Scalable Epidemiological Workflows to Support COVID-19 Planning and Response

Dustin Machi*, Parantapa Bhattacharya*, Stefan Hoops*, Jiangzhuo Chen*, Henning Mortveit*, Srinivasan Venkatramanan*, Bryan Lewis*, Mandy Wilson*, Arindam Fadikar‡, Tom Maiden†, Christopher L. Barrett* and Madhav V. Marathe*

* University of Virginia

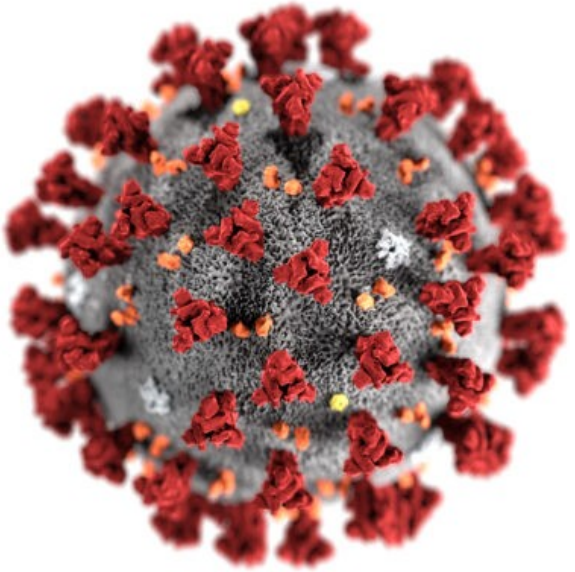
‡ Argonne National Laboratory

†Pittsburgh Supercomputing Center

IPDPS 2021

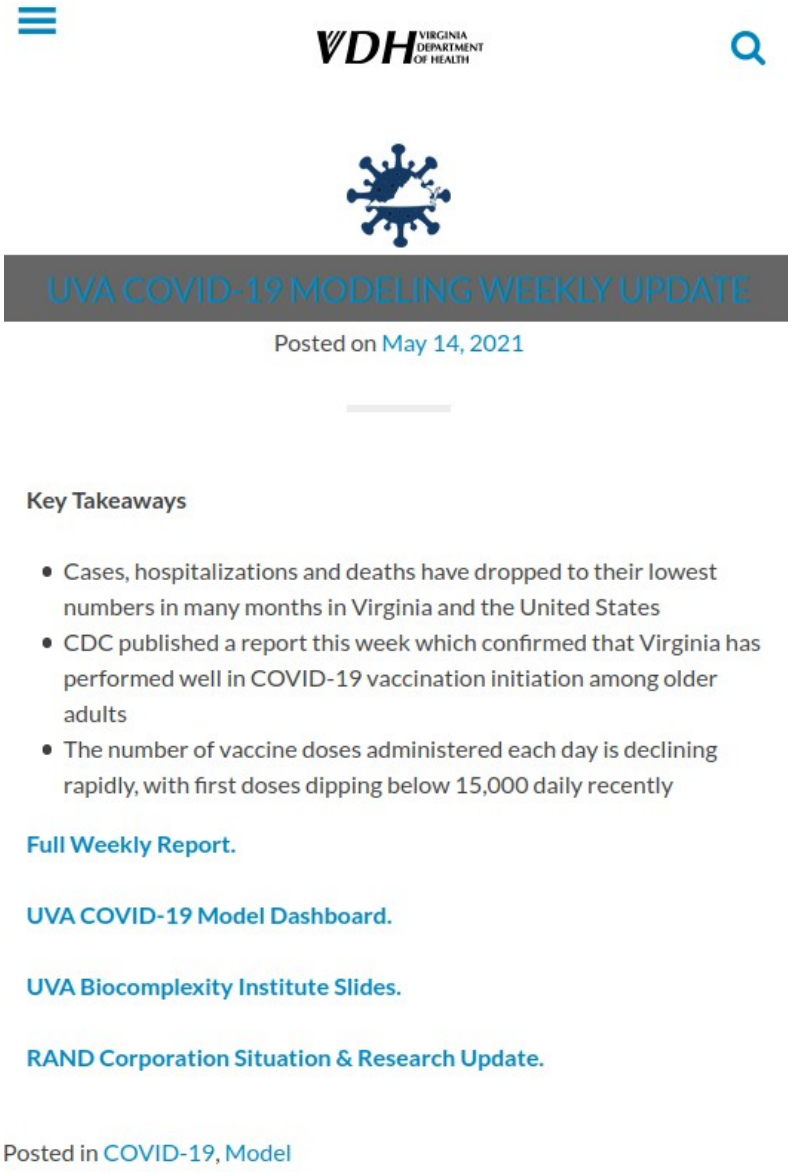
The Covid-19 Pandemic

- The most significant epidemic event since the 1918 influenza epidemic
- Over 164 million confirmed cases worldwide
- Over 3.39 million confirmed deaths worldwide
- Estimated economic burden of over 9 trillion US dollars



