Droughts and Despair: A Cutting-Edge Tool for Assessing Climate Change's Economic Impact on Mental Health in Farming Communities

Team: Olatunji-Fairbrass-Lucas

Wellcome Data Science Ideathon 2023

12/07/23

The Team

- Olatunji Johnson – University of Manchester
  - Statistical methods in health
- Alison Fairbrass – University College London
  - Environmental economics
- Tim Lucas – University of Leicester
  - Environmental health

Climate and health challenge 1

- Demonstrate how **data science approaches** can be used for prediction of the economic **impacts of climate change mitigation policy** on health.

- To predict impacts of **climate change mitigation policy** on health requires robust inference of causal effects.

- Causal inference is the appropriate **data science approach.**

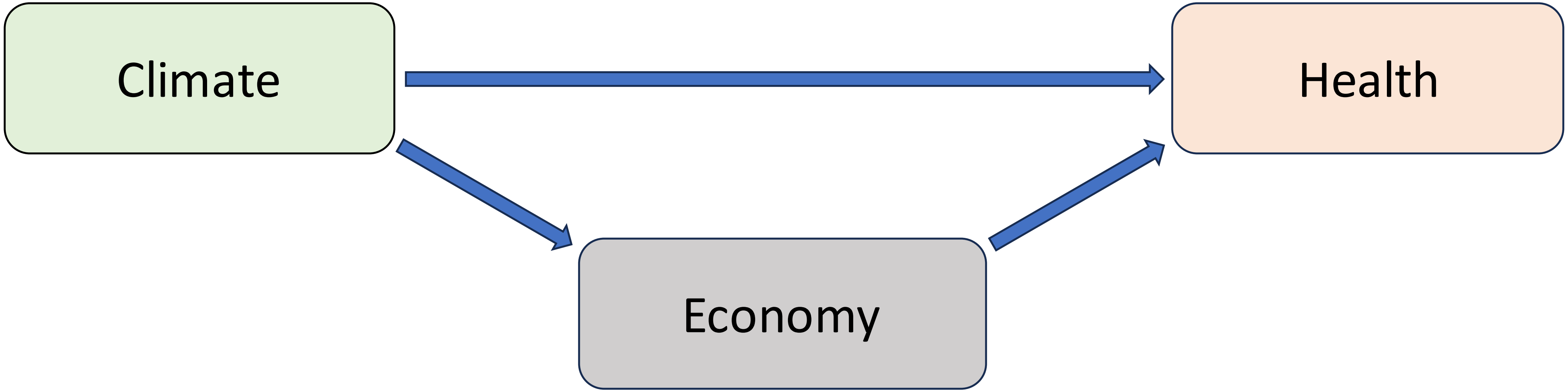
Our approach

- Robustly estimate and communicate effects of policy at climate-economy-health interface.
- Estimating causal effects from observational data is usually **biased**.
- Novel insight from causal inference:
  - Bias is small if direct effect is small.
- Case study of our approach: suicide in agricultural workers.
- WebApp for policy makers to explore policy implications
  - Robust, unbiased policy assessment.

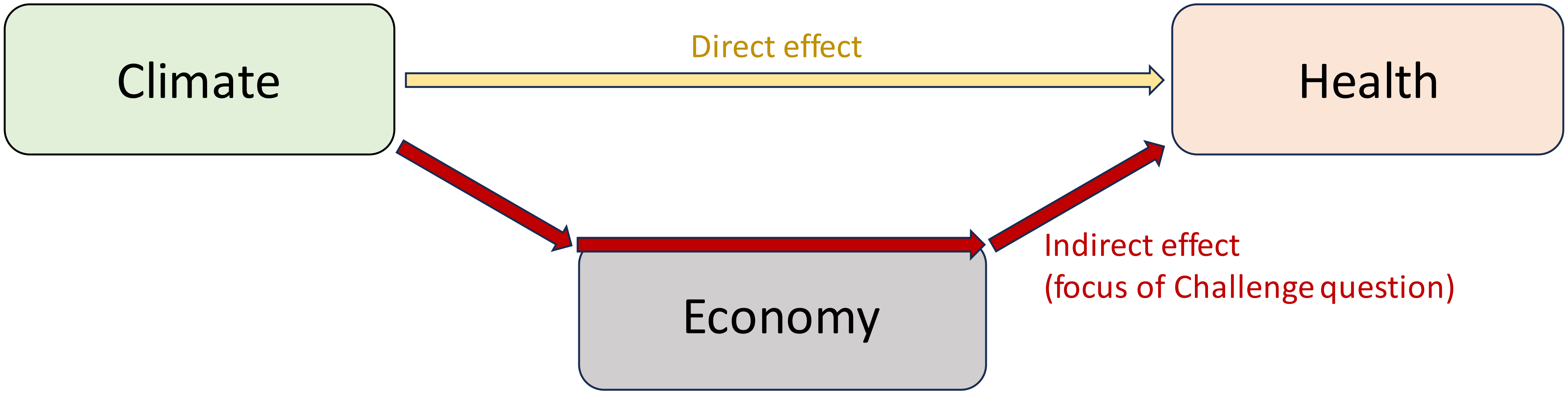
Our approach

- "Indirect effect" is a metric
- Tells the policy maker how effective policies based on economic pathway will be.
- Without such metrics, policies that are focused on irrelevant pathways will seem reasonable.

