RIPA2TAN epidemiological study
Project update

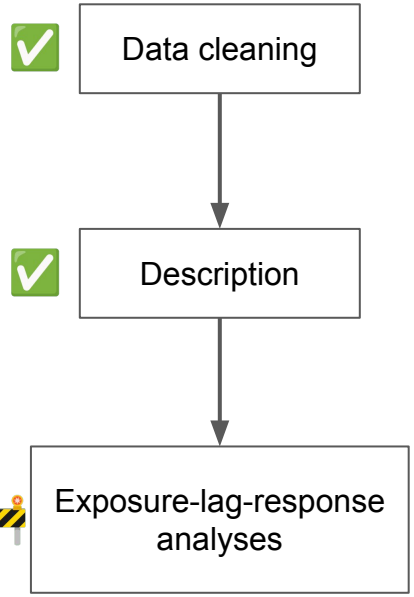
Alejandro Navarro Martínez

Computational Biology Group - Life Sciences Department

RIPA2TAN project meeting 6th August, 2024

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Overview



Environmental + disease + population

Describe **trend** of the environmental variables and disease incidence rates

Implemented GAMs that decompose the time series into trend and seasonal components

—

Used models to **impute** the missing environmental variables in the period 2014-2021

$g[E(\text{Incidence rate})] \propto f(\text{exposure, lag})$

Implement the GAMs form the previous step, adding DLNM terms

Issues:

47 diseases

8 exposures

Select variables; frame questions

Univariate models for each exposure, or multivariate models?
Need causal reasoning

Descriptive models

$$g[\mathbb{E}(y_i)] = \underbrace{\alpha_{d[i]} + f_{1,d[i]}(t_i)}_{\text{Trend}_i} + \underbrace{f_{2,d[i]}(m_i) + f_{3,d[i]}(t_i, m_i)}_{\text{Seasonal}_i}$$

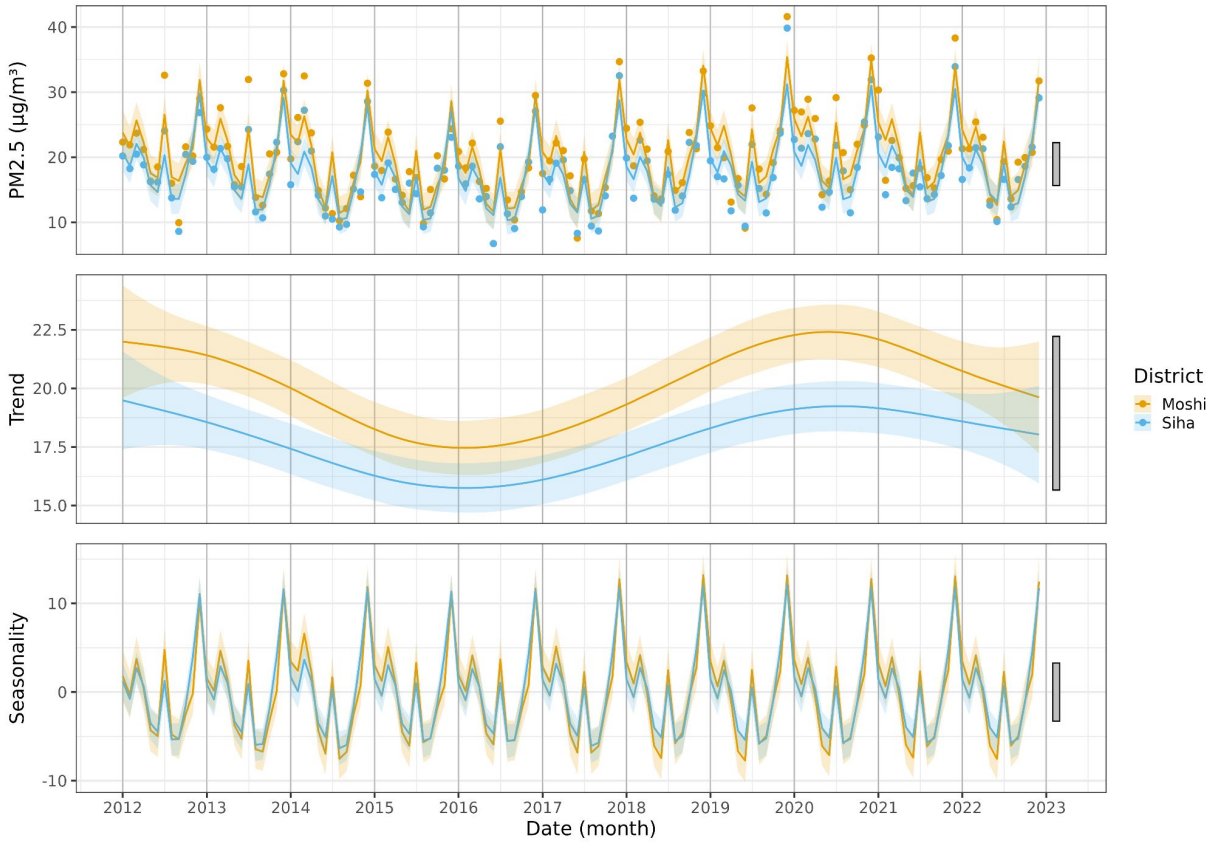
$$d_i \in \{\text{Moshi}, \text{Siha}\}$$
$$t_i = 1, \ldots, T$$
$$m_i = 1, \ldots, 12$$