Supporting Policy Makers

- We have been supporting decision makers March 25, 2020
- We provided weekly forecasts to:
  - Center for Disease Control and Prevention (CDC)
  - US Department of Defense (DoD)
  - Virginia Department of Health (VDH)
  - State Hospital Referral Regions (HRR)
  - University of Virginia
- Answering what if questions
  - Non-pharmaceutical Interventions (NPIs)
  - Mask mandate, school closures, local shutdowns, …



VDH VIRGINIA DEPARTMENT OF HEALTH

UVA COVID-19 MODELING WEEKLY UPDATE

Posted on May 14, 2021

Key Takeaways

- Cases, hospitalizations and deaths have dropped to their lowest numbers in many months in Virginia and the United States
- CDC published a report this week which confirmed that Virginia has performed well in COVID-19 vaccination initiation among older adults
- The number of vaccine doses administered each day is declining rapidly, with first doses dipping below 15,000 daily recently

Full Weekly Report.

UVA COVID-19 Model Dashboard.

UVA Biocomplexity Institute Slides.

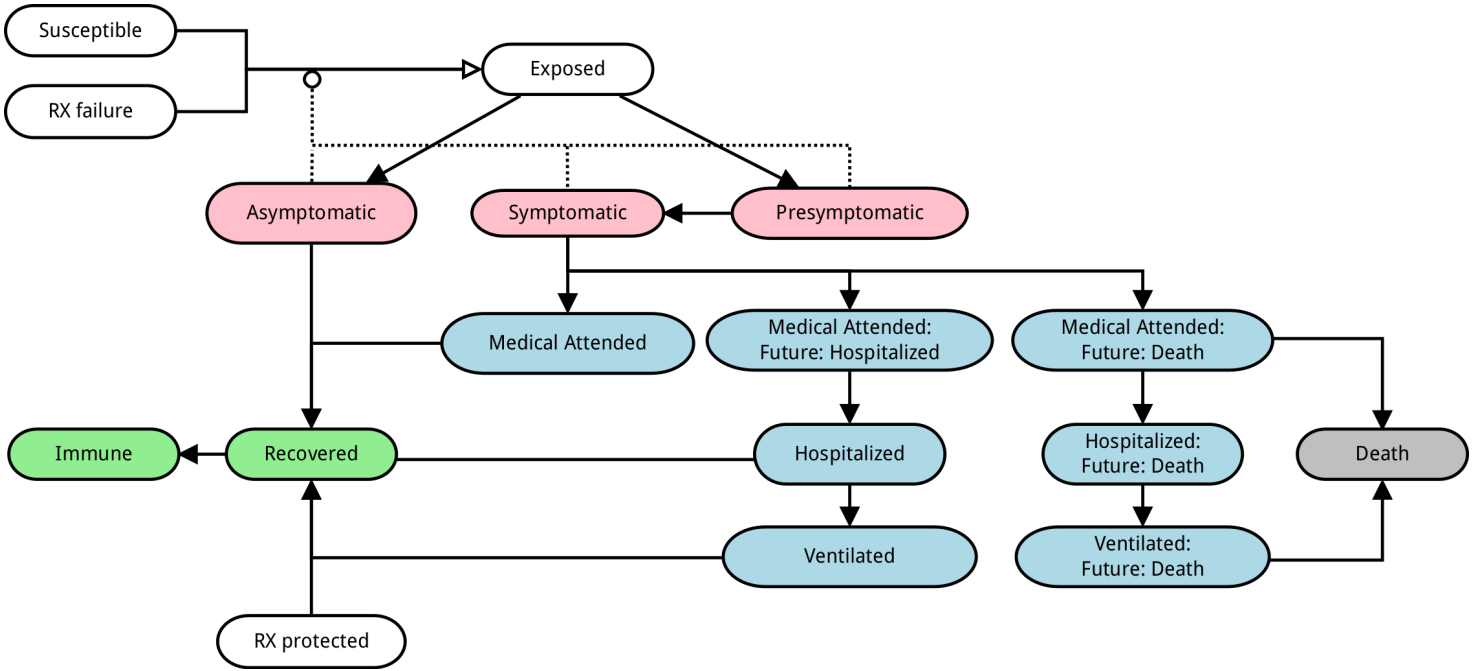
RAND Corporation Situation & Research Update.

