A Inferential Analysis of Waterborne Diseases in India: A Data Science Perspective

Shweta P
Department of Mathematics(SAS)
Vellore Insititue of Technology
Vellore,India
shweta.p2023@vitstudent.ac.in

Greeshma C
Department of Mathematics(SAS)
Vellore Insititue of Technology
Vellore,India
greeshma.c2023@vitstudent.ac.in

Pradeep kumar B
Department of Mathematics(SAS)
Vellore Insititue of Technology
Vellore,India
pradeepkumar.b2023@vitstudent.ac.in

***Abstract*— Over the years, India has witnessed an alarming rate of waterborne diseases mainly diarrhea, typhoid and cholera and has a huge disease burden due to the large number of cases and deaths especially among children. In India, diarrhea is one of the leading causes of under-5 deaths. Cholera cases are also on the rise however the occurrence is seasonal. It continues to remain under-reported and under-recognised in India. Typhoid is also a major public health concern with a high incidence rate in urban areas. Despite having several policies and schemes to address this issue, India is still facing an alarming incidence rate of these diseases.**

This research paper aims to identify the trends and patterns in the number of cases reported and mortality rates of the three major waterborne diseases in India, namely, diarrhea, typhoid and cholera. We use data from the National Health Profile published by the Central Bureau of Health Intelligence, India and employs descriptive statistics measures such as mean, rank correlation and proportion ranking for the characterization of each disease individually as different diseases have different attributes. We also perform an inferential analysis to identify the hotspots using scan statistics(using SatScan) for facilitating resource optimization.

Keywords*— Waterborne diseases, Diarrhea, Typhoid, Cholera, trends and patterns, hotspots

I. INTRODUCTION

The goal 6 of United Nations Sustainable Development Goals (SDGs) focuses on 'Clean Water and sanitation for all' by 2030[1]. However, according to UNICEF, on the global front, 2.2 billion people do not have access to safe drinking water, 3.6 billion people which is nearly half of the world do not have access to safe sanitation and 494 million people practise open defecation which are the major contributors of waterborne diseases (WBD) [2]. According to the United Nations, water borne diseases affect 37.7 million Indians annually and is estimated to cause an economic burden of 600 million USD. Poor sanitation and hygiene practices along with consumption of unsafe and contaminated water are the major contributors to the spread of waterborne diseases.

According to a report by the World Bank, about 21% of communicable diseases in India are water-related[4]. In spite of having several schemes and policies to address the inadequate access to safe drinking water and sanitation, the incidence rate of waterborne diseases are concerning. Hence it is important to study the trends and patterns in waterborne diseases as well as identify the major hotspots in India for early diagnosis, timely management and effective policy implementation.

.

II. LITERATURE REVIEW

The study looked at slum population, living circumstances, cleanliness, income, and sanitation to see how these affected illness and death rates. A survey of 1070 families in Mumbai found that inadequate sanitary conditions were responsible for at least 30% of sickness. The study discovered a link between sea surface temperature and Cholera outbreaks, with yearly incidences of Diarrhea, Typhoid, and Malaria estimated to be 614, 68, and 126 per thousand people, respectively. Water-related illnesses accounting for 30% of total morbidity, with selected water-borne disease incidence rates in Malad and Rajiv Gandhi Nagar.[4]

Soorya et al., 2021 analysed the trends and patterns in waterborne diseases in Kerala and its districts along with the seasonality of waterborne diseases from 2011-2019 using data from Directorate of Health Services Kerala portal. The study examines waterborne diseases like Leptospirosis, Hepatitis A, Cholera, Typhoid and Diarrhoea and concluded that more than 97% of waterborne diseases are due to diarrhoea with no deaths reported. Leptospirosis constitutes only 0.2 to 0.4 % of the total waterborne diseases however its death rate is very high. The study also discusses the seasonality of waterborne diseases which was studied using ANOVA test and concluded that there is a surge in waterborne diseases during the rainy season. The researchers also concluded that diarrhoea is the main reason for morbidity due to waterborne diseases.[5]

Using The research looks at the factors that influence drinking water in the Uttar Pradesh districts of Lucknow and Kanpur, concentrating on income, family size, education, occupation, and caste. The 2011 Census presented a varied picture of drinking water sources, with piped tap water dominating in Lucknow and handpumps/bore wells dominating in Kanpur. Both governmental and private sources developed disproportionately, with the most diversified access coming from SC/ST populations and low-income households. Education also has a favorable impact on improved sanitary facilities, with better access to upgraded

facilities available to better-educated families. This underscores the importance of improved sanitary services for all homes.[6]

III. Methodology

A. Data Collection

The National Health Profile was the source of the data. The data includes the number of cases and fatalities related to water-borne illnesses such cholera, typhoid, and acute diarrheal illness. The data used in this study was collected between 2013 and 2021.

