Formal scientific modeling: a case study in global health

organized by
Nina Fefferman, Tim Hosgood, and Mary Lou Zeeman

Workshop Summary

Executive summary

The workshop "Formal Scientific Modeling: A Case Study in Global Health" convened researchers and practitioners from mathematics, epidemiology, computer science, public health, behavioral science, philosophy of science, and software engineering to examine how formal mathematical structures can improve scientific collaboration, model transparency, and decision support in complex global health settings. Participants explored the possibility that many persistent barriers to interdisciplinary modeling arise not merely from disciplinary differences in terminology, but from deeper incompatibilities in the formal structures through which assumptions, evidence, uncertainty, and causal mechanisms are represented.

The workshop focused particularly on the challenge of creating modeling systems capable of supporting collaborative reasoning across heterogeneous forms of expertise. Existing epidemiological and public health modeling approaches frequently require researchers to manually reconcile incompatible assumptions, data structures, scales of analysis, and computational frameworks. This process can be labor-intensive, opaque, and difficult to reproduce or audit. To address this, participants investigated how category-theoretic and compositional approaches to modeling could provide formal mechanisms for integrating diverse model types while preserving interpretability and scientific rigor. Rather than treating models solely as isolated predictive tools, participants explored the possibility of representing models as formally composable objects whose assumptions, interfaces, and transformations could themselves be mathematically specified. This perspective has implications not only for public health modeling, but also for scientific communication, reproducibility, software infrastructure, and collaborative decision-making more broadly.

Across five days of intensive collaboration, the workshop emphasized small-group problem solving, iterative framework design, and the production of concrete research artifacts. Working groups developed candidate formal structures, outlined future software architectures, identified unresolved conceptual challenges, and articulated priorities for future collaboration. The meeting generated both technical and organizational momentum toward a broader research programme focused on formal scientific modeling as an infrastructure for interdisciplinary science.

Context

The increasing complexity of global health challenges has exposed limitations in many existing scientific modeling workflows. Pandemic preparedness, antimicrobial resistance, environmental health interactions, and emerging zoonotic diseases all require integration of biological, ecological, behavioral, logistical, and sociopolitical processes that are often modeled independently using incompatible methodologies. While interdisciplinary collaboration is now widely recognized as essential, the mechanisms by which researchers combine models, assumptions, and evidence across domains remain underdeveloped. Distinct disciplines often encode causality, uncertainty, time scales, validation standards, and intervention effects differently. As a result, integrating models across domains can introduce ambiguity regarding interpretation, comparability, and reliability. The workshop therefore approached "modeling" itself as an object worthy of formal analysis.

The choice of global health as a motivating case study reflected both the urgency of pandemic preparedness and the unusually rich modeling ecosystem already present within epidemiology and public health. Infectious disease systems require simultaneous reasoning across biological transmission processes, human behavior, surveillance systems, intervention logistics, and policy constraints, making them an especially demanding environment in which to test new formal approaches.

Goals

The workshop was organized around several interrelated objectives intended to guide both immediate collaborative work and longer-term research development.

A primary goal was to develop a shared conceptual vocabulary for discussing formal scientific modeling across disciplinary boundaries. Participants sought to identify common structural elements underlying otherwise disparate modeling traditions and to determine where category-theoretic approaches could clarify relationships among models, assumptions, and data.

A second objective was the identification and prioritization of open research problems. Participants worked collaboratively to compile challenge areas requiring further theoretical, computational, or organizational development. These included questions related to compositional semantics, uncertainty propagation across linked models, representation of social and behavioral systems, software interoperability, and mechanisms for model validation in heterogeneous modeling ecosystems.

The workshop also emphasized practical computational pathways. Participants explored how formal model structures might support future software tools enabling collaborative model construction, modularity, automated reasoning about assumptions, and improved transparency in scientific workflows. Discussions considered both foundational theoretical requirements and pragmatic implementation challenges.

Several recurring questions guided discussions throughout the workshop:

- How can formal mathematical structures support interdisciplinary scientific reasoning without oversimplifying disciplinary nuance?
- What types of compositional relationships among models are most useful for real-world public health decision making?
- How can formal approaches improve transparency regarding assumptions, uncertainty, and model limitations?
- What computational infrastructures are necessary to make formal scientific modeling accessible and useful to applied researchers and policymakers?

These questions helped structure both plenary discussions and smaller working-group activities throughout the week.