Posted in COVID-19, Model

Contributions and Significance

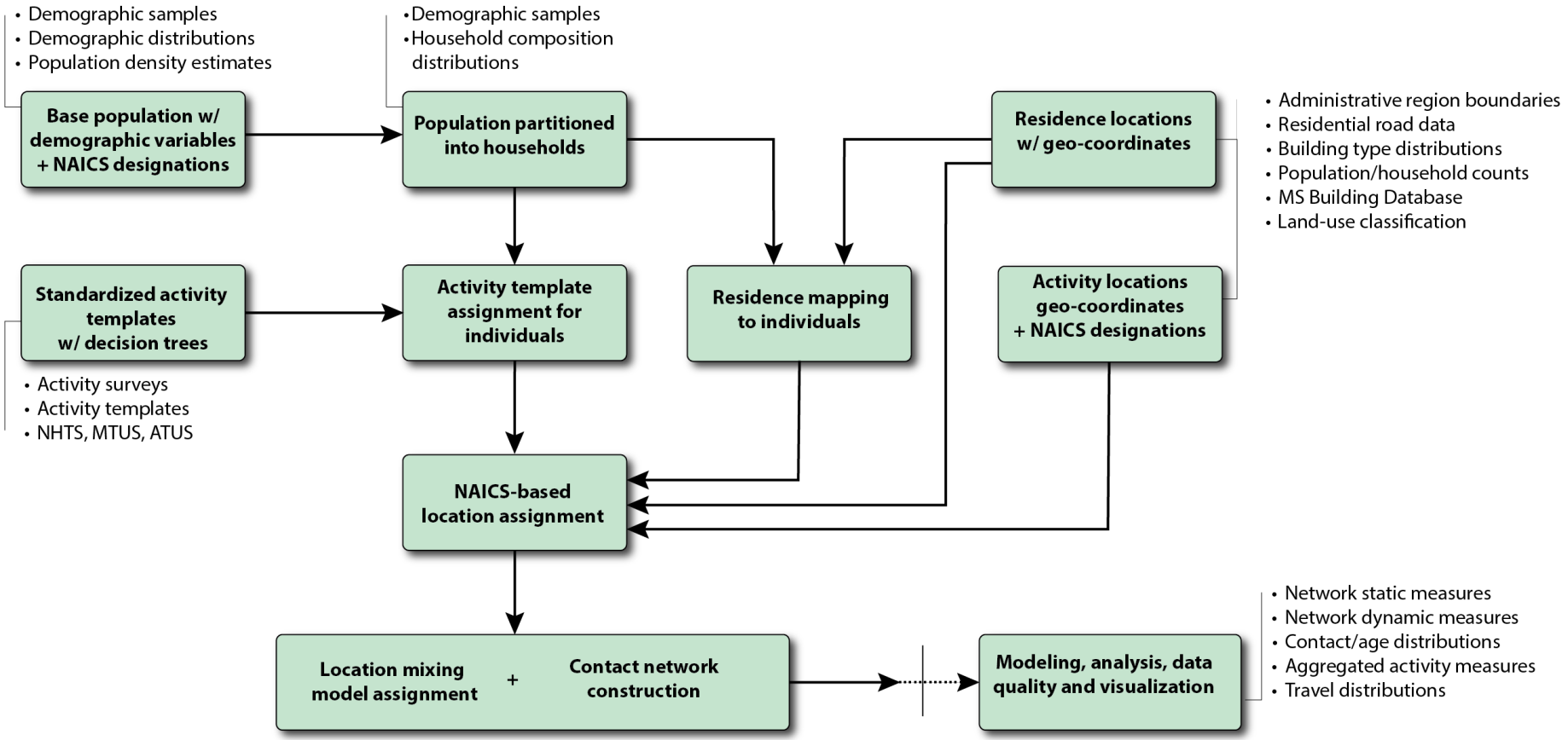
- A novel high performance computing (HPC) approach for executing epidemiological workflows
  - High resolution agent based models
  - Realistic representation of national scale social contact networks
    - 300 million nodes, 7.9 billion edges, 50 states + Washington DC
  - 5000–17,900 simulations per night
    - County level forecasts for 3140 counties
  - Split across two HPC clusters

Covid-19 Disease Model



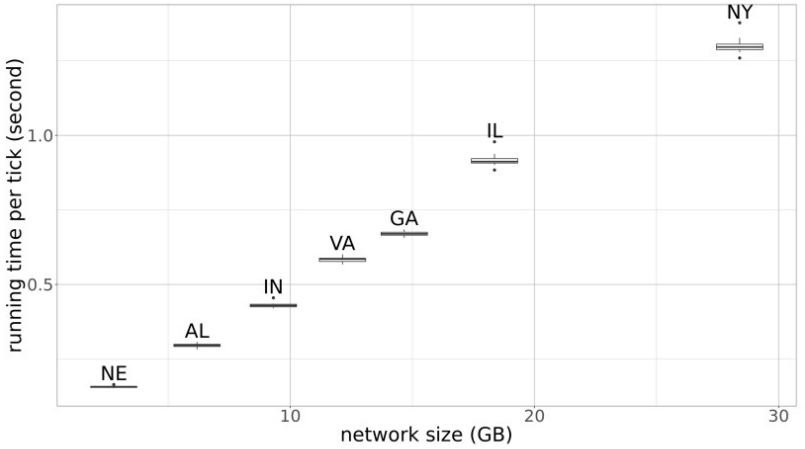
We built very detailed Covid-19 disease models
utilizing publicly available information from multiple sources

