Environmental Determinants of Enteric Infectious Disease

PI: BENJAMIN ZAITCHIK, JOHNS HOPKINS UNIVERSITY
CO-I: MARGARET KOSEK, UNIVERSITY OF VIRGINIA
CO-I: HAMADA BADR, JOHNS HOPKINS UNIVERSITY
CO-I: JIM NELSON, BRIGHAM YOUNG UNIVERSITY
POSTDOC: JOSH COLSTON, UNIVERSITY OF VIRGINIA
Mortality

Globally, diarrhea kills 2,195 children every day

This is 1 out of 9 child deaths, worldwide

It is more than AIDS, malaria, and measles combined

It is the second leading cause of death in children less than five years old
Morbidity

Impaired **cognitive development**

**Stunting**

Reduced **vaccine response**

https://borgenproject.org/what-causes-stunting/
EID are preventable and treatable

In some cases, **vaccines** are available

Improved Water, Sanitation and Hygiene (**WASH**) infrastructure and behavior is critical

Those suffering from diarrhea can be treated with **oral rehydration therapy**
Project goal

Establish the feasibility of Earth Observation-informed EID risk mapping, monitoring, and prediction systems

We are doing this through collaboration with multiple EID studies performed at sites around the world
Current list of collaborating studies
Earth Observation data

None of these infection studies included collection of data on climate or environment.

Earth Observations offer an opportunity to fill this gap.
Accomplishments

PY1:
1. Evaluated EO performance at MAL-ED sites, and published results collaboratively with MAL-ED site PIs (Colston et al., 2018)
2. Generated a preliminary rotavirus prediction model based on MAL-ED site data and Earth Observations

PY2:
1. Published the results of the rotavirus model collaboratively with site PIs (Colston et al., 2019)
2. Performed preliminary regionalization based on rotavirus predictors
3. Built template visualization app in Tethys
4. Participated in NASA’s pilot commercial data buy program
Rotavirus transmission pathways

Colston et al. (2019)
Accomplishments

PY3:

1. Performed a targeted study of ENSO influence on EID at our Peru MAL-ED site (Colston et al., 2019)
2. Substantially enhanced our database of predictor variables
3. Nearly completed models of Shigella
4. Produced maps of Shigella risk for dissemination to partners
Impacts of the 2011-2012 La Nina floods on enteric infections in Santa Clara, Peru

Colston et al. (2020)
Shigella modeling results
Correlations between shigella and its predictors

Shigella
Variable Importance Plot

Impurity-corrected Random Forest; Unconditional permutation scheme; cross-validated results
Shigella risk maps
Understanding mechanism
Understanding mechanism
## Timeline and Risks

### Risks

COVID has slowed our research and communication with health system partners.

Slower feedback might limit the number of pathogens we can address by the end of the project.

<table>
<thead>
<tr>
<th>Objective 1</th>
<th>PY1</th>
<th>PY2</th>
<th>PY3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perform and evaluate retrospective LDAS simulations</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td>Create unified database of EID predictors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop and evaluate statistical EID models</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 2</th>
<th>PY1</th>
<th>PY2</th>
<th>PY3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perform EID-specific regionalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characterize regionalization uncertainty</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 3</th>
<th>PY1</th>
<th>PY2</th>
<th>PY3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate maps of EID potential by disease and season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implement monitoring/warning systems for selected EID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Produce projections of future EID potential</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 4</th>
<th>PY1</th>
<th>PY2</th>
<th>PY3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Tethys app for display and analysis of EID database</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrate HiClimR to Tethys</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td>Present preliminary system to MAL-ED community</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refine and operationalize system</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ARL

Current: ARL 5 for Shigella / ARL 4 for others

Expectation: ARL 7 for some pathogens by early 2021, but further COVID-related slippage is possible

Goal: ARL 7
COVID-19

BEN ZAITCHIK, HAMADA BADR, LAUREN GARDNER, JUSTIN LESSLER – JHU
MARGARET KOSEK, JOSH COLSTON - UVA
Why?

A short and unreliable COVID-19 data record

Inconsistent and sometimes inappropriate definition of response variable

Inconsistent and sometimes inappropriate scales of analysis

Difficulty of accounting for non-meteorological predictors: behavior, policy, demographics, cultural practices, etc.

Differences between climate zones

Diverse and sometimes questionable methodologies

Challenge of isolating climate influence early in the pandemic
Creating a unified, reliable data record

- Maps all geospatial units globally into a unique standardized ID.
- Standardizes administrative names and codes at all levels.
- Standardizes dates, data types, and formats.
- Unifies variable names, types, and categories.
- Merges data from all credible sources at all levels.
- Cleans the data and fixing confusing entries.
- Integrates hydrometeorological variables at all levels.
- Optimizes the data for machine learning applications.

https://github.com/hsbadr/COVID-19

* NUTS 1 level represents groups of subregions (or equivalent) for some European countries (e.g., Italy).
** NUTS 2 level represents subregions (or equivalent) for some European countries (e.g., Italy).
Creating a unified, reliable data record

red = Admin0, blue = Admin1, green >= Admin2
Selecting a response variable

Following graphs were generated in R using the following epidemiological parameters:
- Daily new case data (JHU CSSE)
- Serial interval, incubation period, generation time
- Reporting delay
Hydrometeorological data

We have a “long list” of EO to incorporate, but for now we are focused on hydrometeorology drawn from reanalysis (NLDAS, MERRA2/FIPS and ERA5) and satellite-derived data (GPM).

Daily, spatially averaged to unit of COVID-19 case data in the harmonized database.
Non-environmental covariates

### Demographics

<table>
<thead>
<tr>
<th>Country</th>
<th>cases</th>
<th>deaths</th>
<th>excess deaths</th>
<th>population</th>
<th>pop. Density</th>
<th>incidence score</th>
<th>Accessibility</th>
<th>intervention data</th>
<th>Facebook Mobility</th>
<th>Google mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colombia</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>*1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Peru</td>
<td>3</td>
<td>3</td>
<td>0*</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>*0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Chile</td>
<td>2</td>
<td>1*</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>*~1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ecuador</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>*1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note: * denotes the highest score.
Non-environmental covariates

Demographics

Non-pharmaceutical interventions

**scientific data**

HIT-COVID, a global database tracking public health interventions to COVID-19


*Scientific Data 7, Article number: 286 (2020) | Cite this article*
Non-environmental covariates

Demographics

Non-pharmaceutical interventions

Comorbidity:
- diabetes, obesity, HIV, hypertension, smoking, COPD, cardiovascular disease index all compiled at Admin0 globally and Admin1 for the US

Mobility
- High resolution mobility data for the US and selected European countries
- International air travel data
Current analyses:

1. Global analysis at national level, covering ~187 countries

2. Higher-resolution global analysis at Admin 1

3. Detailed analysis for selected countries: US, Germany, Italy, Colombia, Peru, Ecuador, Chile
Collaboration across the AST!

Our EO database has been leveraged for:

- COVID-19 stay at home orders and heat-related illness by Suwei Wang (Julia Gohlke’s group)
- County-level analysis of COVID-19 risk for the US by Bill Pan’s group
Thank You