

Developing a Statistical Early-Warning System for Cholera using Socio-Economic indicators

Howard Lei, Farnaz Ganjeizadeh, Shervin Baharmand, and Andishe Almasi

Introduction

- Early warning systems for disease outbreaks help predict medication demand, with business and supply-chain management implications
- Use of socio-economic indicators in a statistical framework a low-cost, minimally invasive approach
- Up to 2-year ahead predictions made for 10 African countries across 8-years (2004-2011)

Countries examined

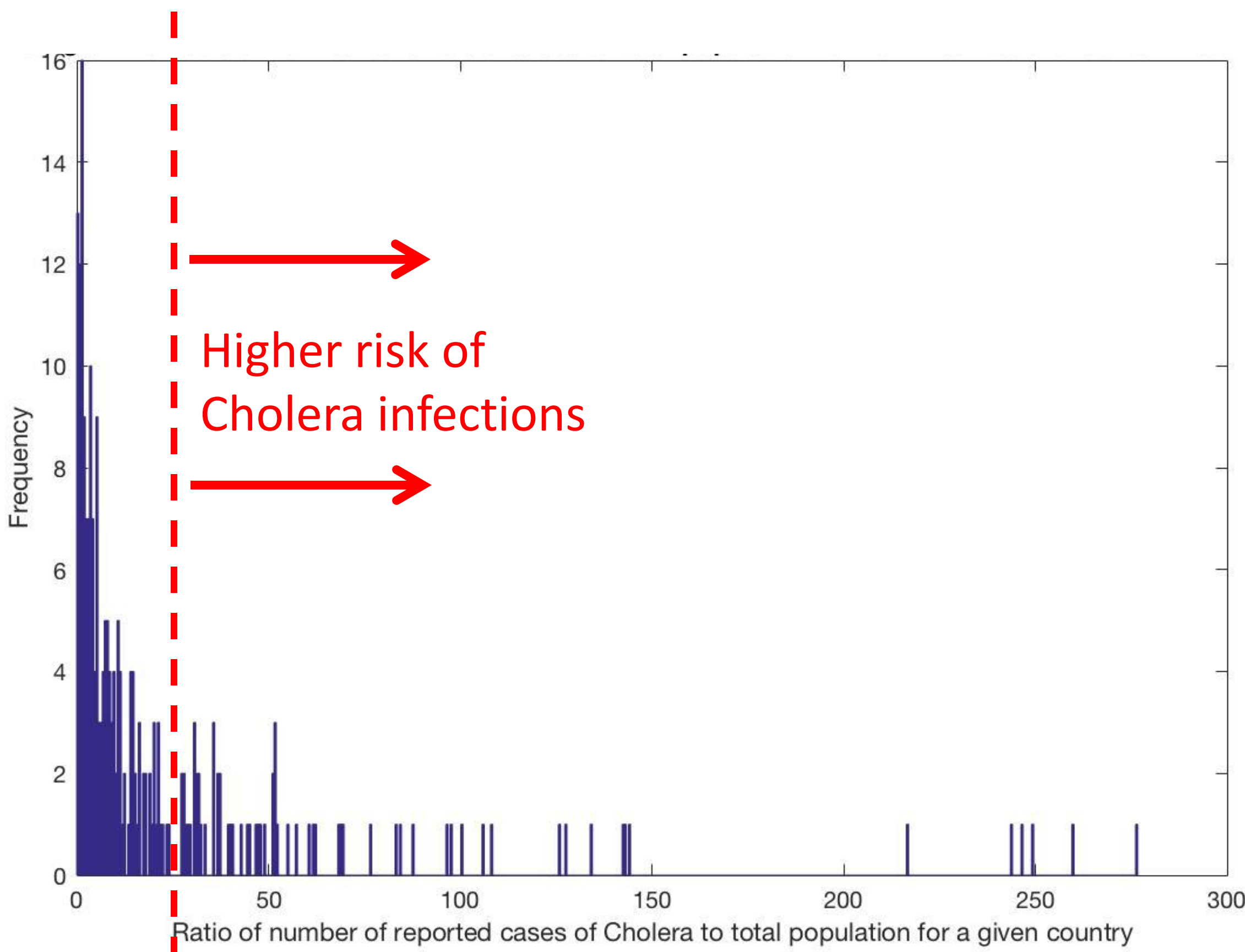
- Benin
 - Burundi
 - Cameroon
 - Democratic Republic of Congo
 - Ghana
- Malawi
 - Mozambique
 - Nigeria
 - Tanzania
 - Togo