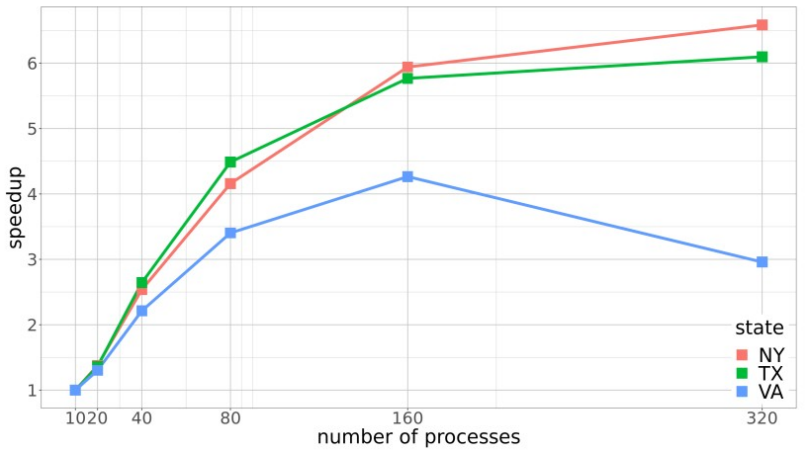
Detailed Synthetic Populations of USA



EpiHiper: A distributed epidemic simulator

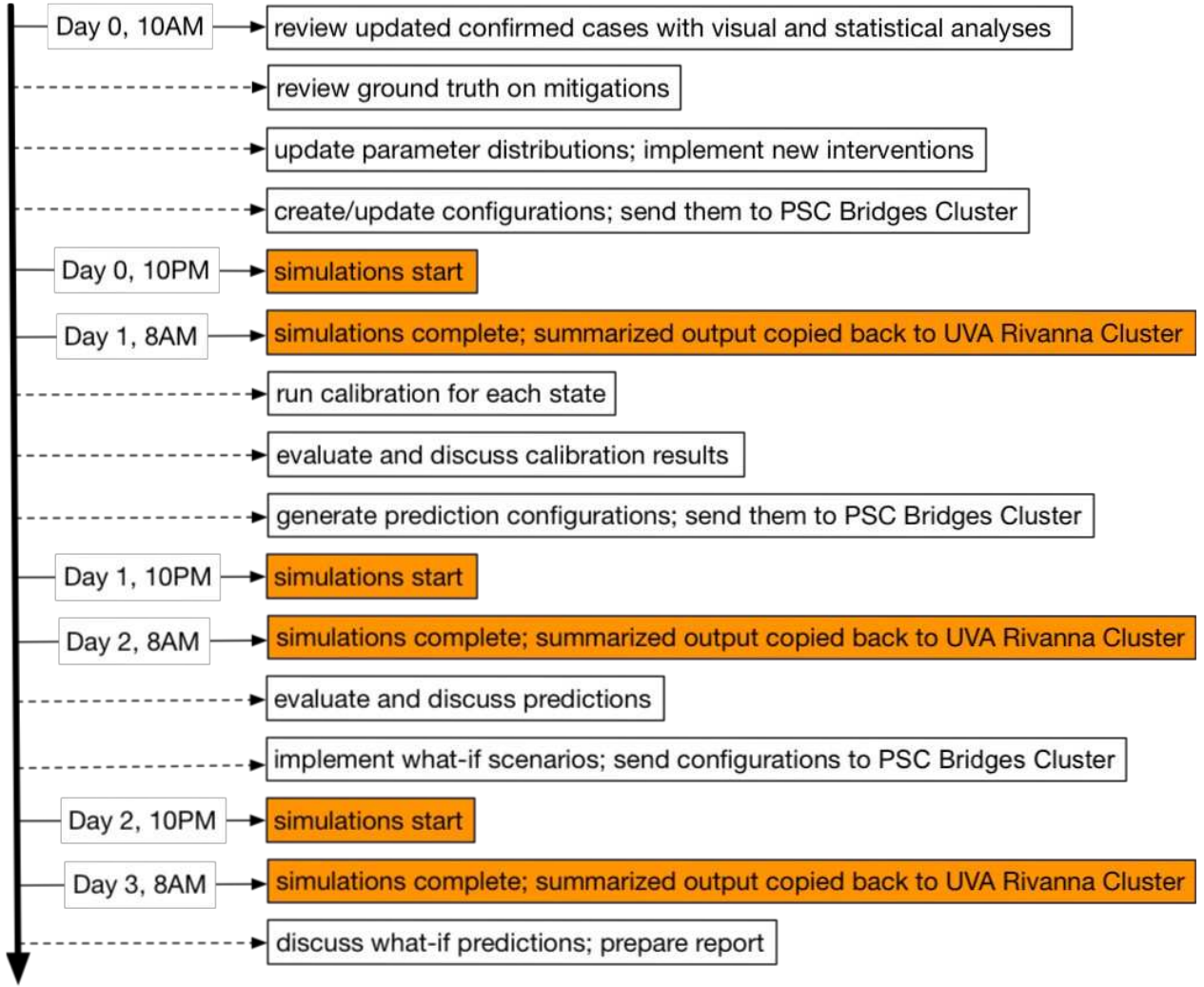
- An agent based discrete time simulation model
- Inputs: Disease model + Social contact network
- Provides custom domain specific language for programming NPI scenarios:
  - Voluntary home isolation, school closure, stay-at-home, pulsating shutdowns, partial reopening, custom vaccination schedules …
- A distributed memory program written in C++/MPI
- Contact network is partitioned onto MPI ranks
- Shared inputs served via PostgreSQL database





