, it too is not without limitations. By examining Yemen as a whole, population movements are neglected from the analysis. Migrations within the country could lead to regional effects as a sick child becoming infected and receiving treatment in different regions would lead to a decrease in the percent coverage in the area where the child was infected and an increase in the percent coverage where the patient was treated. Internally displaced person (IDP) status is also associated with poor living conditions including poorer hygiene and sanitation and limited access to health services, potentially increasing both the incidence of under-five diarrheal disease and the risk of mortality from this condition in areas where large populations of IDPs are living.

Given that the model was designed to be easily tunable and applicable to other areas of conflict with suboptimal data quality and quantity, future iterations could look at expanding the model into these regions or other diseases / conditions in Yemen. In the current version of the model, the only information required is time series data on a region's transportation infrastructure, reports on the number of people in a subpopulation treated for the disease (or types of disease) and the outcome of their treatment. With this information, the model can be combined with publicly available information (such as population size and mortality) and provide predictions for a disease or age group in question. Expanding the model into other regions, especially on a subnational level or other diseases could even serve as a further means of validation, thus improving its functionality not only in each new region, but in all areas where it is being used.

Conclusion

To assist health care workers and policy makers alike, a computational Markov Model was created for Yemen that monitored the spread of diarrheal diseases in children under five years of age. The model was able to recreate various trends recorded in the clinical register health facility data collected by a third party on the number of children treated for diarrheal diseases, as well as for the treatment mortality and efficacy rates, and for data previously published in the literature on the incidence rate and the overall mortality rate of this population. With this viewed as sufficient validation for the model, a predictive treatment coverage map was created for the region analyzed on a weekly basis over the course of a year.

Based on this predictive chart for 2019, it was determined that the percent of children receiving treatment varied between 1.73% during the worst of the mid-year rains to 5.2% during the last week of the year. As the number treated is always lower than the number needing treatment, increasing the population in this state directly will lead to the most drastic increase in coverage percentage, while also serving to indirectly decrease the total sick population as well. Thus, to attempt to increase this value and prevent a decrease in future years, our model recommends that the number of community outreach programs should be increased just before and during these weeks of increasing disease burden and declining proportion of the population in need that seeks treatment.

Declarations

Ethical Approval and Consent to Participate

All human based data was deidentified and shared by UNICEF following ethical guidelines.

Consent for Publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from UNICEF but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of UNICEF.

Code used to reach these conclusions is available upon request.

Competing Interests

The authors state that the institutions of Boston University (BU) and the United Nations International Children's Fund (UNICEF) have entered into a financial partnership during the course of this publication.

Funding

The completion of this study was made possible with funding provided by UNICEF.

Authors' Contributions

MPS was a major contributor to the analysis and interpretation of the data, created and coded the model, and was a major contributor in the writing of the manuscript. AG assisted in the analysis of data.

MKD compiled and provided data as well as provided guidance on the creation of the manuscript.

FS and MHZ provided major guidance during the study.

MPS, MKD, FS, and MHZ all provided edits to the manuscript.

All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Authors' Information

MKD and FS have both worked on the ground in Yemen and are directly familiar with the current realities the country and its children are currently facing.

References

1. Health system in Yemen close to collapse. (2015). Retrieved from https://www.who.int/bulletin/volumes/93/10/15-021015.pdf
2. UN Humanitarian. (2019). Eleven facts about the Yemen crisis. Retrieved from https://unocha.exposure.co/eleven-facts-about-the-yemen-crisis

