The Reference Model Interface with ClinicalTrials.Gov

ClinicalTrials.Gov is an NIH project that accumulates clinical trial data

Jacob Barhak Ph.D.

http://sites.google.com/site/jacobbarhak/

IMAG Multiscale Modeling (MSM) Consortium Meeting
22-24 March 2017

In a Nutshell: The Reference Model, that accumulates other models and validates them against observed outcomes, can now import data from ClinicalTrials.Gov

This work builds upon a decade of development with key publications in the following list:

[1] J. Barhak, The Reference Model for Disease Progression uses MIST to find data fitness. PyData Silicon Valley 2014 held at Facebook Headquarters:
Presentation: http://sites.google.com/site/jacobbarhak/home/PyData_SV_2014_Upload_2014_05_02.pptx
Video: https://www.youtube.com/watch?v=vyvxiljc5vA

[2] J. Barhak, A. Garrett, Population Generation from Statistics Using Genetic Algorithms with MIST + INSPYRED. MODSIM World 2014, April 15 - 17, Hampton Roads Convention Center in Hampton, VA.
Paper: http://sites.google.com/site/jacobbarhak/home/MODSIM2014_MIST_INSPYRED_Paper_Submit_2014_03_10.pdf
Presentation: http://sites.google.com/site/jacobbarhak/home/MODSIM_World_2014_Submit_2014_04_11.pptx

[3] J. Barhak, Object Oriented Population Generation, MODSIM world 2015. 31 Mar – 2 Apr, Virginia Beach Convention Center, Virginia Beach, VA.
Paper: http://modsimworld.org/papers/2015/Object_Oriented_Population_Generation.pdf
Presentation: http://sites.google.com/site/jacobbarhak/home/MODSIM2015_Submit_Jacob_Barhak_2015_03_29.pptx

[4] J. Barhak, The Reference Model for Disease Progression Combines Disease Models. I/IITSEC 2016 28 Nov – 2 Dec Orlando Florida.
Paper: http://www.iitsecdocs.com/volumes/2016
Presentation: http://sites.google.com/site/jacobbarhak/home/IITSEC2016_Upload_2016_11_05.pptx

What is novel?

ClinicalTrials.Gov holds data of Clinical Trials from all around the world. On 12-Feb-2017 there were 236,687 trials registered in this database where 24,251 reported results. This database keeps on growing. The Reference Model new interface with this database provides an essential modeling capability that is traceable and reproducible.
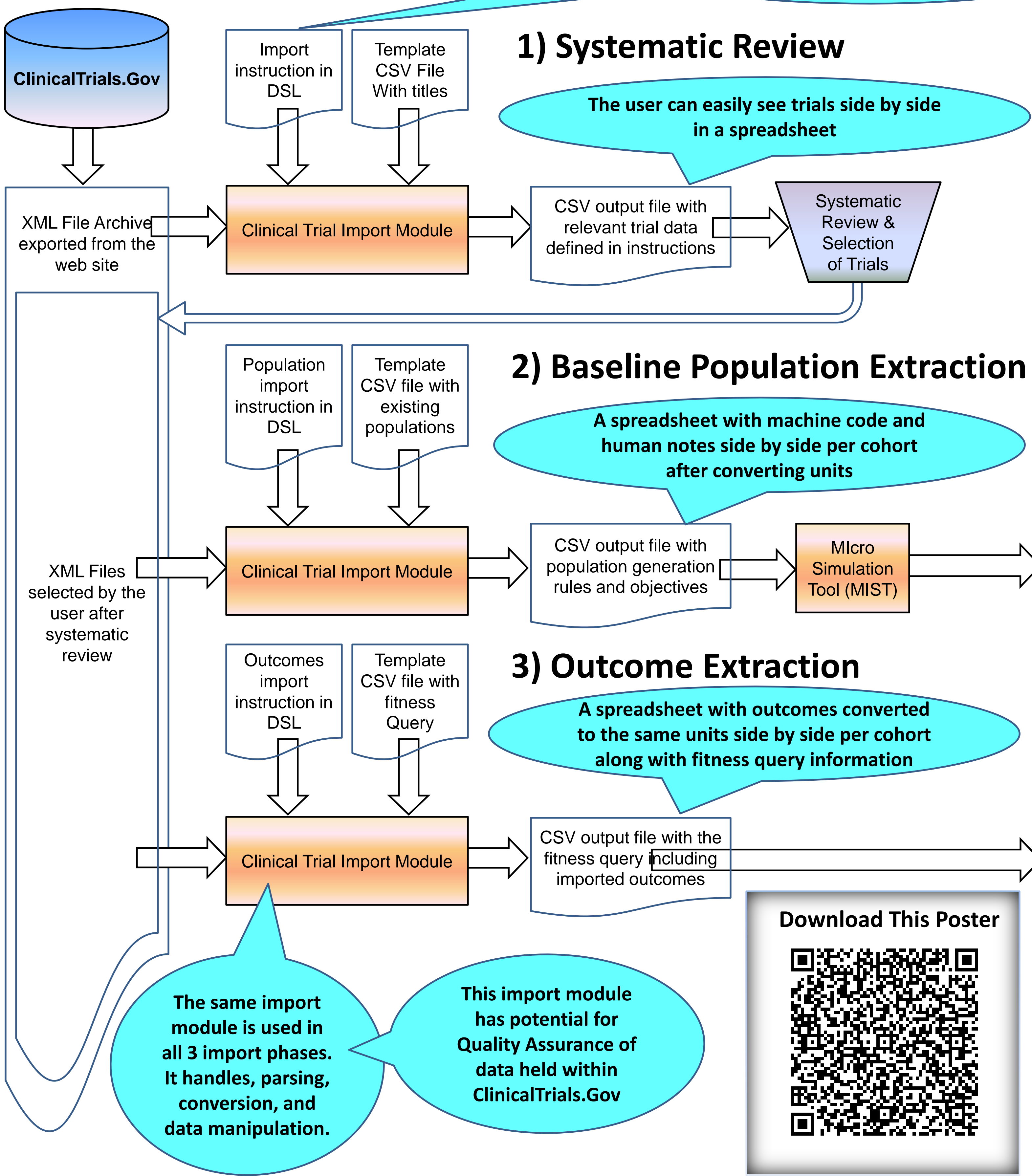
Key Points