Rivanna and Bridges: A multi-cluster setup

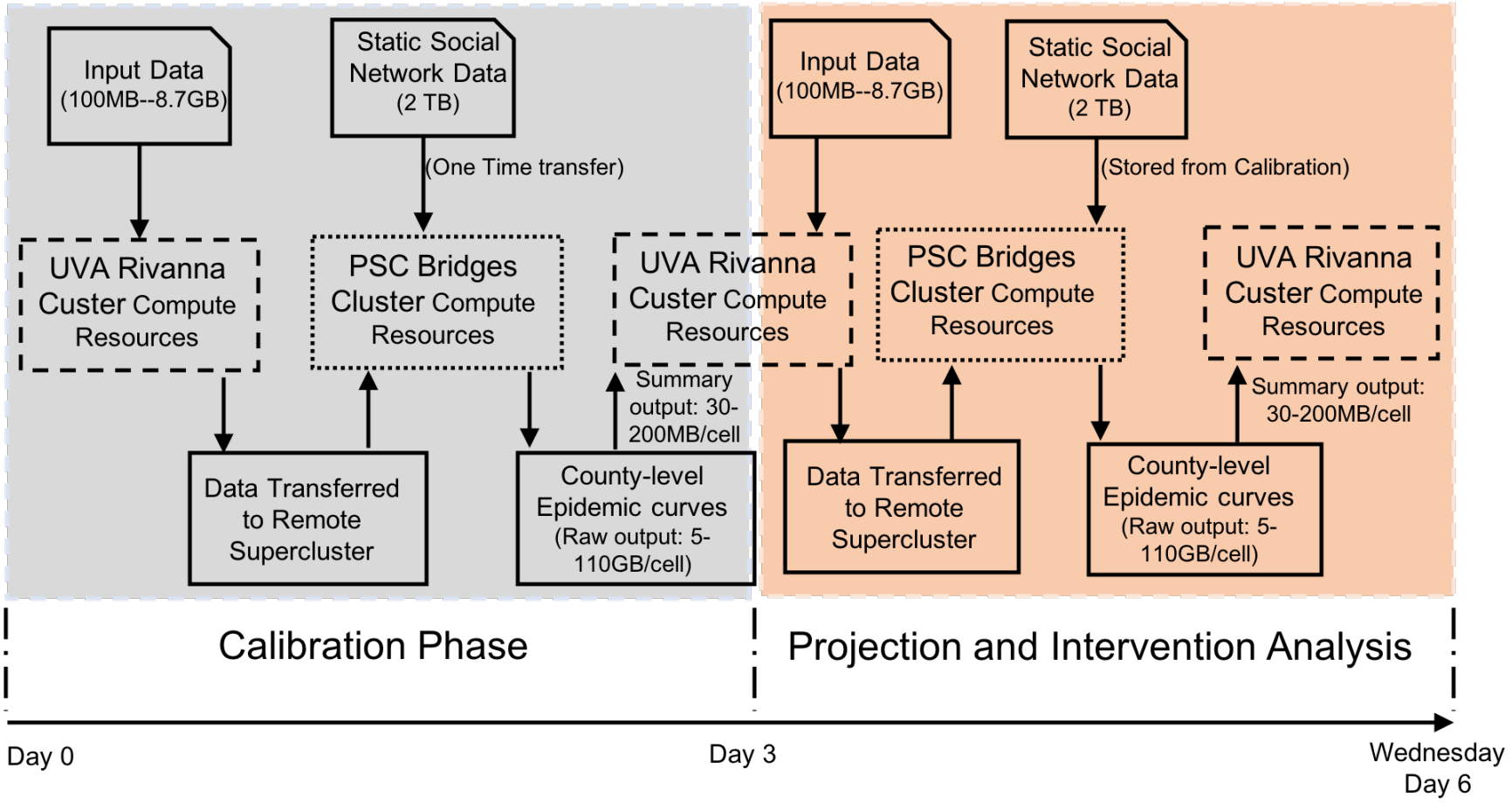
- Rivanna: Home cluster at University of Virginia
  - 50 nodes (40 CPU cores, 384 GB RAM per node)
- Bridges: Remote cluster at Pittsburgh Supercomputing Center
  - Limited access (10pm—8am every night)
  - 720 nodes (28 CPU cores, 128 GB RAM per node)
- Data transfer via Globus
- Both clusters used Slurm for scheduling