3. Camacho, A., Bouhenia, M., Alyusfi, R., Alkohlani, A., Naji, M. A. M., de Radigues, X., . . . Luquero, F. J. (2018). Cholera epidemic in Yemen, 2016-18: An analysis of surveillance data.*The Lancet. Global Health, 6*(6), e680-e690. doi: S2214-109X(18)30230-4 [pii]
4. UN News. Yemen. Retrieved from https://news.un.org/en/focus/yemen
5. Madhok, R., & AlMarani, S. (2017). Falling through the cracks: Yemen's forgotten children in a cholera crisis. Retrieved from https://www.unicef.org/stories/falling-through-cracks-yemens-forgotten-children-cholera-crisis
6. Bajait, C., & Thawani, V. (2011). Role of zinc in pediatric diarrhea. *Indian journal of pharmacology*, *43*(3), 232–235. doi:10.4103/0253-7613.81495
7. Walker, C. L., & Black, R. E. (2010). Zinc for the treatment of diarrhea: Effect on diarrhea morbidity, mortality and incidence of future episodes.*International Journal of Epidemiology, 39 Suppl 1*, i63-9. doi:10.1093/ije/dyq023 [doi]
8. Lam, F., Wentworth, L., Cherutich, P., Migiro, S., Abdala, K., Musyoka, M., … Schroder, K. (2019). An evaluation of a national oral rehydration solution and zinc scale-up program in Kenya between 2011 and 2016. *Journal of global health*, *9*(1), 010505. doi:10.7189/jogh.09.010505
9. World Health Assembly, 64. (2011). Cholera: mechanism for control and prevention: report by the Secretariat. World Health Organization. https://apps.who.int/iris/handle/10665/3361
10. Ministry of Public Health and Population - MOPHP/Yemen, Central Statistical Organization - CSO/Yemen, Pan Arab Program for Family Health - PAPFAM, and ICF International. 2015. Yemen National Health and Demographic Survey 2013. Rockville, Maryland, USA: MOPHP, CSO, PAPFAM, and ICF International.
11. El Bcheraoui, C., Jumaan, A. O., Collison, M. L., Daoud, F., & Mokdad, A. H. (2018). Health in Yemen: losing ground in war time. Globalization and health, 14(1), 42. doi:10.1186/s12992-018-0354-9
12. World Health Organization. (2017). Diarrheal Disease. Retrieved from https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease
13. Amood, A. K., Al-Areqi, L., Mujally, A., Alkarshy, F., Nasser, A., & Jumaan, A. O. (2016). Diarrheal Diseases Hospitalization in Yemen before and after Rotavirus Vaccination. *Scientifica*, *2016*.
14. Getachew, A., Tadie, A., G.Hiwot, M. *et al.*Environmental factors of diarrhea prevalence among under five children in rural area of North Gondar zone, Ethiopia. *Ital J Pediatr***44,** 95 (2018) doi:10.1186/s13052-018-0540-7
15. N Barbolosi, D., Crona, J., Serre, R., Pacak, K., & Taieb, D. (2018). Mathematical modeling of disease dynamics in SDHB- and SDHD-related paraganglioma: Further step in understanding hereditary tumor differences and future therapeutic strategies. *PloS one*, *13*(8), e0201303. doi:10.1371/journal.pone.0201303
16. World Health Organization. (n.d.). Global Health Observatory data repository Life tables by country - Yemen. Retrieved September 9, 2019, from http://apps.who.int/gho/data/view.main.61840?lang=en.
17. Mokomane, M., Kasvosve, I., de Melo, E., Pernogistica, J. M., & Goldfarb, D. M. (2018). The global problem of childhood diarrheal diseases: emerging strategies in prevention and

management. *Therapeutic advances in infectious disease*, *5*(1), 29–43. doi:10.1177/2049936117744429

18. Centers for Disease Control and Prevention. Epidemiology and Prevention of Vaccine-Preventable Diseases. Hamborsky J, Kroger A, Wolfe S, eds. 13th ed. Washington D.C. Public Health Foundation, 2015.
19. The World Bank. (n.d.). World Bank Climate Change Knowledge Portal. Retrieved December 3, 2019, from https://climateknowledgeportal.worldbank.org/country/yemen.
20. Yemen Data Project. (2019). Airwar. Retrieved from https://yemendataproject.org/data.html
21. Engelke, Wilfried. 2017. *Yemen transport sector (English)*. Yemen policy note. Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/636961508411397037/Yemen-transport-sector
22. Al-Shikh, H. N. S., & Raju, V. (2017). Logistics in Yemen: Optimal intermodal freight transportation.*International Journal of Humanities and Cultural Studies, 4*(2), 15.
23. USAID. (2016). Yemen fact sheet #9. Retrieved from https://www.usaid.gov/sites/default/files/documents/1866/yemen_ce_fs09_04-28-2016.pdf
24. Logistics Cluster. In World Food Program (Ed.), *Yemen access constraints map*
25. OCHA. Cholera response dashboard. Retrieved from https://www.humanitarianresponse.info/en/operations/yemen/cholera-response-dashboard
26. Yemen Central Statistical Organization. (n.d.). 2018 Population Estimates.
27. Ambler, G., Berry, J., Berry, S., Castle, B., Clarence, C., Clark, A., … Schenker, I. (2018, November 12). Proposal to include an additional listing of copackaged ORS and zinc for management of diarrhea in children on the WHO Model List of Essential Medicines for Children. Retrieved December 3, 2019, from https://www.who.int/selection_medicines/committees/expert/22/s17.5_ORS-zinc.pdf.