Penalized spline functions:

$$\begin{cases} f_{1,d[i]}(t_i) = \sum_{j=1}^{J} \beta_{d[i],j}\, b_{1,j}(t_i) \\ f_{2,d[i]}(m_i) = \sum_{k=1}^{12} \beta_{d[i],k}\, b_{2,k}(m_i) \\ f_{3,d[i]}(t_i, m_i) = \sum_{j=1}^{J} \sum_{k=1}^{12} \beta_{d[i],j,k}\, b_{1,j}(t_i)\, b_{2,k}(m_i) \end{cases}$$

Descriptive models: environmental variables

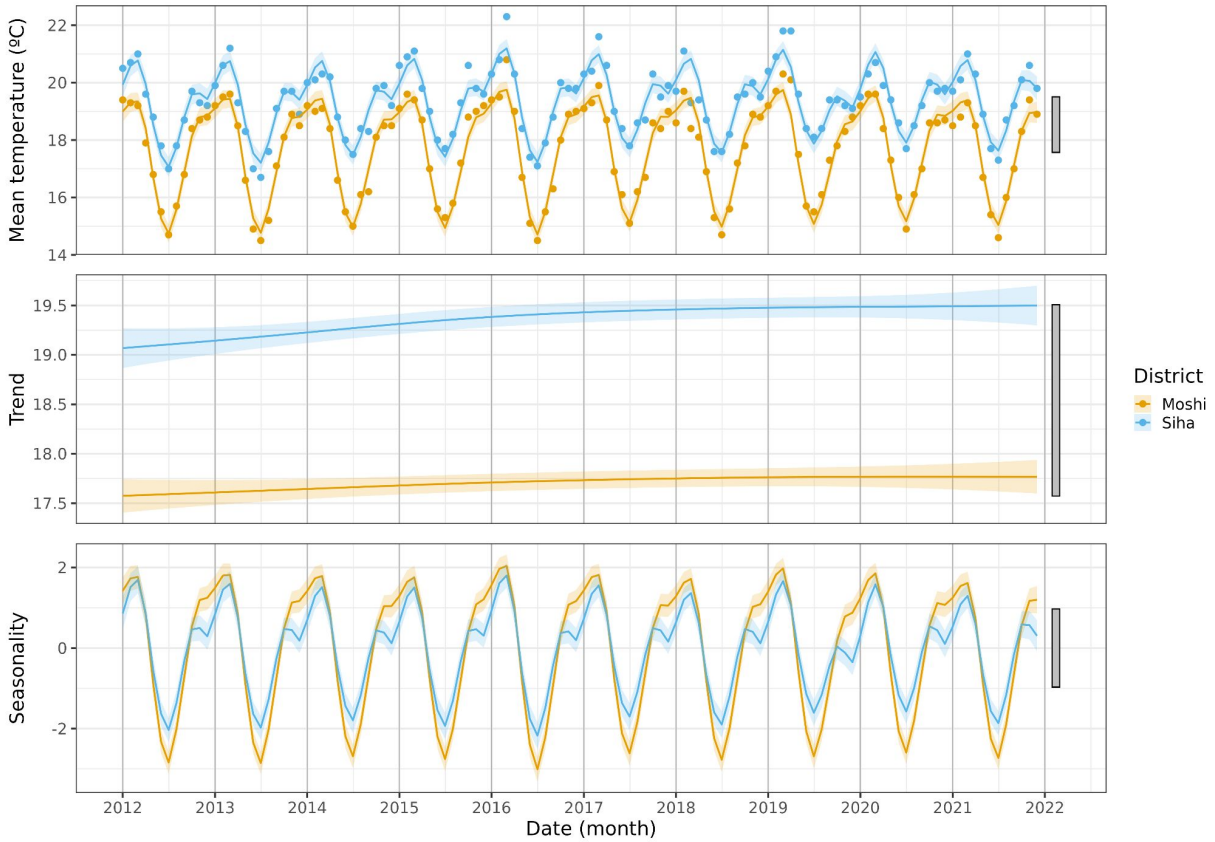
- pm2p5
- temp_min
- temp_mean
- temp_max
- total_rainfall
- n_raindays
- greenness
- utci