B. Data Preprocessing

The data collected from the National health profile was in the form of pdf format. To make the data usable for the analysis, preprocessing step was undertaken. This step involved in the extraction of the information from the pdf and converting it into an excel file, ensuring that data was in a structured format. Handling outliers and null values was done with great care during this step. This step was useful to enhance the accuracy and dependablity of the dataset.

C. Descriptive Statistics

- This step is to understand the Central tendancy and dispersion of the data by calculating the Mean and Standard deviation. For each diseases we calculate the average number of cases and deaths from 2013 to 2021.Each disease is calculated separately, making it possible to identify differences and variation in the number of cases and deaths from a given condition year over year. The Standard deviation is calculated for each disease, providing information on the degree of dispersion.

D. Exploratory data analysis

- In this step, with the help of Descriptive statistics, We generated various plots which includes bar plot and histogram, aiming to gain initial understanding of the dataset. The graphical depictions of the overall number of cases and deaths resulting from each disease for specific years have been plotted. The graphs have been plotted for all the three diseases-typhoid, acute diarrheal and cholera-across all years.

E. Inferential analysis

- The main objective of Inferential analysis is to generalize results from a small sample to the larger population by providing a framework for conclusions and judgements that go beyond the observed facts. This involves making predictions, about events or making supported decisions based on statistical information.
- Spearman's rank correlation is used to examine the rank correlation between the number of cases and mortality for three diseases: acute diarrhea, cholera, and typhoid. This statistical technique is used to evaluate the degree and direction of relationships, between the ranked variables.
- Moreover, the highest ranking states in India for cases and deaths from each of the three diseases are determined by applying Spearman's rank correlation. This analytical method contributes to a more comprehensive understanding of India's health situation by identifying the areas with the greatest impact on illness incidence and mortality.

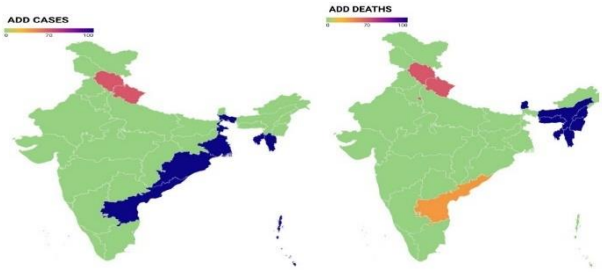
F. Spatial analysis

SatScan is a software tool used for spatial and spatiotemporal cluster detection in epidemiological studies. The utilisation of this software facilitates the identification of hotspots by allowing researchers to identify regions with markedly elevated illness incidence. The software makes it possible to identify geographic clusters where disease rates differ significantly from expectations, providing important information about areas of increased health concern.

IV. Results and Conclusion

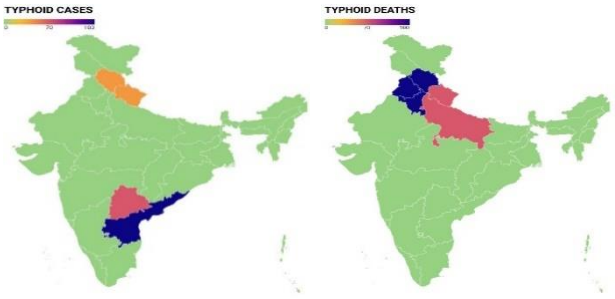
A. Descriptive Measures

- The mean diarrheal cases in India from 2013-2021 is 2983579 with a standard deviation of 3149158. The skewness 1.76 indicates that the distribution of diarrheal cases is positively skewed, i.e, the cases in majority of the states lie above the mean value.
- The mean diarrheal deaths is 382.77 with a standard deviation of 518.54 and skewness 1.92(the deaths in the majority of states lie above the mean value).
- The mean typhoid cases in India is 493858 with a standard deviation of 654543. The skewness 2.7 indicates that the distribution of typhoid cases is positively skewed.
- The mean typhoid deaths is 92.15 with a standard deviation of 242.29 and skewness 5.27 indicating that the distribution is positively skewed.
- The mean cholera cases in India is 187.59 with a standard deviation of 349.87. The skewness 2.52 indicates that the distribution of cholera cases is positively skewed.
- The mean cholera deaths is 1.05 with a standard deviation of 1.89 and skewness 2.23 indicating that the distribution is positively skewed.



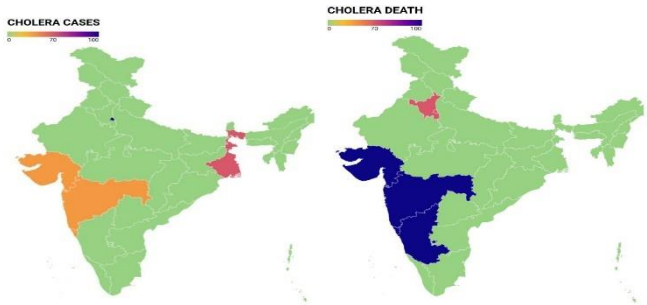
A B

Fig.1. A -Shows the hotspots of Diarrhea Cases
B- Shows the hotspots of Diarrhea deaths