Figures

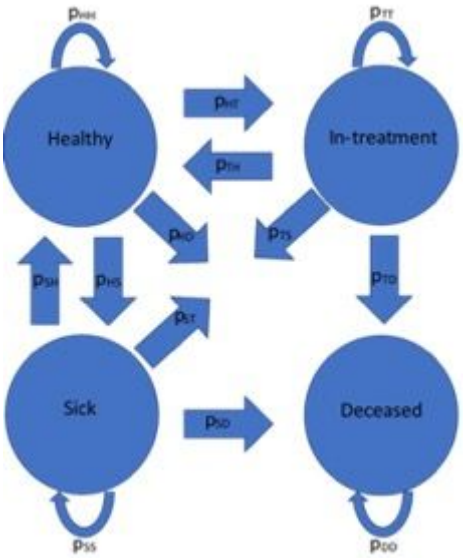


Figure 1

Model schematic. Overview of the possible interactions between the four different states in the model.

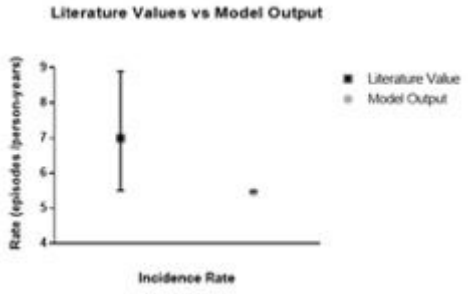


Figure 2

Incidence rate comparison. The comparison between data published by El Bcheraoui et al. and the average output of the model after 7 iterations for 2016. Error bars represent the upper and lower bounds of the 95% confidence interval analyzed in GraphPad. Although numerically close, the model's output for the incidence rate is deemed to be statically different as it falls outside the confidence interval of the target estimation data.

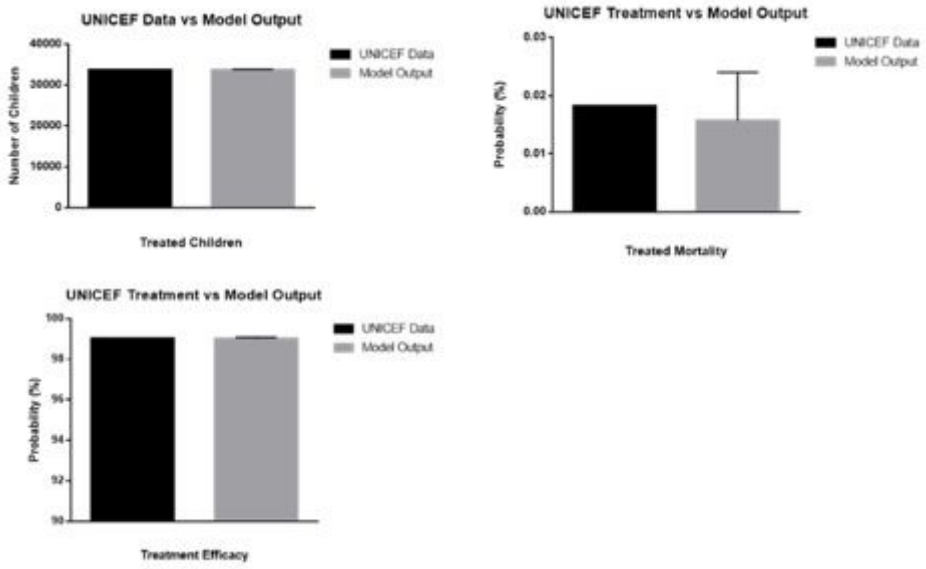
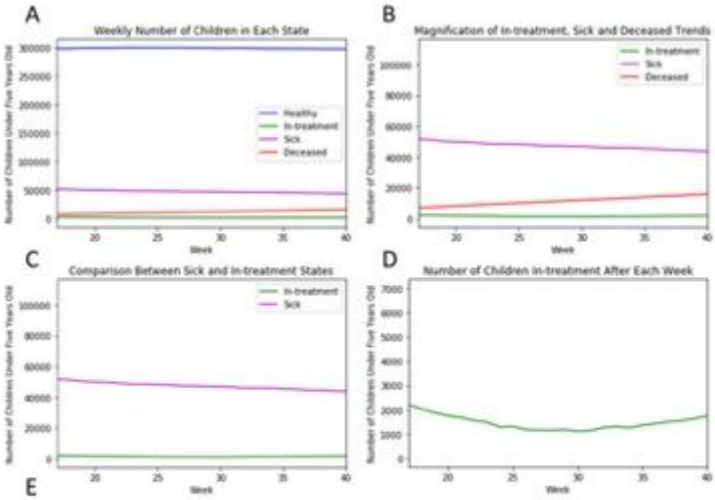


Figure 3

Comparison to Number of Children Treated (A), Treatment Mortality (B) and Treatment Efficacy (C). Comparison of the model against validation data showing no significant difference between the A) total number of children under 5 treated (P=.8864) and B) the treatment mortality rates from 2018 health facility monitoring data collected through a third party (P=.4479). C) The efficacy of Zn/ORS intervention from the same data set (P=.8450) was also not deemed not significant. Significance was assessed with a one sample T test in GraphPad. Error bars shown as standard deviation for n=7 code iterations.