$$g[\mathbb{E}(y_i)] = \mathbb{E}(y_i) = \text{Trend}_i + \text{Seasonal}_i$$

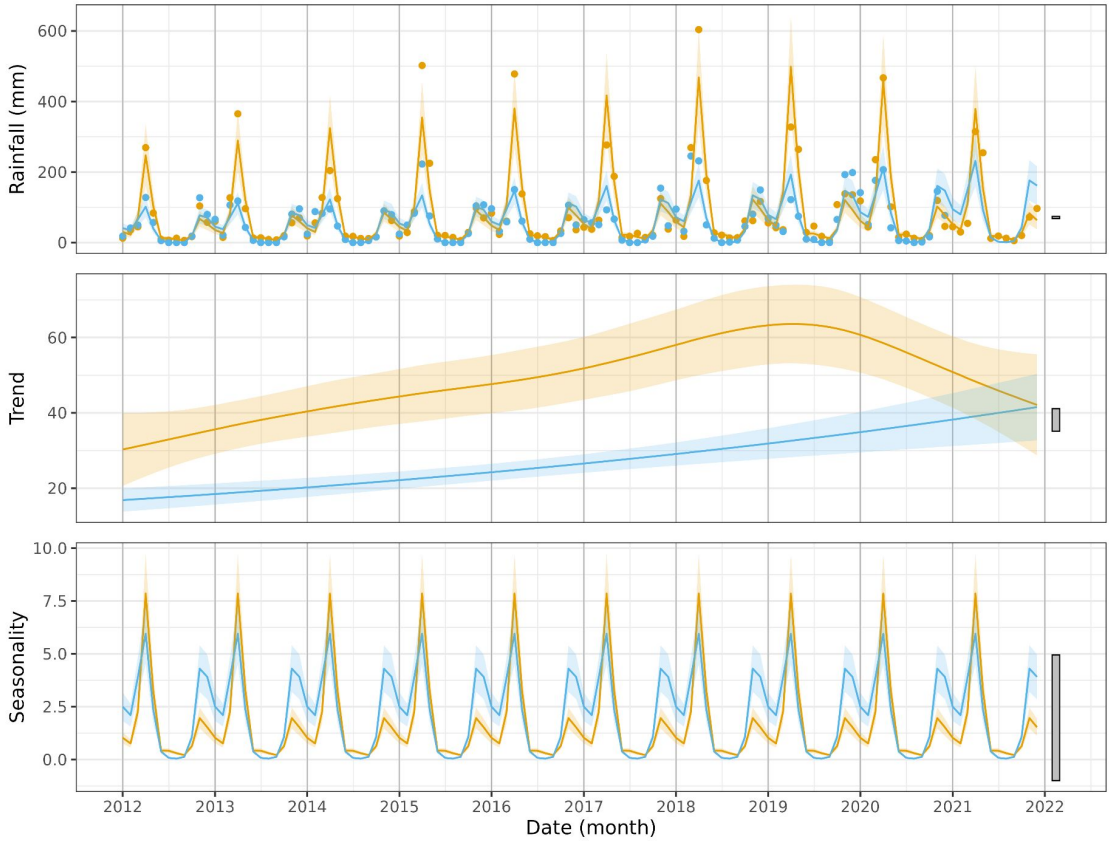
Descriptive models: environmental variables

