Really Big Data from HPC-enabled Biomedical Agent-based Modeling: Pathways to Precision Medicine

Chase Cockrell and Gary An, Department of Surgery, University of Chicago

"Big Data" in Biomedicine: The Real Problem

BIOMEDICINE IS TOO DATA POOR FOR SUSTAINED "BIG DATA" APPROACHES => MACHINE LEARNING + EVOLUTIONARY COMPUTING

- Model from Physical Sciences => Integration of Simulated Data into Investigatory Workflow
- Contra 1: Our Models not "Good Enough"
- Reply 1: *Criteria for "Good Enough" not suitable to use of M&S in Biomedicine*
- **PROPOSED: Use of sufficiently complex MSM to identify investigatory boundaries to constrain/ direct Experimental/Clinical Investigation**
- EXAMPLE: How controllable is Sepsis, and what would it take to deliver Precision Medicine => "Right Drug, Right Patient, Right Time"

Embracing Heterogeneity: Biology as Parameter Space

BIOLOGY = HETEROGENEITY OF PHENOTYPE FROM COMMON STRUCTURE

CONTRADICTION:
- Experimental Biology => Reduce Outliers => "brittle results"
- Biology => Science of Outliers (Need for Evolution to work)

PROBLEM:
- Current standards of Calibration, Validation and Prediction in M&S drawn from Physical Sciences w/ binding Natural Laws
- Biology doesn't have corresponding binding Natural Laws

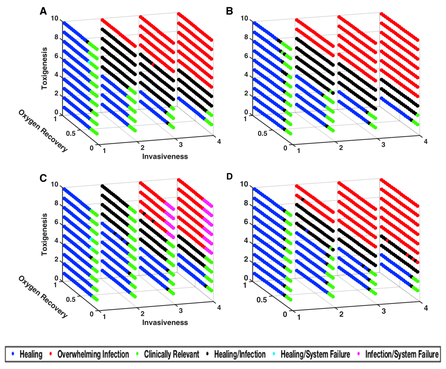
RESULT: **Models of Experimental Systems that don't transfer/ link across systems/populations => "Crisis of Reproducibility"**

CONCLUSION: Current approach to evaluation Biomed M&S **not appropriate** to characterize Biology

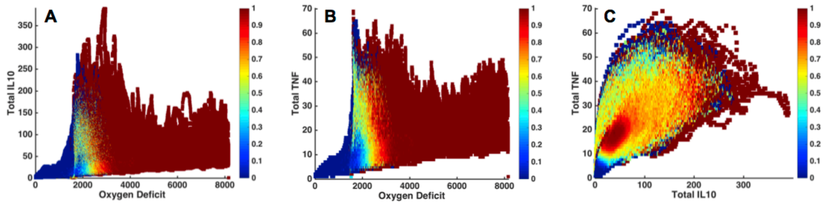
SOLUTION: ***Use M&S to characterize the "denominator" of biosystem heterogeneity => Parameter Space Exploration (vs. Parameter Fitting)***

EXAMPLE: Space of Sepsis
- No effective therapies
- Clinical Data Sparse
- Exp Models don't Map

HPC Simulation of Sepsis ABM
- ➢ 74 million "Patients"
- ➢ "Plausible" Parameter Sets
- ➢ "Denominator" of Sepsis

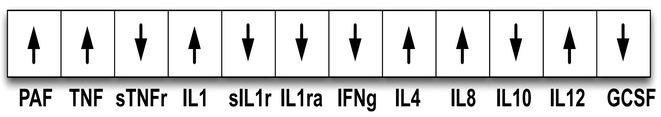


Predictability from State ID (Data)? => **Nope.** Random Dynamical System with Probabllistic Basins of Attraction (PBoAs)



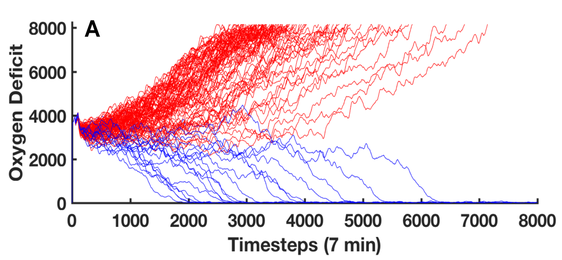
Determining Controllability: GA for Multimodal Control

Genetic Algorithm (GA):
- Efficiently search a large parameter space using evolutionary principles (Fitness/Inheritance/ Mutation)
- 4 inhibition strengths, 4 augmentation strengths, 300 billion combinations/sequential intervention, 10^{91} for 8 sequential interventions

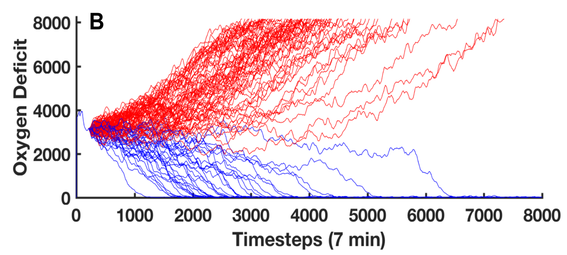


- Run simulation with intervention
- Test intervention for "fitness"
- Top 50% fittest individuals breed w/ mutation $\mathbb{P}$
- Repeat until convergence to small set of solutions
- GA trained on 1 individual, RNG was reseeded at start of Intervention