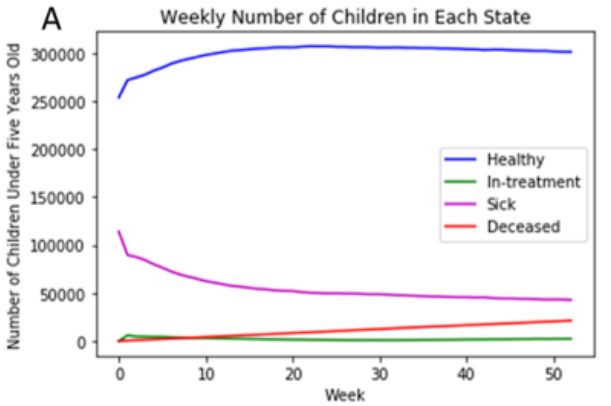


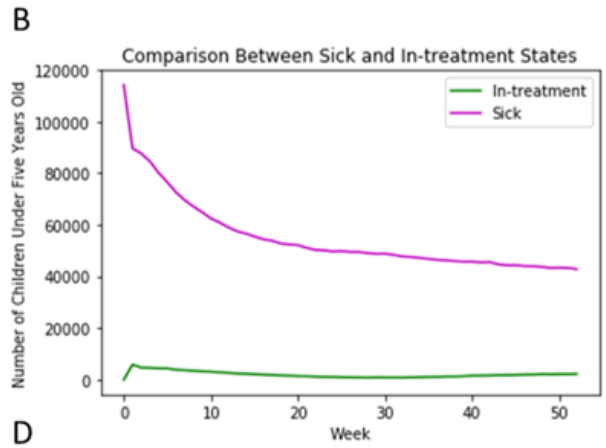
E

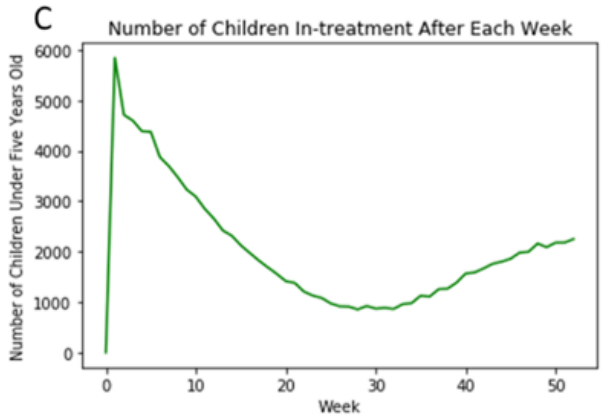
Week 18-28	3.67	3.49	3.37	3.14	3.04	2.66	2.70	2.45	2.42	2.44	2.49
Week 29-39	2.37	2.43	2.72	2.85	2.73	2.96	3.15	3.33	3.48	3.67	3.95

Figure 4

Health trends for children under five in Yemen during May - September of 2018. (A-E) The week number is the number of weeks since the start of the examination period began on May 1st. A) Overall weekly distribution of the health states for the estimated child population and (B) the zoomed in to see the bottom half of the graph. C) Plot of the total Sick state population vs time compared against the In-treatment state vs time, with the In-Treatment state enlarged in (D). E) The percent coverage was listed week by week.







D

Week 1-9	5.12	5.26	5.25	5.19	5.46	5.07	5.10	5.00	4.82
Week 10-18	4.76	4.56	4.36	4.10	4.03	3.76	3.56	3.35	3.14
Week 19-27	2.96	2.70	2.65	2.37	2.24	2.16	1.96	1.84	1.84
Week 28-36	1.73	1.88	1.78	1.81	1.79	2.01	2.06	2.39	2.37
Week 37-45	2.71	2.74	3.02	3.43	3.47	3.67	3.86	4.03	4.20
Week 46-52	4.47	4.54	4.91	4.78	5.03	5.02	5.20		

Figure 5

Predictions for each state (A), no Healthy (B), or Sick states (C) and weekly percent coverage (D). A) Tracked changes over all states B) Comparison between Sick and In-treatment states throughout 2019 and C) a magnified view of the Sick state D) The weekly percent treatment coverage during the year.