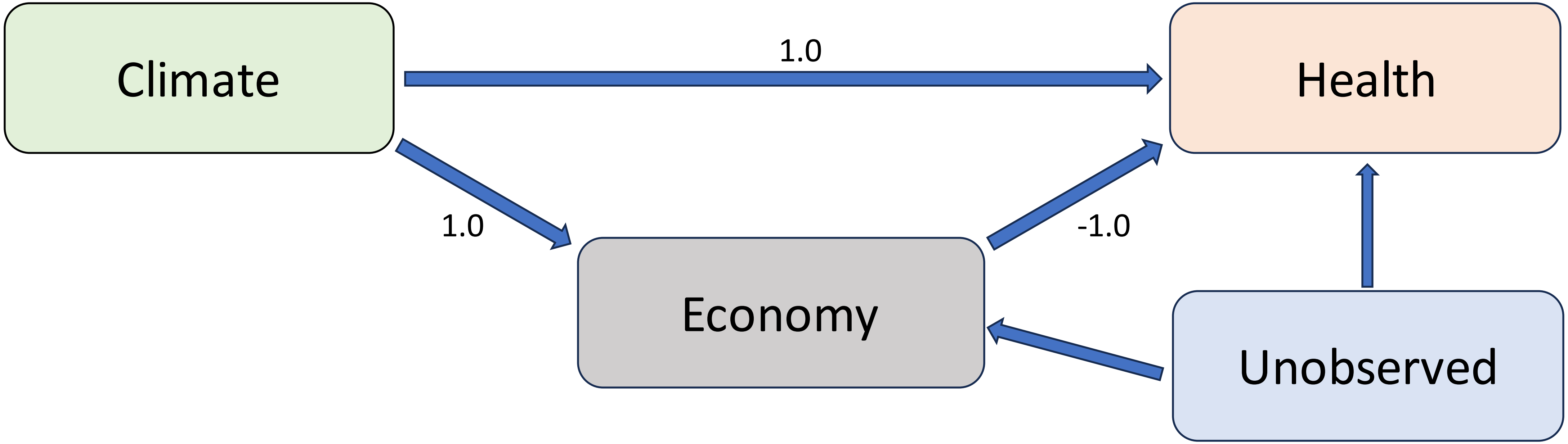
Prototype part 1:
Causal inference framework



Prototype part 1:
What are we trying to estimate?

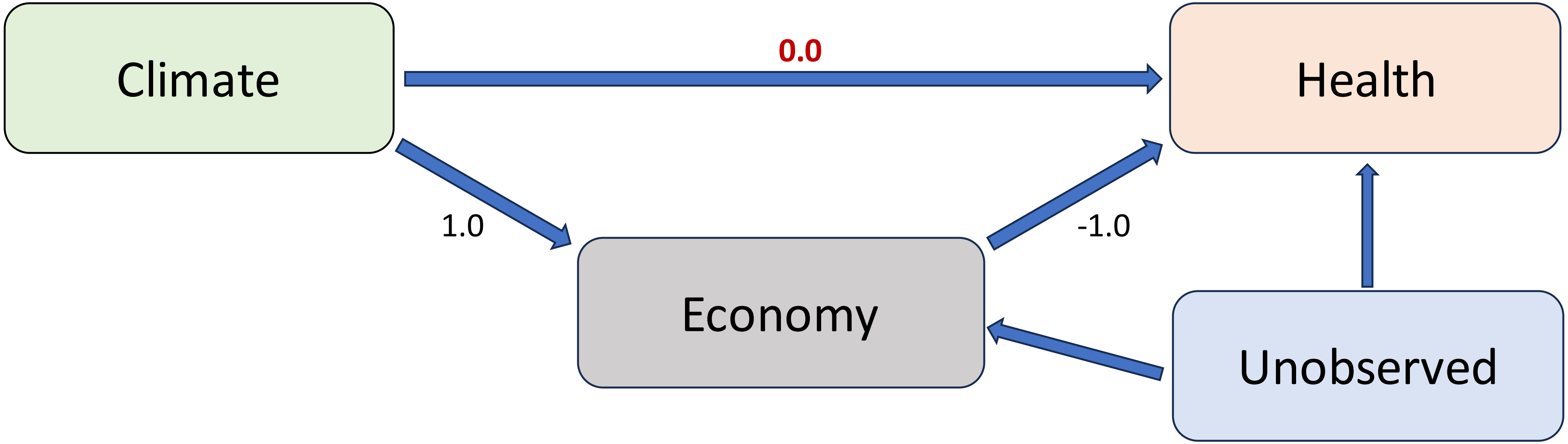


Prototype part 1:
There's always unobserved variables!