- Ensemble model
- Accumulates knowledge from:
  - Existing models
  - Observed outcomes
- Modeling at the population level
- Simulation at the individual level
- Focuses on summary data
  - Avoids individual data restrictions
  - Larger merged population base
- Flexible Import from ClinicalTrials.Gov
- Applicable for other disease processes
- Traceable and reproducible

Abstract:

The Reference Model for Disease Progression allows merging information from multiple existing disease models and from observed results from clinical trials. This is a powerful way of accumulating clinical knowledge, both cognitive knowledge embedded within computational models, and evidence based knowledge provided by clinical studies. The Reference Model development started half a decade ago and accumulated observed data published in clinical trial reports such as baseline demographics and outcomes. However, entry of such data was tedious and prone to errors since it involved copying printed data manually and transferring it into programmatic modeling code.

Detached from this project, the NIH started accumulating clinical trial data in the last few years in the ClinicalTrials.Gov web site. The web site provides a registry for clinical trials. Recently, the registry was adapted to hold clinical trial results that contain essential information for modeling. The registry web site was designed to be easy to interface with, and it can export data using XML scheme – a widely used data transfer format.

The Reference Model can now interface with ClinicalTrials.Gov allowing it to convert the registered knowledge into modeling code. The work will describe the basics of this interface that will allow the model to accumulate more clinical information in the future to better understand disease progression.
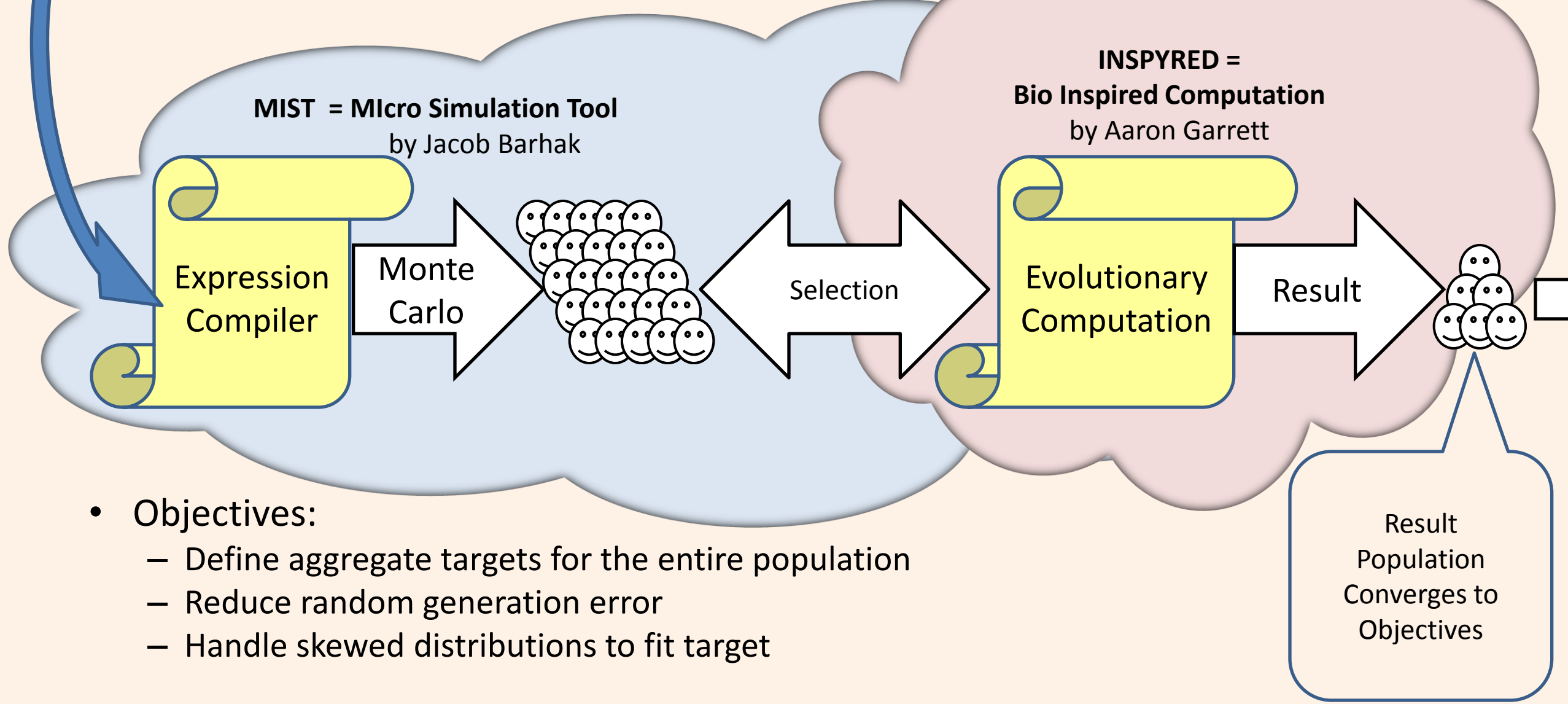
3 Phases of Import

DSL = Domain Specific Language that is dedicated to extract data from ClinicalTrials.Gov

1) Systematic Review

The user can easily see trials side by side in a spreadsheet

ClinicalTrials.Gov

Import instruction in DSL

Template CSV File With titles

XML File Archive exported from the web site

Clinical Trial Import Module

CSV output file with relevant trial data defined in instructions

Systematic Review & Selection of Trials

2) Baseline Population Extraction

A spreadsheet with machine code and human notes side by side per cohort after converting units

Population import instruction in DSL

Template CSV file with existing populations

XML Files selected by the user after systematic review

Clinical Trial Import Module

CSV output file with population generation rules and objectives

MIcro Simulation Tool (MIST)

3) Outcome Extraction

A spreadsheet with outcomes converted to the same units side by side per cohort along with fitness query information

Outcomes import instruction in DSL

Template CSV file with fitness Query

Clinical Trial Import Module

CSV output file with the fitness query including imported outcomes

Download This Poster

The same import module is used in all 3 import phases. It handles, parsing, conversion, and data manipulation.