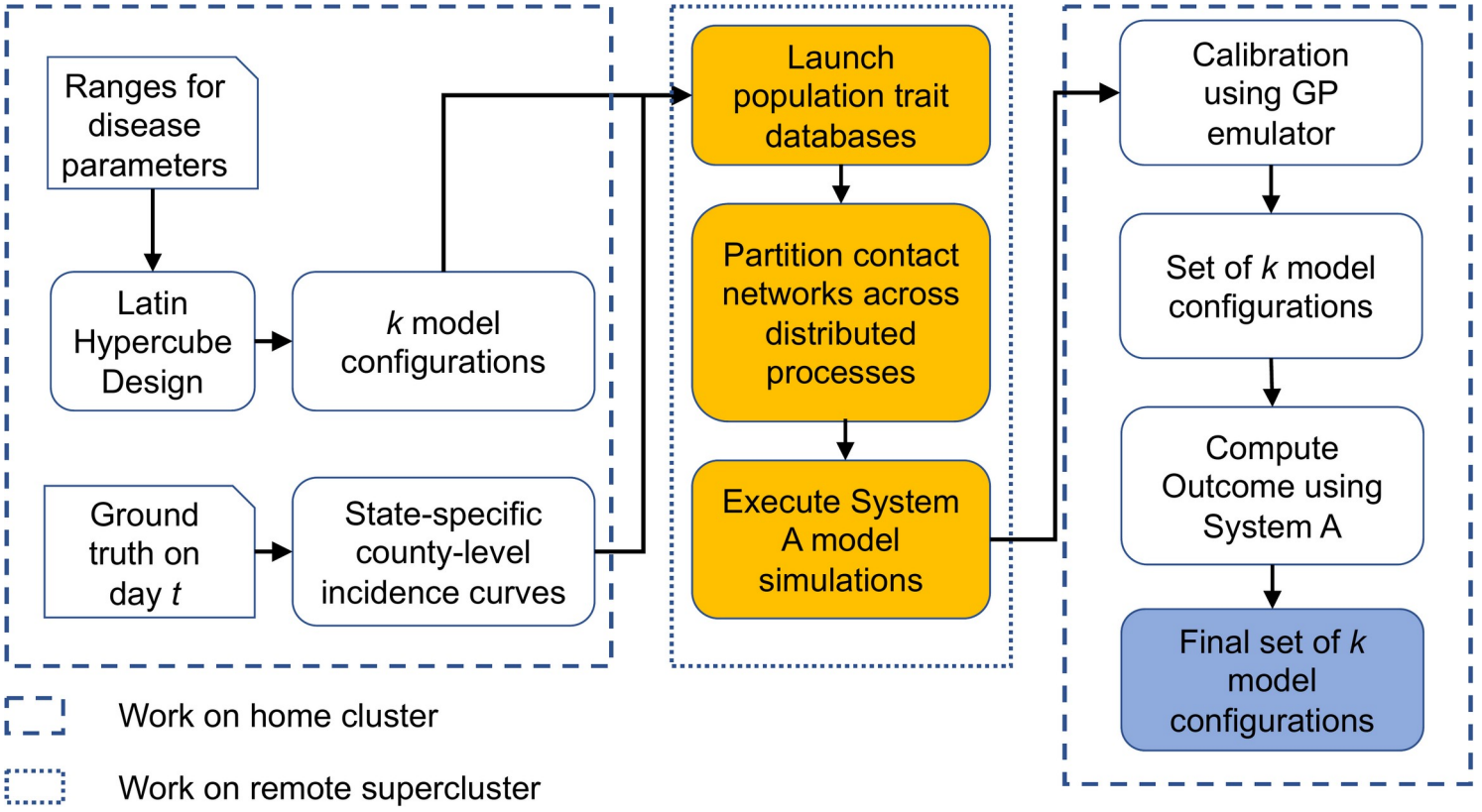
A typical timeline of tasks involving human efforts.

Orange boxes are automated.