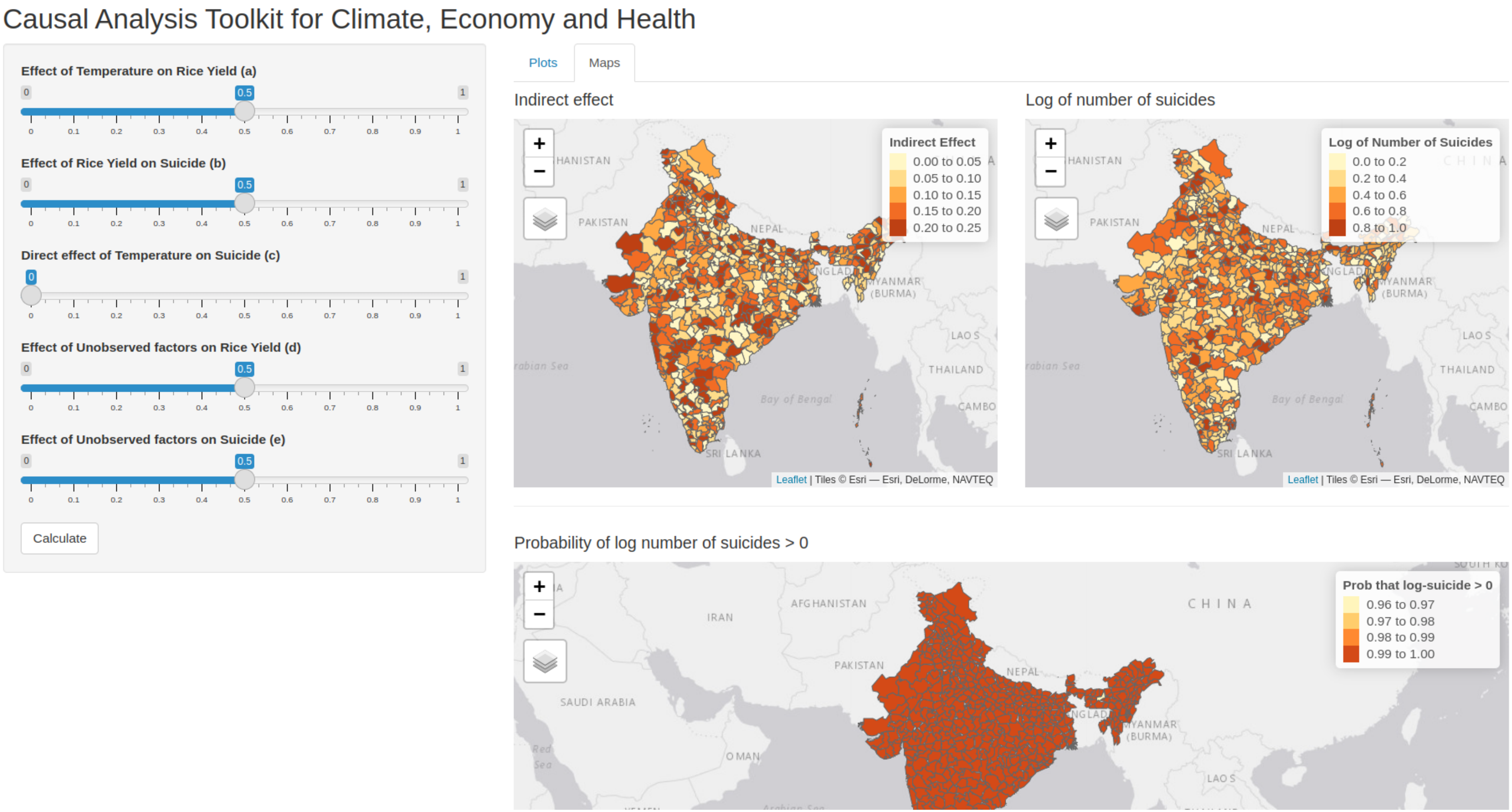
True indirect effect = -1.0
Estimated indirect effect = **-0.004** **(Biased!)**

Prototype part 1:
Unbiased estimate of direct effect, if direct effect is small.