- pm2p5
- temp_min
- temp_mean
- temp_max
- total_rainfall
- n_raindays
- greenness
- utci



Descriptive models: environmental variables

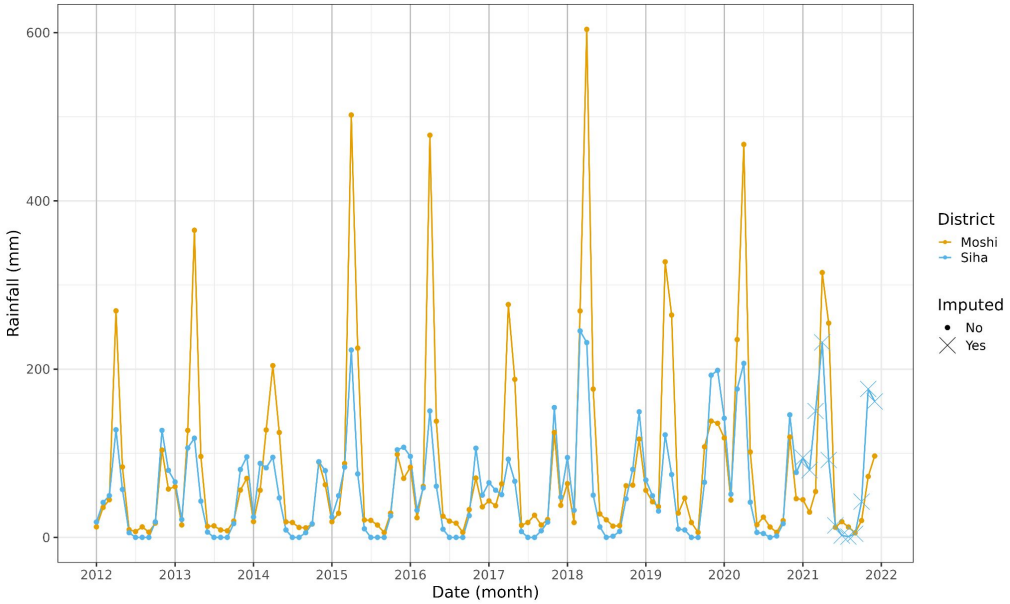
- pm2p5
- temp_min
- temp_mean
- temp_max
- total_rainfall
- n_raindays
- greenness
- utci

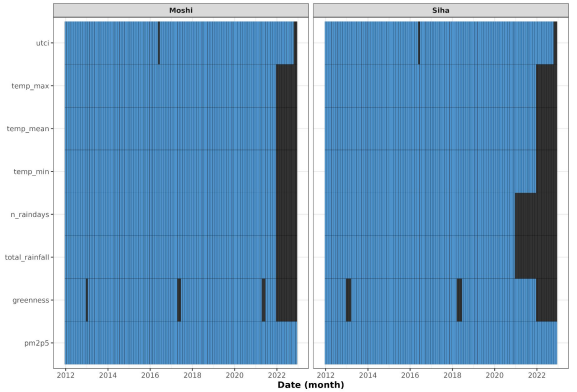


+ Imputation for Siha 2021

Descriptive models: environmental variables

- pm2p5
- temp_min
- temp_mean
- temp_max
- total_rainfall
- n_raindays
- greenness
- utci