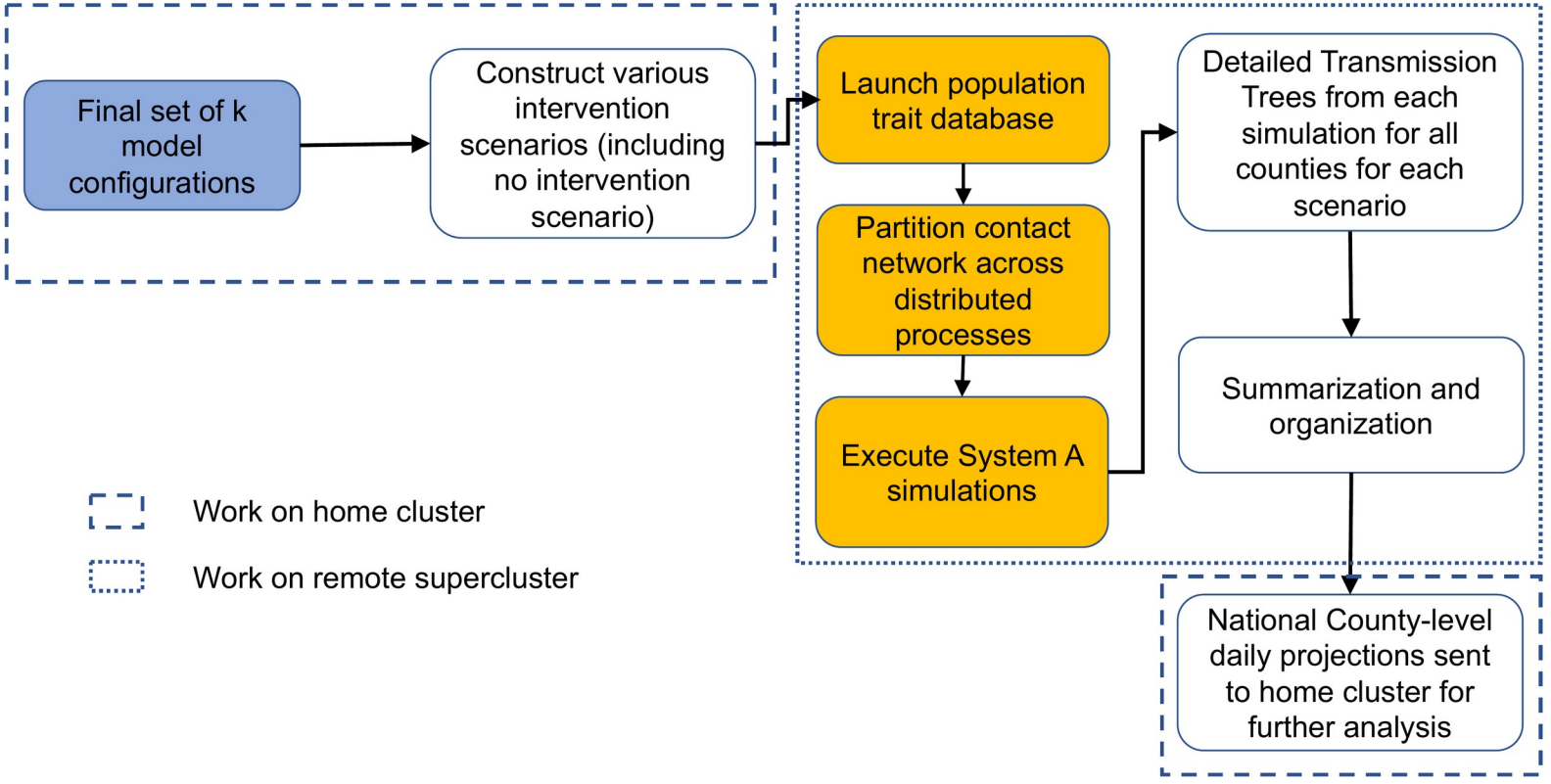
Combined workflow: End to end timeline



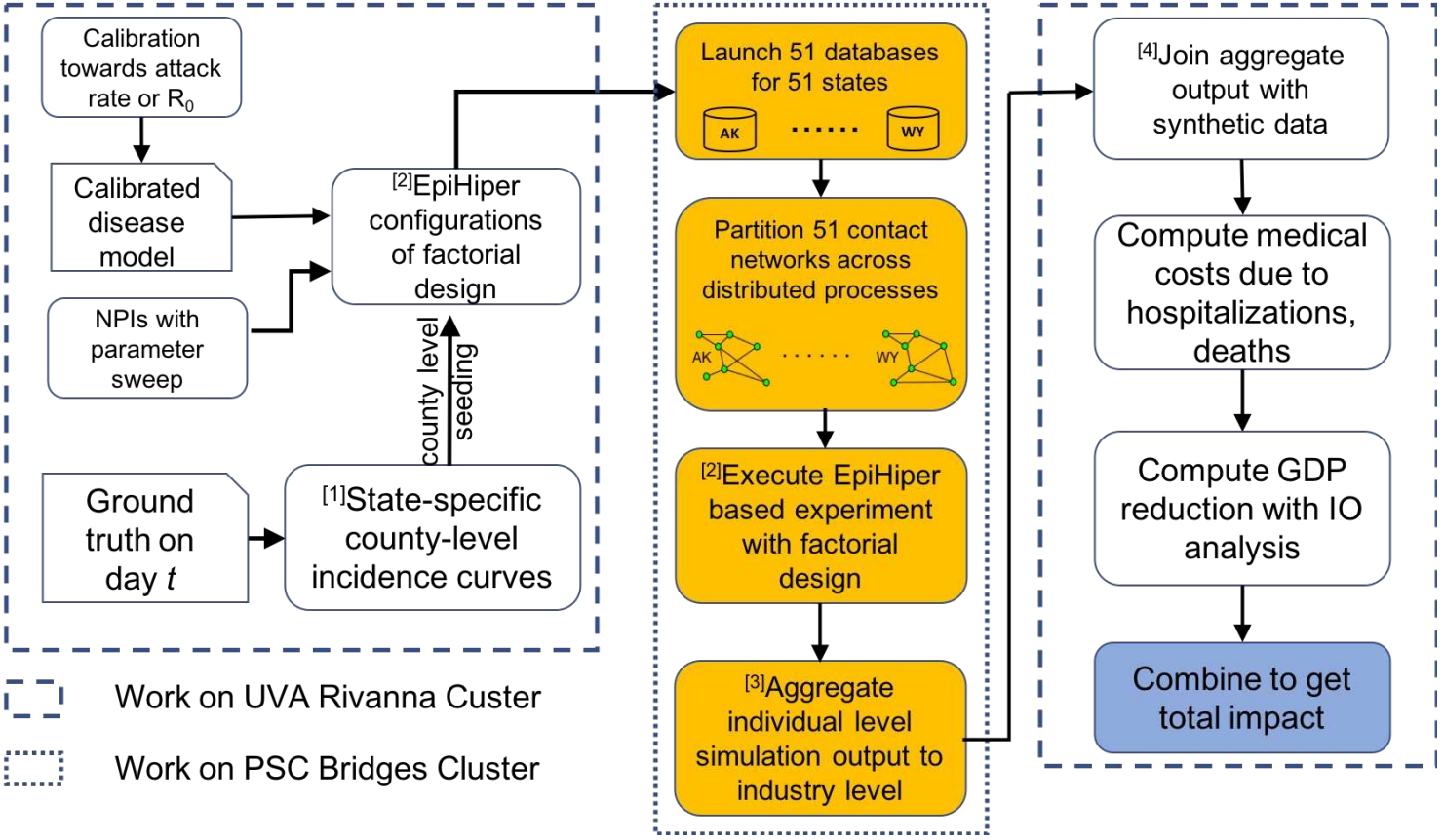
Calibration Workflow



Prediction Workflow



Economic Workflow



Case Study 1: Medical Costs of Covid-19

scientific reports

View all journals | Search | Login

Article | Open Access | Published: 28 October 2020

Medical costs of keeping the US economy open during COVID-19

Jiangzhuo Chen, Anil Vullikanti, Stefan Hoops, Henning Mortveit, Bryan Lewis, Srinivasan Venkatramanan, Wen You, Stephen Eubank, Madhav Marathe, Chris Barrett & Achla Marathe

Scientific Reports **10**, Article number: 18422 (2020) | Cite this article

2237 Accesses | **3** Citations | **6** Altmetric | Metrics

Abstract

We use an individual based model and national level epidemic simulations to estimate the medical costs of keeping the US economy open during COVID-19 pandemic under different counterfactual scenarios. We model an unmitigated scenario and 12 mitigation scenarios which differ in compliance behavior to social distancing strategies and in the duration of the

Scale of Simulations

Workflow	# Cells	# States	# Replicates	# Simulations	Raw Output	Summarized output
Calibration	300	51	1	153,00	5.0 TB	4.0 GB
Prediction	12	51	15	9180	1.0 TB	2.5 GB
Economic	12	51	1	9180	3.0 TB	5.0 GB

Large number of simulations are needed to explore the parameter space and to generate confident predictions for decision support purposes.

Workflow Orchestration

- Given:
  - A set of simulations to run
  - A set of compute nodes to run them on
- Objective:
  - Generate job ordering for Slurm
  - Minimize the total run time
- The problem can be mapped to
  - 2D Bin packing problem
  - a variant of the coloring problem (r-relaxed-coloring)

- Heuristics tested
  - First-fit decreasing time with database access constraints (FFDT-DC)
  - Next-fit decreasing time with database access constraints (NFDT-DC)
- Metric
  - System utilization
- Heuristic performance
  - FFDT-DC performs better (96.6% median utilization)
  - NFDT-DC performs worse (55.5% median utilization)

Conclusion

- We developed a novel HPC oriented workflow in order to support planning and response to pandemics such as Covid-19.
  - We used two geographically separated supercomputing facilities
  - Incorporated daily county-level surveillance data and policy data
  - National and high resolution agent-based simulations
- Real-time data driven high resolution epidemics science at national scale is indeed possible.

Thank You!