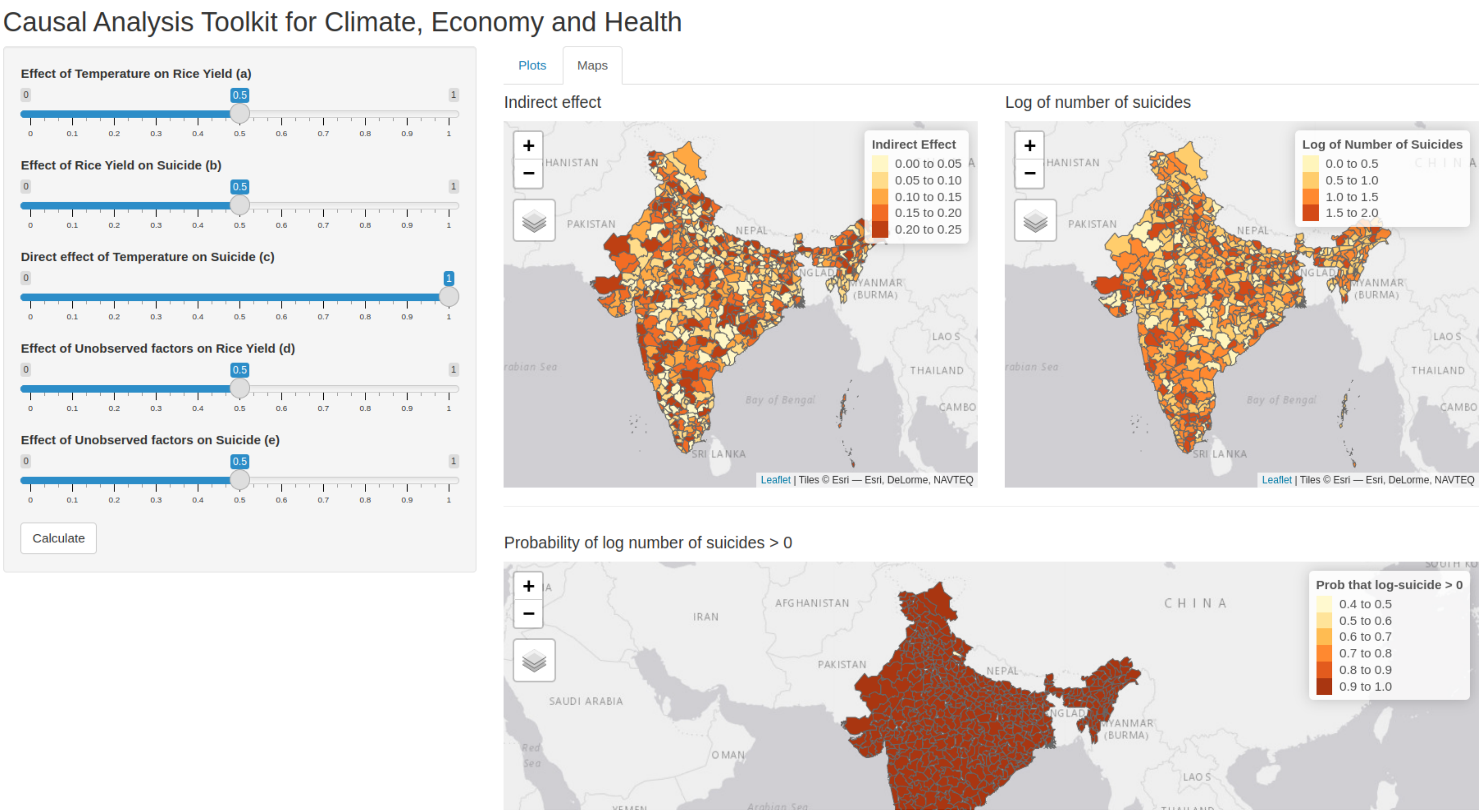


True indirect effect = -1.0
Estimated indirect effect = -1.0006 (Unbiased)

Prototype part 2: Communicating policy assessment

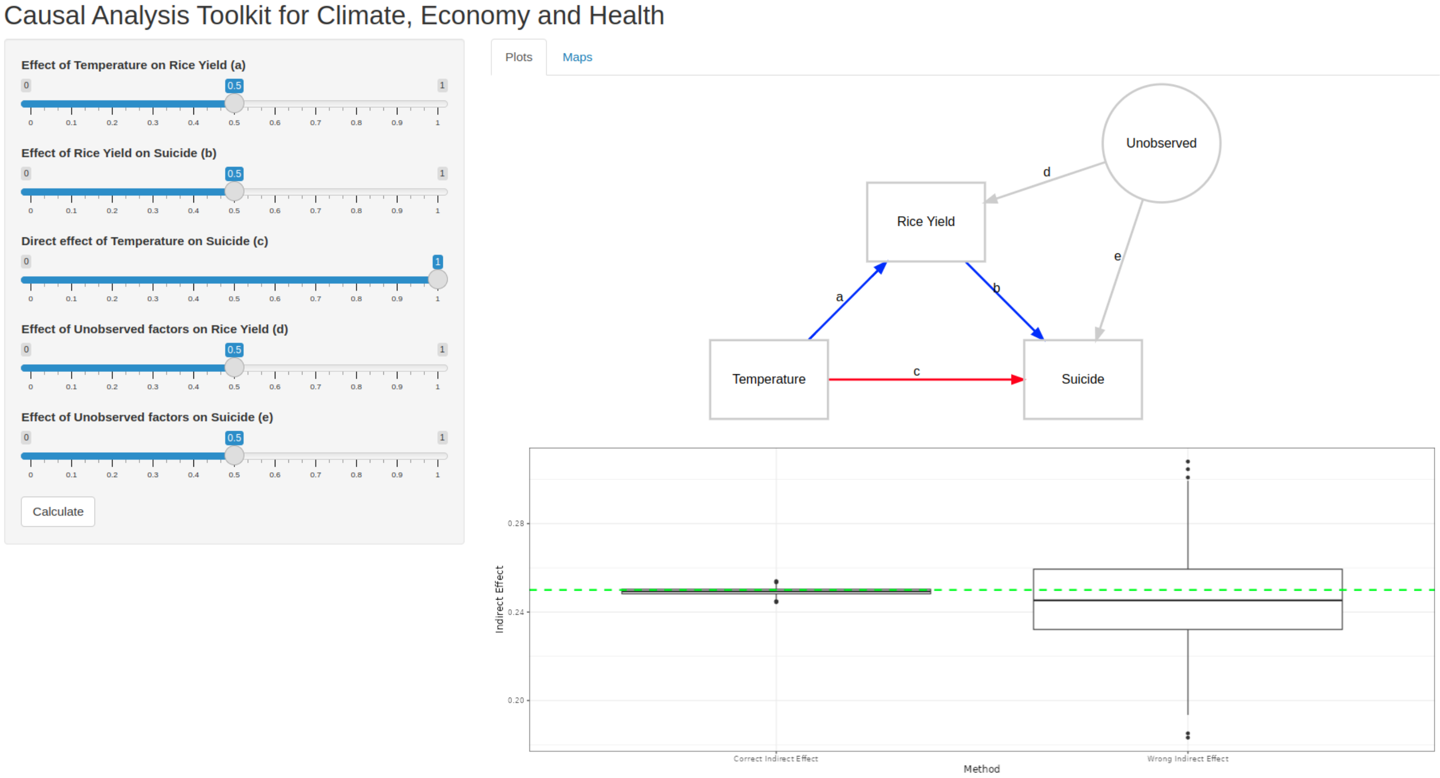


Prototype part 2:
Allow policy maker to "input policy".