2014-2021 complete now

Descriptive models: disease incidence rates

#47 diseases

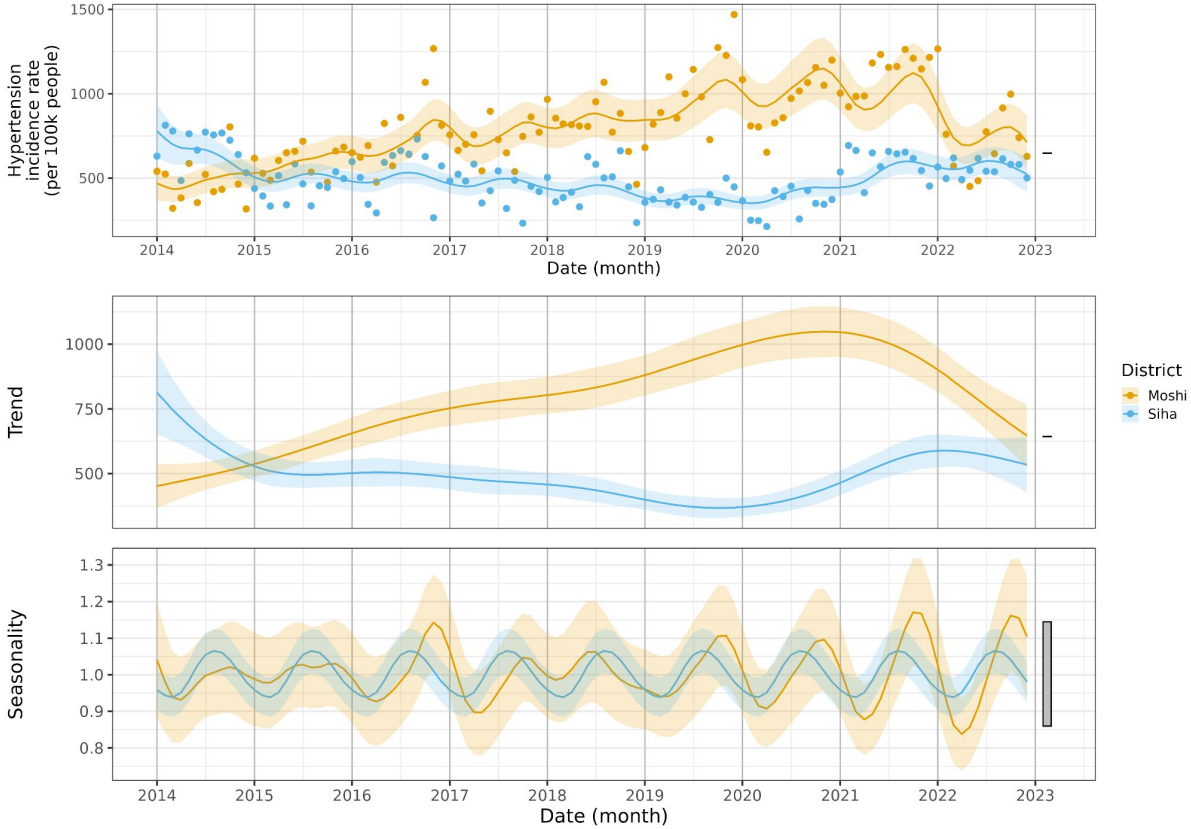
Bronchial Asthma

$$\log[\mathbb{E}(y_i)] = \text{Trend}_i + \text{Seasonal}_i$$
$$\mathbb{E}(y_i) = \exp(\text{Trend}_i)\ \exp(\text{Seasonal}_i)$$

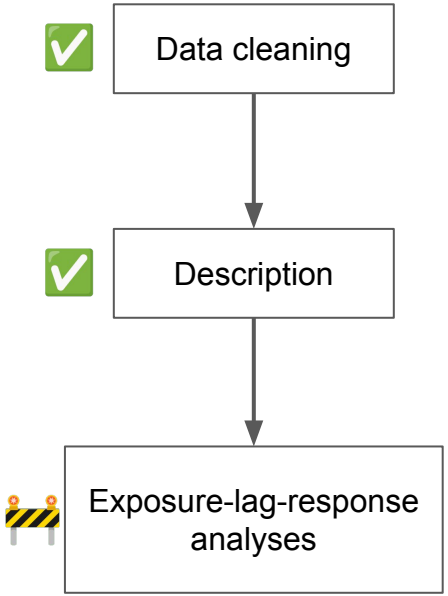
Descriptive models: disease incidence rates

#47 diseases

Hypertension



Work ahead


✅ Data cleaning
Environmental + disease + population
✅ Description
Describe **trend** of the environmental variables and disease incidence rates
Implemented GAMs that decompose the time series into trend and seasonal components
—
Used models to **impute** the missing environmental variables in the period 2014-2021
Exposure-lag-response analyses
g[E(Incidence rate)] ∝ f(exposure, lag)
Implement the GAMs form the previous step, adding DLNM terms
Issues:
47 diseases
8 exposures
Select variables; frame questions
Univariate models for each exposure, or multivariate models?
Need causal reasoning