Format

Consistent with the collaborative format characteristic of AIM workshops, the meeting emphasized sustained interaction among participants rather than formal presentations. Short introductory talks established common conceptual foundations and identified major challenge areas, after which most workshop time was devoted to collaborative working sessions.

Participants organized into flexible working groups focused on themes such as compositional model semantics, formal representations of epidemiological systems, software architecture and interoperability, uncertainty and validation frameworks, and applications to pandemic response decision-making. Groups frequently reconfigured throughout the week as participants moved among related problem areas and synthesized emerging ideas, and researchers with expertise in pure mathematics, epidemiology, philosophy, software engineering, and public health were able to iteratively refine terminology, assumptions, and objectives in real time.

On Wednesday a practice challenge was issued to participants. An imagined scenario was presented, in which cases of an infectious disease had been reported in Pasadena and three different research groups were tasked with developing models to address the following prompts:

(1) There is an effective post-infection treatment, but it costs money and there are economic limitations. Monitor the outbreak and decide what quantity of treatment medicines to order.
(2) There is a vaccine, but it costs money and there are economic limitations. Who should be vaccinated? Develop an effective strategy combining vaccination and education campaigns to minimize consequences of the outbreak.
(3) There is funding, but it is limited. How should it be allocated among vaccination, treatment, and education to minimize consequences of the outbreak?

This exercise became a highlight of the week. It was extremely effective in conveying a variety of implementation challenges for such interdisciplinary modeling, and how those challenges translate to research questions in applied category theory. Of particular interest, for example, was how it helped to clarify the relationship between the communication channels needed among modelers across the disciplines, as well as aspects of composition in formal scientific modeling.

Throughout, the workshop environment encouraged exploratory thinking and intellectual risk-taking. Participants repeatedly returned to foundational questions regarding what constitutes a "model", how models should be compared or composed, and which aspects of scientific reasoning are most amenable to formalization. This process was essential for building shared understanding across disciplinary perspectives that have, traditionally, been intellectually isolated from each other

Planned outputs

The workshop generated a series of planned deliverables intended to support both immediate dissemination and longer-term collaborative research development.

One major deliverable was the creation of a structured problem list documenting key open questions in formal scientific modeling. Participants categorized problems according to conceptual, mathematical, computational, and organizational dimensions, with the goal of establishing a shared research agenda for future collaborations and funding opportunities.

A second set of deliverables focused on formal modeling frameworks and representational structures. Participants drafted candidate approaches for expressing relationships among models, assumptions, datasets, and interventions using compositional mathematical formalisms. These efforts included preliminary discussions of interfaces capable of linking traditionally distinct modeling paradigms, including systems of differential equations, agent-based models, network models, and behavioral frameworks.

The workshop also produced plans for future software development efforts. Participants outlined desired capabilities for computational tools that could support modular model construction, metadata-aware composition, collaborative editing, and formal tracking of assumptions and transformations across linked models.

At the heart of all the discussions was the collaboration between category-theory based software developers and application based modelers. Groups for ongoing cross-disciplinary collaboration were established on topics including:

- Survey of types of model composition: examples and open questions.
- Composing models with different time scales.
- Composition types and energy flows in physiologically based demographic models (PBDMs).
- Developing diagrams for visualizing composition in ODEs and PDEs
- Calculating R_0 via flow network migration.
- Patch / multi-layer / stratified models from a category-theoretic perspective.
- Educational roles for the CatColab software suite.

Summary

The workshop demonstrated both the promise and the difficulty of developing formal approaches to interdisciplinary scientific modeling. Participants identified substantial opportunities for improving transparency, interoperability, and collaborative reasoning in global health modeling systems through the use of compositional and category-theoretic frameworks. At the same time, discussions highlighted important unresolved challenges involving scalability, usability, semantics, validation, and communication across disciplines. Perhaps most importantly, the meeting established a growing interdisciplinary community organized around the shared objective of formalizing scientific modeling practices in ways that remain both mathematically rigorous and practically useful. The collaborative environment fostered new intellectual connections among participants who would not ordinarily share common methodological or disciplinary frameworks.

Future work will focus on refining formal structures, expanding computational implementations, and testing proposed approaches against real-world global health modeling challenges. Participants also identified the need for broader engagement with applied practitioners, policymakers, and scientific software developers to ensure that future tools remain responsive to operational realities while preserving explainable theoretical foundations.

The workshop served as a foundational step toward a broader long-term research program centered on formal scientific modeling as a new infrastructure tool for interdisciplinary science.