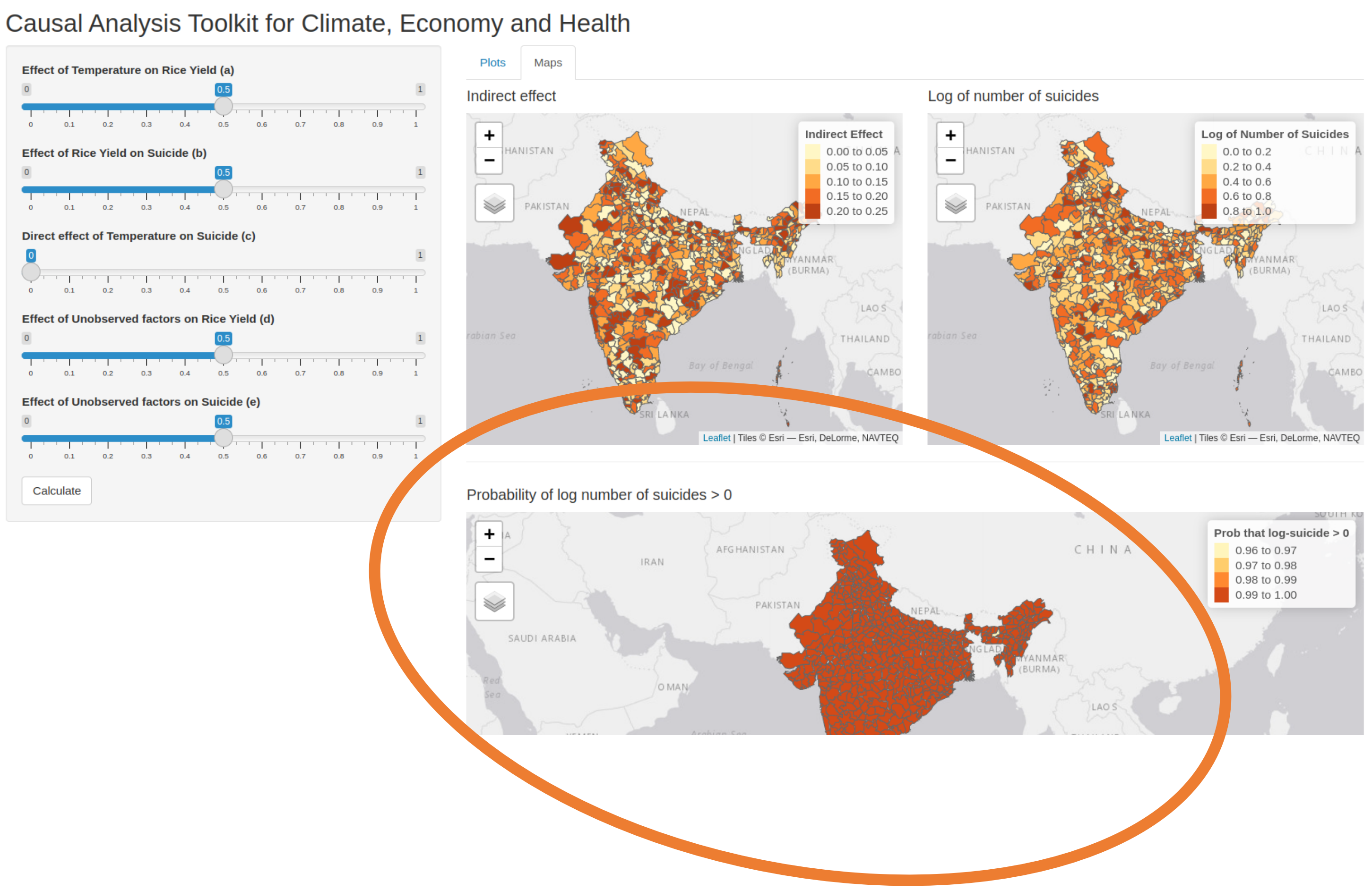
Prototype part 2: Communicating policy assessment