Example of exposure-lag-response function

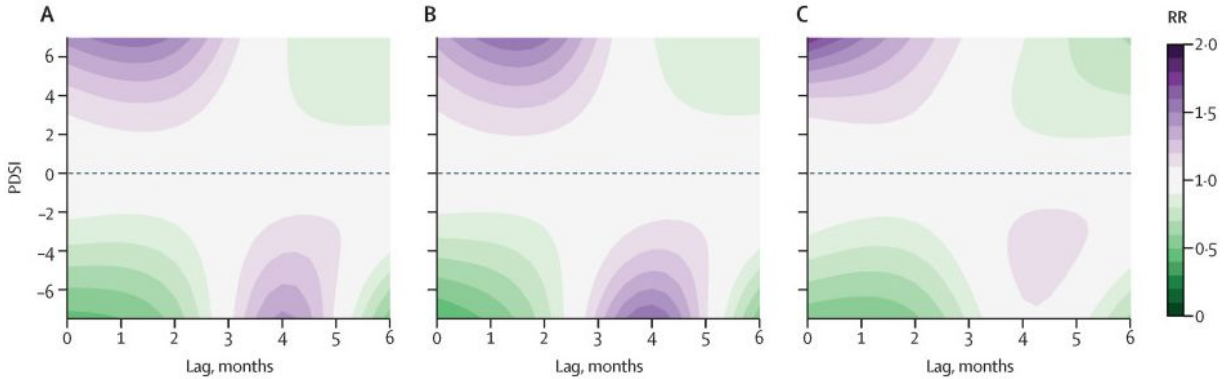
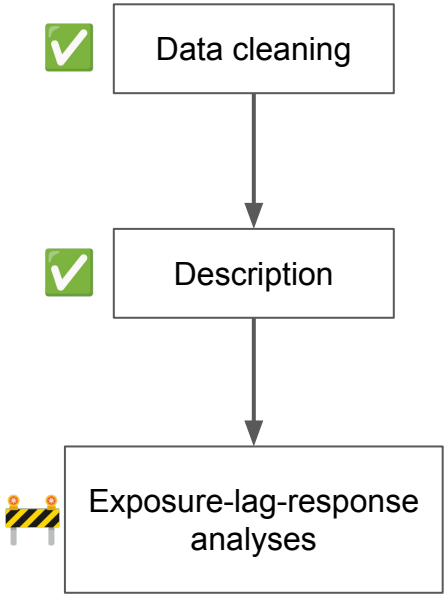


Figure 4 Association between risk of dengue and drought severity index at different time lags overall, and for high and low levels of urbanisation

Lowe, R., et al. (2021). Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: A spatiotemporal modelling study. *The Lancet Planetary Health*, *5*(4), e209–e219. https://doi.org/10.1016/S2542-5196(20)30292-8

Issues



Environmental + disease + population

Describe **trend** of the environmental variables and disease incidence rates

Implemented GAMs that decompose the time series into trend and seasonal components

—

Used models to **impute** the missing environmental variables in the period 2014-2021

g[E(Incidence rate)] ∝ f(exposure, lag)

Implement the GAMs form the previous step, adding DLNM terms

Issues:

47 diseases

8 exposures

Select variables; frame questions

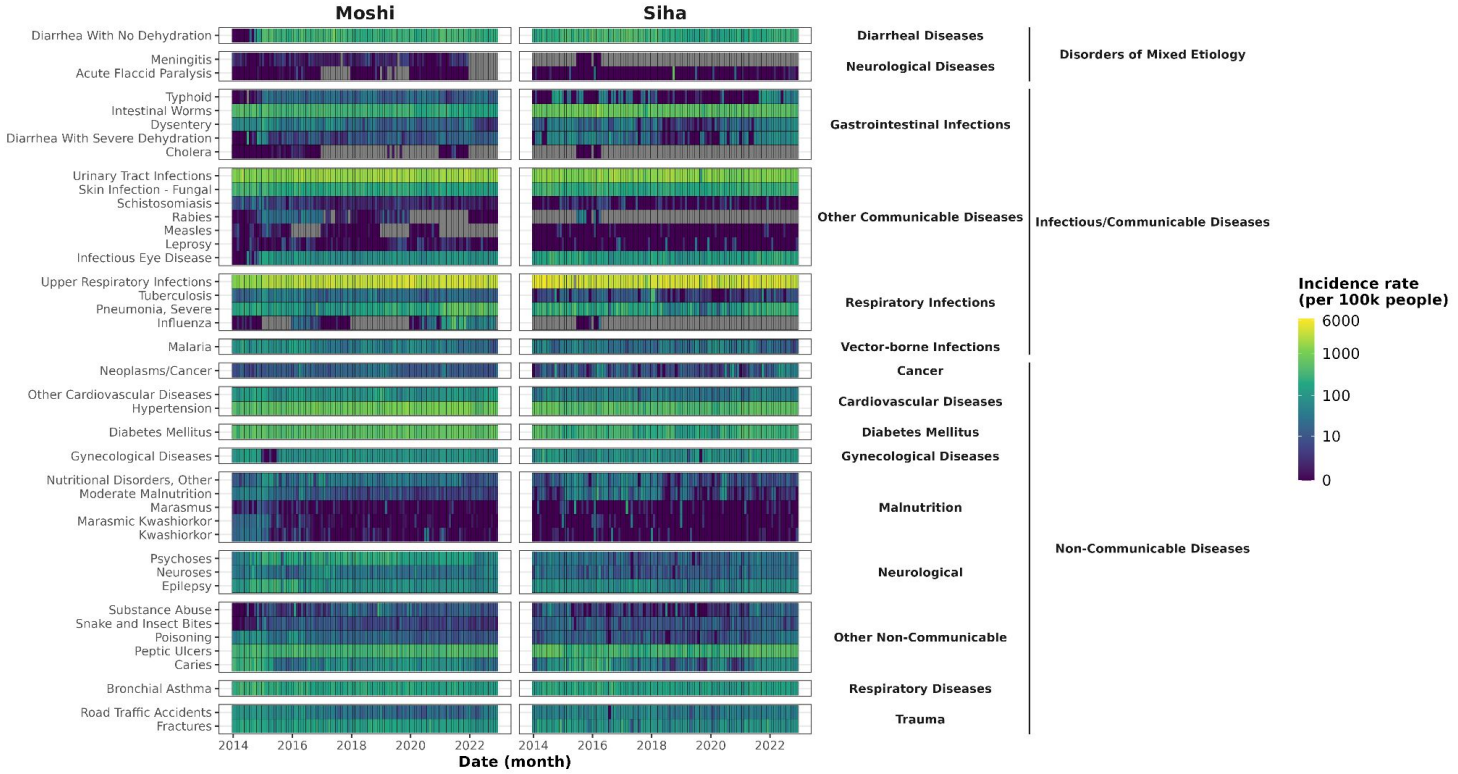
Univariate models for each exposure, or multivariate models?
Need causal reasoning

Select variables; frame questions

(Chrysanthi's grouping)