A B

Fig.2. A-Shows the hotspot of Typhoid Cases
B-Shows the hotspot of Typhoid Deaths

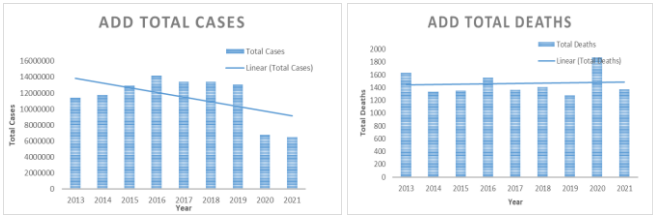


A B

Fig.3. A-Shows the hotspot of Cholera Cases
B-Shows the hotspot of Cholera Deaths

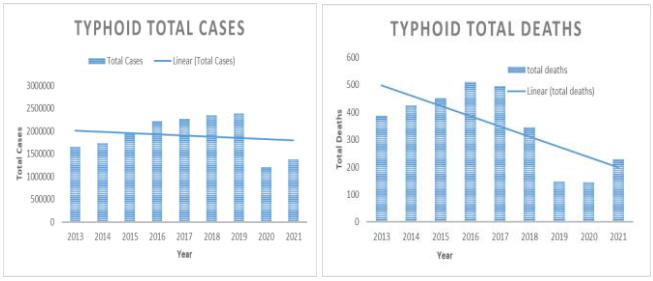
By using Spearman's rank correlation, we found the hotspots for deaths and cases of diarrhea, cholera and typhoid as follows:
The hotspots for:
i) Diarrheal cases: West Bengal, Andhra Pradesh, Uttar Pradesh, Odisha, Rajasthan
ii) Diarrheal deaths: Uttar Pradesh, Assam, Andhra Pradesh, West Bengal, Odisha
iii) Cholera cases: West Bengal, Gujarat, Maharashtra, Delhi, Tamil Nadu
iv)Cholera deaths: Maharashtra, Gujarat, Karnataka, Haryana
v) Typhoid cases: Uttar Pradesh, Bihar, Telangana, Andhra Pradesh, West Bengal
vi) Typhoid deaths: Uttar Pradesh, Odisha, Assam, Delhi, West Bengal

The results from proportion ranking of the states(deaths/total cases) are as follows:
a) Diarrhea: Uttar Pradesh, Assam, Andhra Pradesh, West Bengal, Odisha
b) Cholera: Maharashtra, Gujarat, Karnataka, Haryana
c) Typhoid: Uttar Pradesh, Odisha, Assam, Delhi, West Bengal.



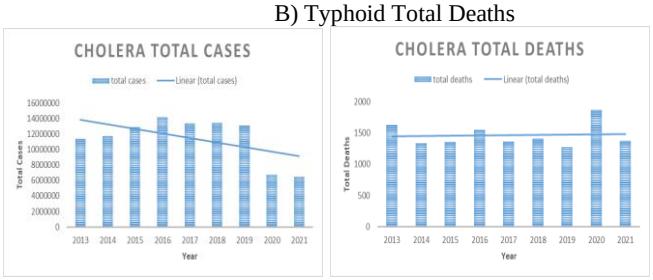
A B

Fig.4.A) Diarrhea Total Cases
B) Diarrhea Total Deaths



A B

Fig.5.A) Typhoid Total Cases
B) Typhoid Total Deaths



A B

Fig.6.A)Cholera Total Cases
B)Cholera Total Deaths

B. *Trend Analysis*

From Fig.4, we can infer that the trend line for total cases of diarrhea is downward sloping ,i.e, a decreasing trend. However the trend line for total diarrheal deaths is upward sloping,i.e, an increasing trend has been observed. Similarly, from Fig.5, it is evident that there has been a slight decrease in the total typhoid cases and a significant decrease in typhoid deaths. Fig.6 indicates a significant decrease in total cholera cases and a slight increase in the total deaths due to cholera.

V. Acknowledgment

We would like to express our heartfelt gratitude to Prof. Dr. Jitendra Kumar, Department of Mathematics, for his invaluable guidance and support throughout our research , which has been instrumental in its successful completion.

VI. Reference

[1] *Goal 6 | Department of Economic and Social Affairs*. (n.d.).
[2] *10 things you didn't know about water*. (n.d.). UNICEF.
[3] World Bank. (2017, April). *Waterlife: Improving Access to Safe Drinking Water in India*. Retrieved November 5, 2023
[4] Ansari, Firdous. (2017). International Journal of Multidisciplinary Research and Development Scenario of some waterborne diseases in slums of Mumbai city. 444-446.
[5] Soorya, V., Raj, J., & R, A. K. K. (2021, July 28). *Trends in water borne diseases in Kerala: an analysis of directorate of health services portal data*. International Journal of Research in Medical Sciences; Medip Academy
[6] Shukla, Rashmi & Nayak, Sanatan. (2017). Drinking Water, Sanitation and Water-borne Diseases. Economic and political weekly. 50. 136-140.