- Requires trust
- Trust is built by making robust, causal estimates of effect



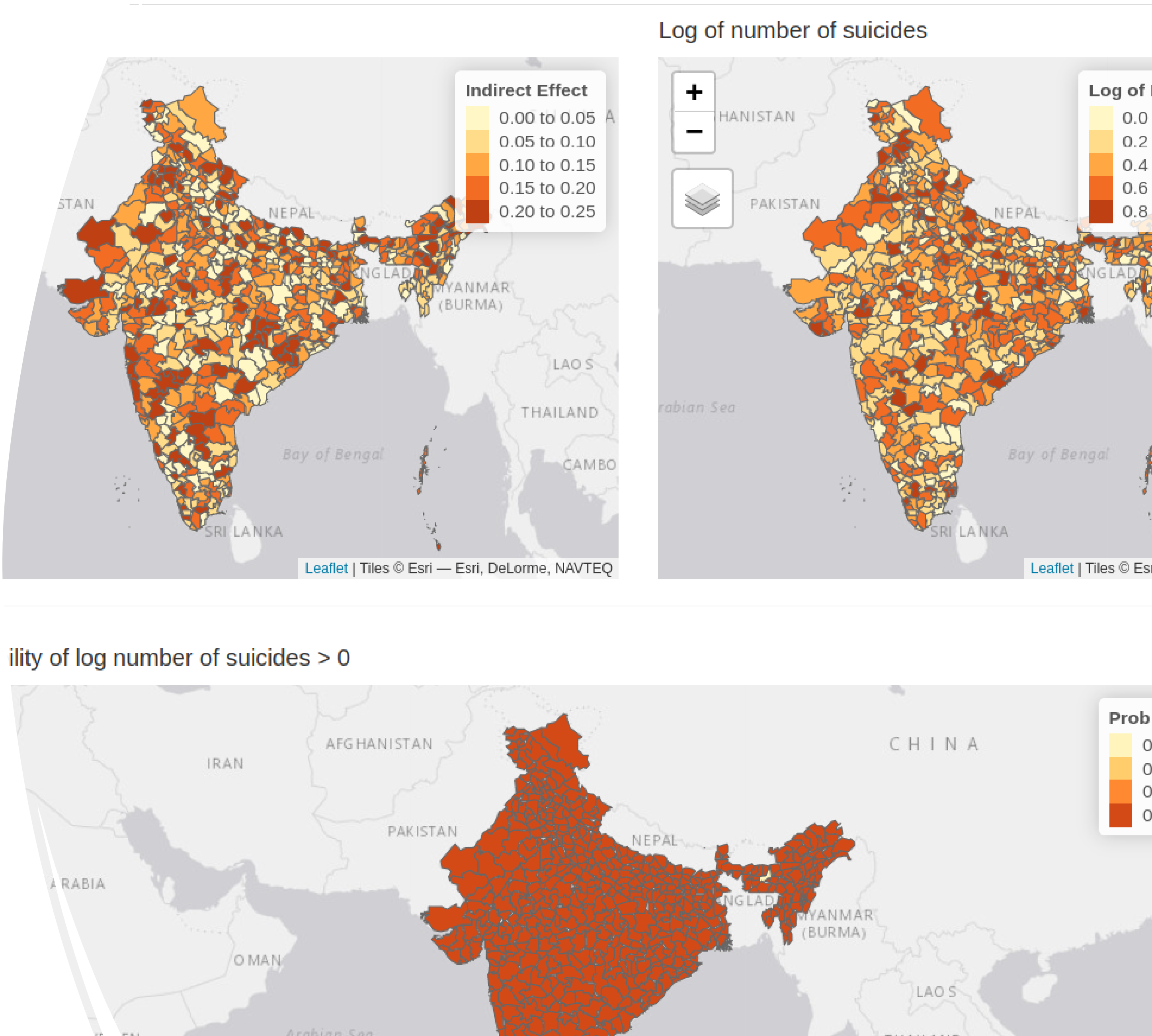
Prototype part 2: Communicating policy assessment

- Communicates uncertainty with *exceedance probability*
- Probability (accounting for uncertainty) that health metric exceeds policy relevant threshold.



Prototype part 2: Communicating policy assessment

- No-code
- Open-access (shinyapps)
- Open-source (GitHub)
- Visualizes causal estimates of indirect effect of climate on health
- Download estimates and uncertainty
    - Confidence intervals, prediction draws or exceedance probability
    - Csv or excel