Socio-economic indicators examined

- Adolescent fertility rate
 - Agriculture, value added (% of GDP)
 - CO2 emissions (metric tons per capita)
 - Exports of goods and services (% of GDP)
 - External debt stocks
 - Fertility rate (births per woman)
 - Foreign direct investment
 - Forest area (sq. km)
 - GDP at market prices (current US\$)
 - GDP growth (annual %)
 - GNI per capita, Atlas method
 - Gross National Income (GNI), Atlas method
 - Immunization (measles)
 - Imports of goods and services (% of GDP)
 - Improved sanitation facilities (% of population with access)
 - Improved water source (% of population with access)
 - Industry, value added (% of GDP)
 - Inflation, GDP deflator (annual %)
 - Life expectancy at birth
 - Merchandise trade (% of GDP)
 - Mobile cellular subscriptions (per 100 people)
 - Mortality rate (under age 5)
 - Net barter terms of trade index
 - Net migration
 - Net ODA received per capita
 - Population growth (annual %)
 - Prevalence of HIV, total (% of population ages 15-49)
 - Services, etc., value added (% of GDP)
 - Surface area (sq. km)
 - Total debt service (% of GNI)
- Green indicators have high-correlation w/ Cholera infections-to-population ratio**

Blue indicators have best classification performance

Cholera infections-to-total population (CTP) ratio histogram



GOAL: Detect countries/years with higher risk of Cholera infections (i.e. higher CTP-ratio)

Predicting the CTP-ratio 1 year later

(Training data: yrs 1990-2003 Test data: yrs 2004-2012)

Classifier	Socio-economic indicators	False Alarm	Miss	Avg False Alarm and Miss
Logistic Regression	10 highest correlation w/ CTP ratio	0.43	0.11	0.27
Logistic Regression	Improved water source CO2 emissions	0.18	0.37	0.28
Linear Regression	10 highest correlation w/ CTP ratio	0.37	0.26	0.32
Linear Regression	Improved water source CO2 emissions	0.18	0.37	0.28
Perceptron	10 highest correlation w/ CTP ratio	0.58	0.11	0.35
Perceptron	Improved water source CO2 emissions	0.52	0.32	0.42
SVM	10 highest correlation w/ CTP ratio	0.36	0.32	0.34
SVM	Improved water source CO2 emissions	0.06	0.42	0.24
Decision Tree	10 highest correlation w/ CTP ratio	0.54	0.26	0.40
Decision Tree	Improved water source CO2 emissions	0.45	0.32	0.39

Predicting the CTP-ratio 2 years later

(Training data: yrs 1990-2003 Test data: yrs 2004-2011)

Classifier	Socio-economic indicators	False Alarm	Miss	Avg False Alarm and Miss
Logistic Regression	10 highest correlation w/ CTP ratio	0.40	0.11	0.26
Logistic Regression	Improved water source CO2 emissions	0.08	0.42	0.25
Linear Regression	10 highest correlation w/ CTP ratio	0.53	0.00	0.27
Linear Regression	Improved water source CO2 emissions	0.08	0.42	0.25
Perceptron	10 highest correlation w/ CTP ratio	0.40	0.21	0.31
Perceptron	Improved water source CO2 emissions	0.20	0.37	0.29
SVM	10 highest correlation w/ CTP ratio	0.35	0.21	0.28
SVM	Improved water source CO2 emissions	0.08	0.42	0.25
Decision Tree	10 highest correlation w/ CTP ratio	0.02	0.89	0.46
Decision Tree	Improved water source CO2 emissions	0.08	0.32	0.20

Summary

- Importance of **2 socio-economic indicators**:

- Improved water source
- CO2 emissions

Note: Bacterium for Cholera transmitted through water

- **Decision tree classifier** with the 2 indicators gives lowest avg. error for 2-year ahead prediction

- Individual classifiers biased towards either false alarm or miss errors

- Minimizing miss error ensures enough medication production
- Minimizing false alarm error results in cost savings, potentially at the risk of not detecting countries/years with higher Cholera infections