This import module has potential for Quality Assurance of data held within ClinicalTrials.Gov
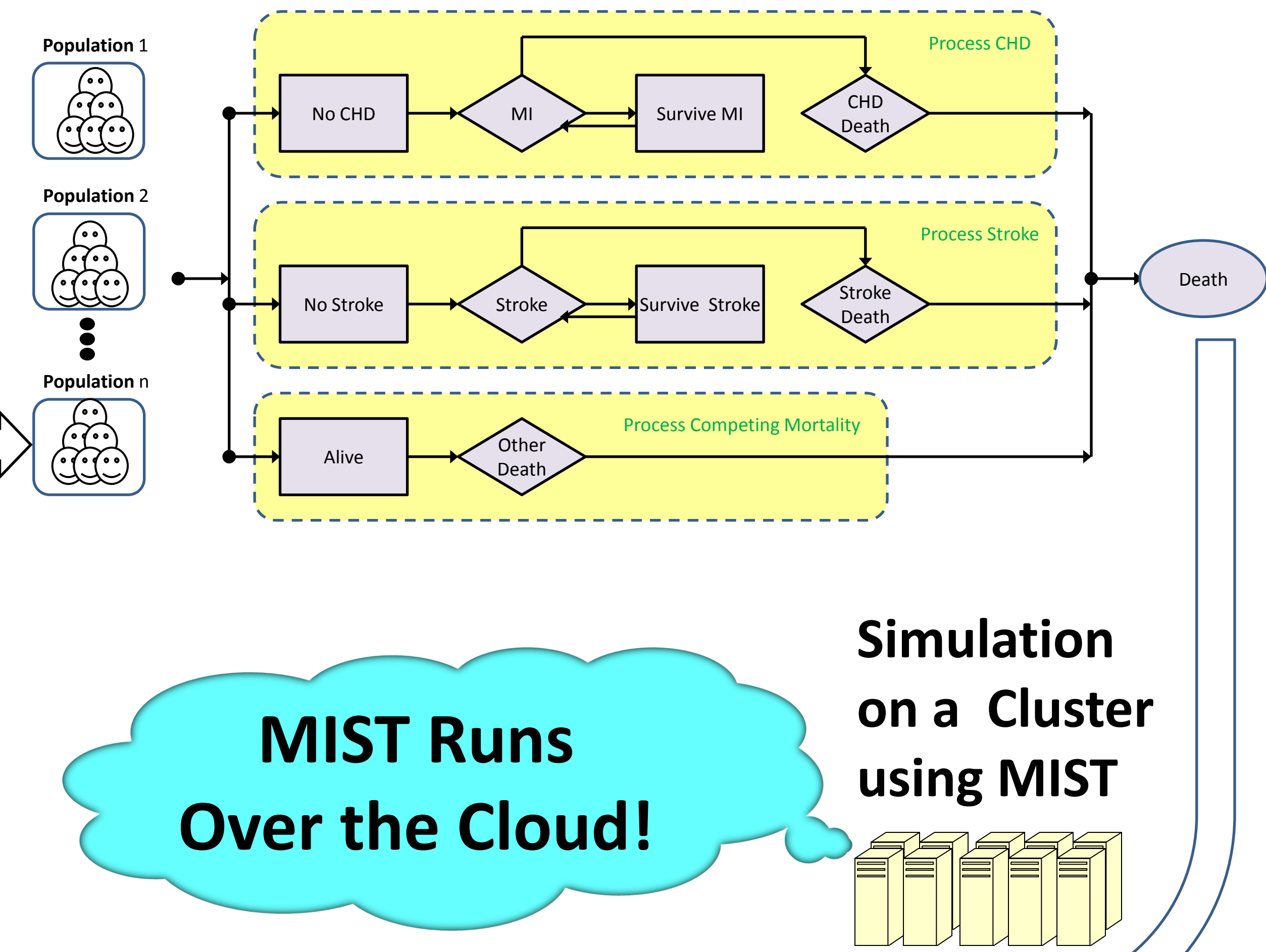
Population Generation Approach

- Generation Expressions:
  - Define how to generate a single individual
  - Test if individual fits the inclusion/exclusion criteria
 - Define ties and correlations between characteristics

MIST = MIcro Simulation Tool by Jacob Barhak

Expression Compiler → Monte Carlo → Selection → Evolutionary Computation → Result

INSPYRED = Bio Inspired Computation by Aaron Garrett

Result Population Converges to Objectives

- Objectives:
  - Define aggregate targets for the entire population
  - Reduce random generation error
  - Handle skewed distributions to fit target

The Reference Model

Population 1
Population 2
...
Population n

Process CHD: No CHD → MI → Survive MI → CHD Death
Process Stroke: No Stroke → Stroke → Survive Stroke → Stroke Death
Process Competing Mortality: Alive → Other Death
Death

MIST Runs Over the Cloud!