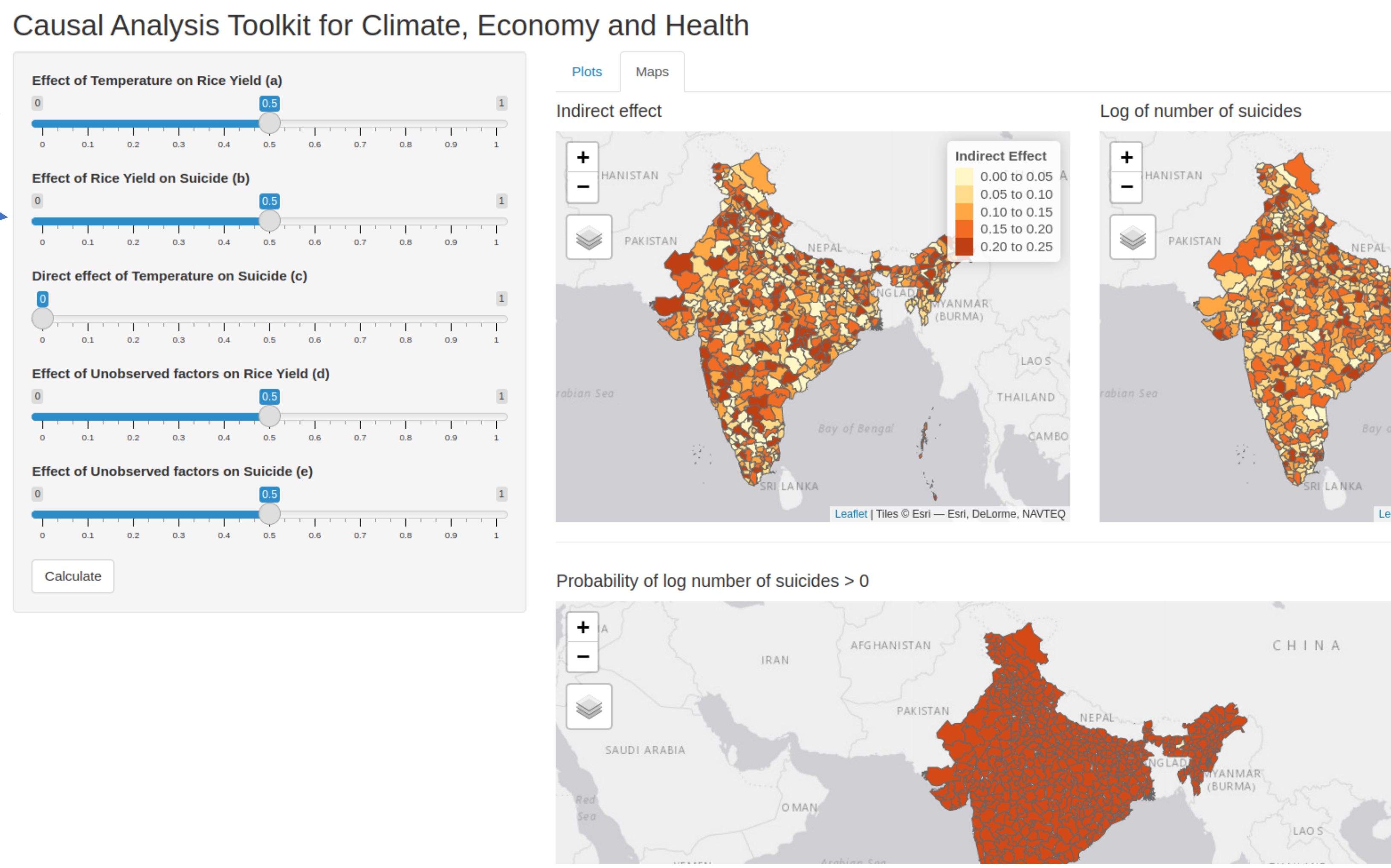


Demonstrating Real-World Utility

Policy A: New rice strain

Policy B: universal basic income

Policy C: Free Counselling



- Explanation of how the platform ensures the usefulness of metrics/models developed on it in real-world scenarios, emphasizing the advisory group's role and the involvement of real-world stakeholders.

- Ensuring Generalization to Updated/New Data (2 minutes)

- Discussion on the platform's ability to adapt to continually updated climate and economic data as well as new data sources, focusing on the platform's flexible and adaptable design.

- Handling Uncertainty and Variability in Climate Data (3 minutes)
- I think we have covered this in prototype

- Explanation of how the platform, particularly its analysis and visualization tools, addresses the inherent uncertainty and variability in climate data, showing examples of this functionality.

Summary

- Great need to robustly assess climate-economy-health policy.
- Must engage stakeholders and build trust.
- Building trust requires robust estimates and assessments of policy.
- Assessments of policy is a **causal** problem.
- We propose using novel insights from causal inference.
- We will build a WebApp that:
  - Communicates robust, causal estimates
  - Allows assessment of policy effects
  - Fully visualizes uncertainty

Thanks for listening

potential structure for our 15-minute presentation:

- Brief introduction of the team and the project.
- Statement of the challenge and the approach taken to address it.

II. Demonstration of the Prototype (7 minutes)

- Explanation of the prototype design and functionalities, emphasizing the data science methods used to correlate climate change and its economic impacts on health.
- Demonstration of how various machine learning approaches, like representation learning, reinforcement learning, unsupervised learning, and supervised learning, have been implemented in the tool.
- Showcasing the no-code, open-access aspect of the platform, demonstrating how users can bring their own datasets, run models and workflows, and visualize, analyze, and download results.
- Explanation of identified climate-economic metrics and how they could be utilized for model development, optimization, and long-term deployment.

III. Answering the Challenge Questions (7 minutes)

- Demonstrating Real-World Utility (2 minutes)
- Explanation of how the platform ensures the usefulness of metrics/models developed on it in real-world scenarios, emphasizing the advisory group's role and the involvement of real-world stakeholders.
- Ensuring Generalization to Updated/New Data (2 minutes)
- Discussion on the platform's ability to adapt to continually updated climate and economic data as well as new data sources, focusing on the platform's flexible and adaptable design.
- Handling Uncertainty and Variability in Climate Data (3 minutes)
- Explanation of how the platform, particularly its analysis and visualization tools, addresses the inherent uncertainty and variability in climate data, showing examples of this functionality.

IV. Conclusion and Q&A (1 minute)

- Recap of the presentation, highlighting the main points and potential impacts.

Notes etc.