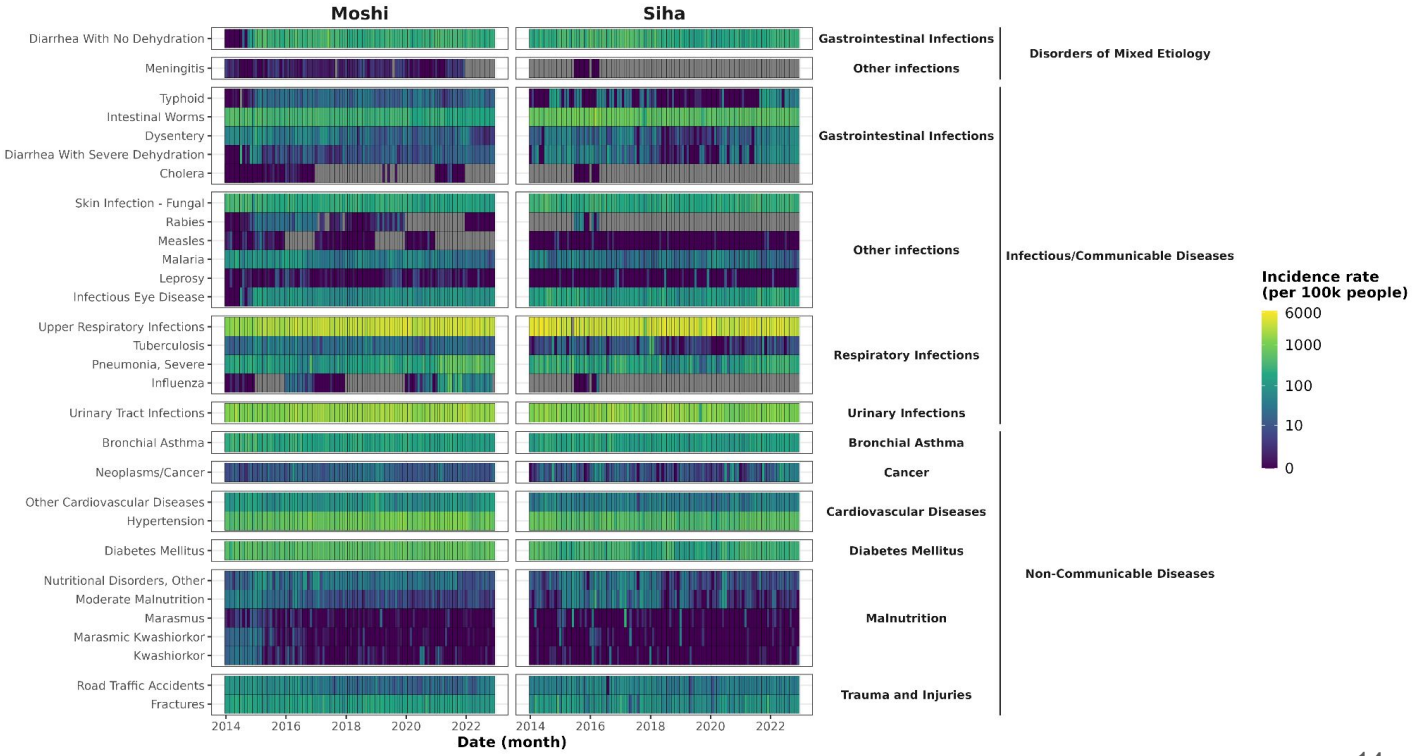
Environmental data

Variable	Type	Values
district	Categorical	{Moshi, Siha}
date	Datum	[2012-01, 2022-12]
pm2p5	Numerical, positive rational	[6, 42]
temp_min	Numerical, rational	[9, 18]
temp_mean	Numerical, rational	[14, 23]
temp_max	Numerical, rational	[18, 28]
total_rainfall	Numerical, positive rational	[0, 604]
n_raindays	Numerical, natural	[0, 19]
greenness	Numerical, rational	[0.24, 0.70]
utci	Numerical, rational	[6, 37]



Select variables; frame questions

(Harald's grouping)



Multivariate exposure models? Causal reasoning needed

AIR POLLUTION

Commentary

Does Air Pollution Confound Studies of Temperature?

Buckley, Jessie P.[a]; Samet, Jonathan M.[b]; Richardson, David B.[a]

Author Information

Epidemiology 25(2):p 242-245, March 2014. | **DOI:** 10.1097/EDE.0000000000000051

https://doi.org/10.1097/EDE.0000000000000051

LETTER • OPEN ACCESS

Heat, humidity and health impacts: how causal diagrams can help tell the complex story

Sidharth Sivaraj[1,2], Jakob Zscheischler[3,4], Jonathan R Buzan[2,5], Olivia Martius[2,6], Stefan Brönnimann[2,6] and Ana M Vicedo-Cabrera[1,2,7]

Published 5 July 2024 • © 2024 The Author(s). Published by IOP Publishing Ltd

Environmental Research Letters, Volume 19, Number 7

Citation Sidharth Sivaraj *et al* 2024 *Environ. Res. Lett.* **19** 074069

DOI 10.1088/1748-9326/ad5a25

https://doi.org/10.1088/1748-9326/ad5a25

RIPA2TAN epidemiological study
Project update

Alejandro Navarro Martínez

Computational Biology Group - Life Sciences Department

RIPA2TAN project meeting 6th August, 2024

Barcelona Supercomputing Center
Centro Nacional de Supercomputación