Simulation on a Cluster using MIST

Validation & Optimization Using an Assumption Engine

- The Assumption Engine allows us to "throw" assumptions at it
  - A model is an assumption!

- The Assumption engine figures out which assumptions work well together to fit observed outcomes data
  - Points to significant models
  - Rejects incompatible models

- Optimizes components in parallel using Gradient Descent
  - Components can cooperate
  - Components can compete = multiple initial guesses

- Define constraints such as:
  - Equations contribute partial information
  - Non negative contribution
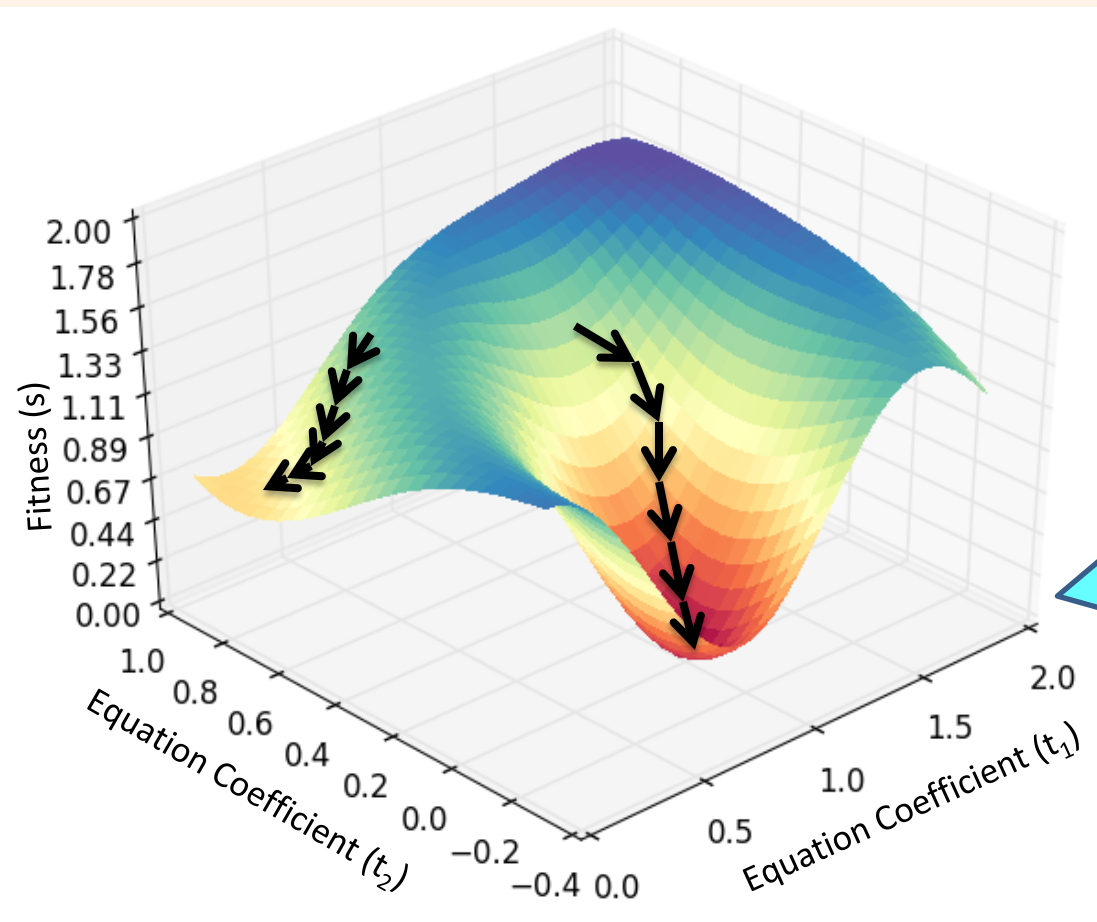
Final result:
1. The model mixture that best fits:
   - The outcomes
   - The query
   - Modeling assumption
2. The fitness/error of each cohort