Writing to be copied into proposal

- Estimating causal effects from observational is difficult, though causal inference gives us a framework to approach this task. Distinguishing between direct and indirect effects (so called mediation analysis) is challenging. A number of assumptions must be made but these assumptions are often unrealistic and impossible to test. Notably, these difficulties exist both in standard statistical frameworks and in the machine learning context.

- This problem can be simplified by looking at disease systems where almost all of the effect is indirect i.e. where we a priori believe there is very little direct effect. In this case, the indirect effect is equal to the combined effect and can be estimated much more easily, with much more robustness to unmet assumptions or model mispecification.

- In this project we will use carefully applied causal inference approaches to estimate the indirect effect of climate on health via economic variables. In order to do this we have selected a disease system in which we believe that the direct effect is very small. We propose to explore diseases of despair (suicide, liver failure and drug overdose) in agricultural workers. We will therefore fully explore the causal pathway from climate change (heatwaves and drought), via economic activity (agricultural employment and yield) through to these important metal health related conditions.

- The assumption that the direct effect is small is important here. However, we can explore this assumption by fitting models to both workers in the agricultural sector and those in other industries. In other industries, where we expect the indirect effect to be weak, the combined effect of climate on health largely represents the direct effect. Therefore, confirming this small effect validates our central assumption.

Writing to be copied into proposal or wherever.

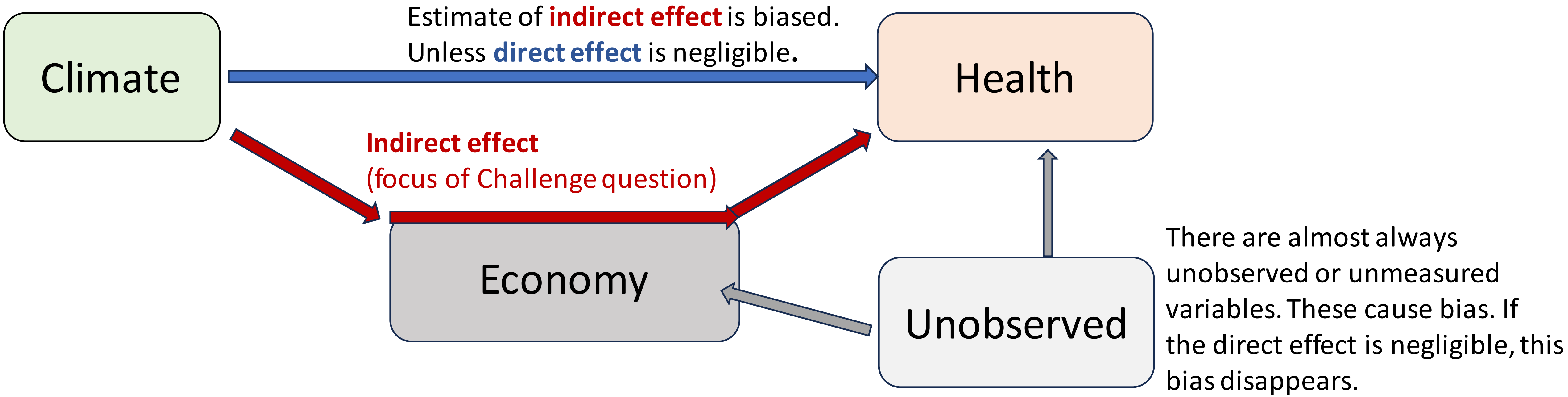
- As part of our prototyping for this project we have conducted a small simulation study to demonstrate the statistical concepts proposed.
- We simulated 10,000 observations of a temperature index from a uniform(0,1) distribution. We also simulated an unobserved confounder from a uniform(0, 1) distribution. We then simulated:
- Rice yield = 1 * temperature + 0.7 * unobserved + N(0, 0.01)
- Suicide rate = -1 * rice yield + 0.7 *unobserved + 0.5 * temperature + N(0, 0.01).
- The indirect effect here is –1 * 1 = -1.
- Estimating the indirect effect (by separately estimating the effect of temperature on rice yield and the effect of rice yield on suicide) gives -0.004, as we cannot control for the unobserved confounder.

Writing to be copied into proposal or wherever.

- In our second simulation we again simulated 10,000 observations of a temperature index from a uniform(0,1) distribution a an unobserved confounder from a uniform(0, 1) distribution. We then simulated:
- Rice yield = 1 * temperature + 0.7 * unobserved + N(0, 0.01)
- Suicide rate = -1 * rice yield + 0.7 *unobserved + 0.0* temperature + N(0, 0.01).
- i.e. the coefficient for temperature on suicide is now 0; there is no direct effect.
- The indirect effect here is again –1 * 1 = -1.
- Estimating the indirect effect (by directly estimating the combined direct and indirect effects) gives -1.0006, which is very close to the correct value. Despite the existence of the unobserved confounder, we can still estimate the indirect effect.

Policy impacts and causal inference

- Understanding the impacts of policy necessarily requires a causal framework. Therefore, the robust causal approach proposed lends itself to evaluating the impacts of policies on the disease system.
- We will create an app that allows policy makers to explore the impacts of various policies on suicide and


Climate
Estimate of **indirect effect** is biased.
Unless **direct effect** is negligible.
Health
Indirect effect
(focus of Challenge question)
Economy
Unobserved
There are almost always unobserved or unmeasured variables. These cause bias. If the direct effect is negligible, this bias disappears.

Project plan