A: single parameter set, 100 stoch replicates

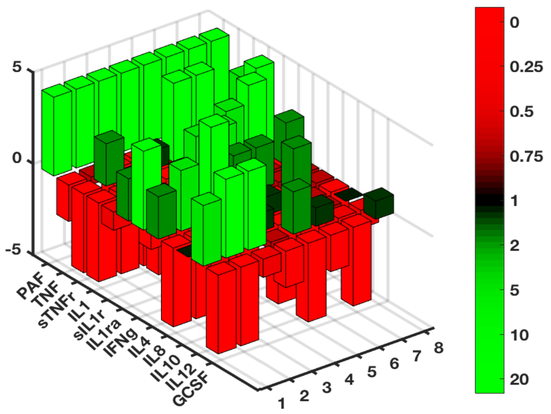


B: single parameter set, single individual, reseed at intervention start.

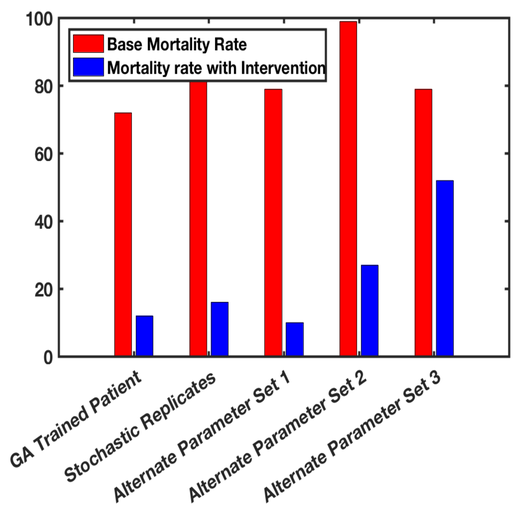


Results: Effective Control for Most, but Lots of Levers

- GA-derived intervention significantly reduces mortality rate in training set and generalizes well to several alternate parameter sets

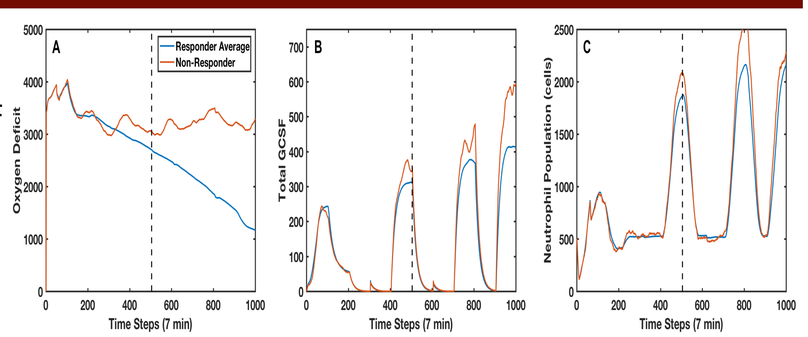


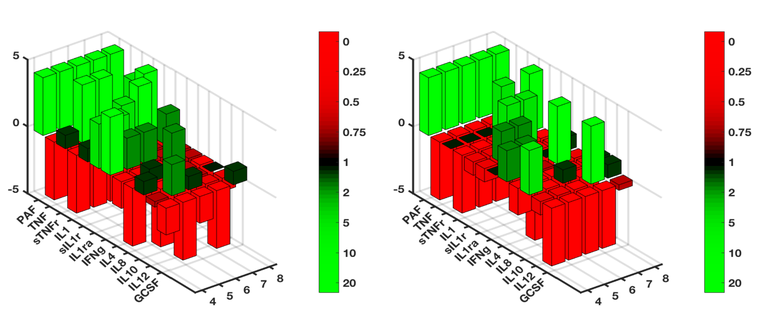
- Poor performance on Alternate Set 3 => due to fixed-length interventions and inability to adapt to non-responders in real-time?



Alternate Treatment for Intervention Non-Responders

- This patient exhibits an enhanced response to GCSF stimulation
- New GA experiment upon significant deviation from the average response leads to personalized intervention sequence, saves non-responder
- Original intervention is shown on the left; Alternate intervention sequence shown on right





Conclusion: Proxy MSMs to Define Epistemic Boundaries

1st Order Conclusions
- Stochastic, dynamical immune system is "controllable"
- Calculated interventions are generalizable (to some degree)
- Personalized interventions are necessary to increase level of success
- GA => too many fixed parameters (intervention length, number of interventions in sequence) => alternate machine learning techniques to be more efficient (deep reinforcement learning)

*N*th Order Conclusions
- Dynamic Systems require Dynamic Control => "True" Precision Medicine is *Hard*
- Real-world data too sparse to use the methods needed => ***Need scale of data generated by Proxy MSM Simulations***
- Proxy MSMs initially used to establish the scope of the problems/solutions => Avoid cul-de-sacs and dead-ends
- Sufficiency of MSMs evaluated by breadth of plausible phenotype coverage =><= to accuracy/precision of prediction

Acknowledgements: This work supported by HPC resources of the National Energy Research Scientific Computing (NERSC) Center, a DoE Office of Science User Facility under Contract No. DE-AC02-05CH1123 and the University of Chicago Computation Institute (Beagle2). Both Drs. An and Cockrell are supported under grant 1S10OD018495-01 from the National Institutes of Health, as well as by funds from Lawrence Livermore National Laboratory under Award #B616283.