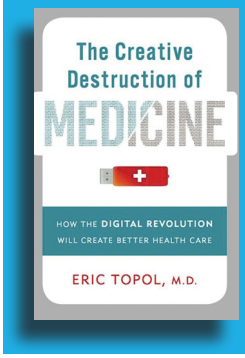
DataSpeaks

Cutting Edge /
Drug Discovery & Development
in Michigan
September 9-10, 2015
East Lansing, MI
Presented by MichBio & the University Research Corridor

Jump-Start Pharmaceutical Industry Productivity

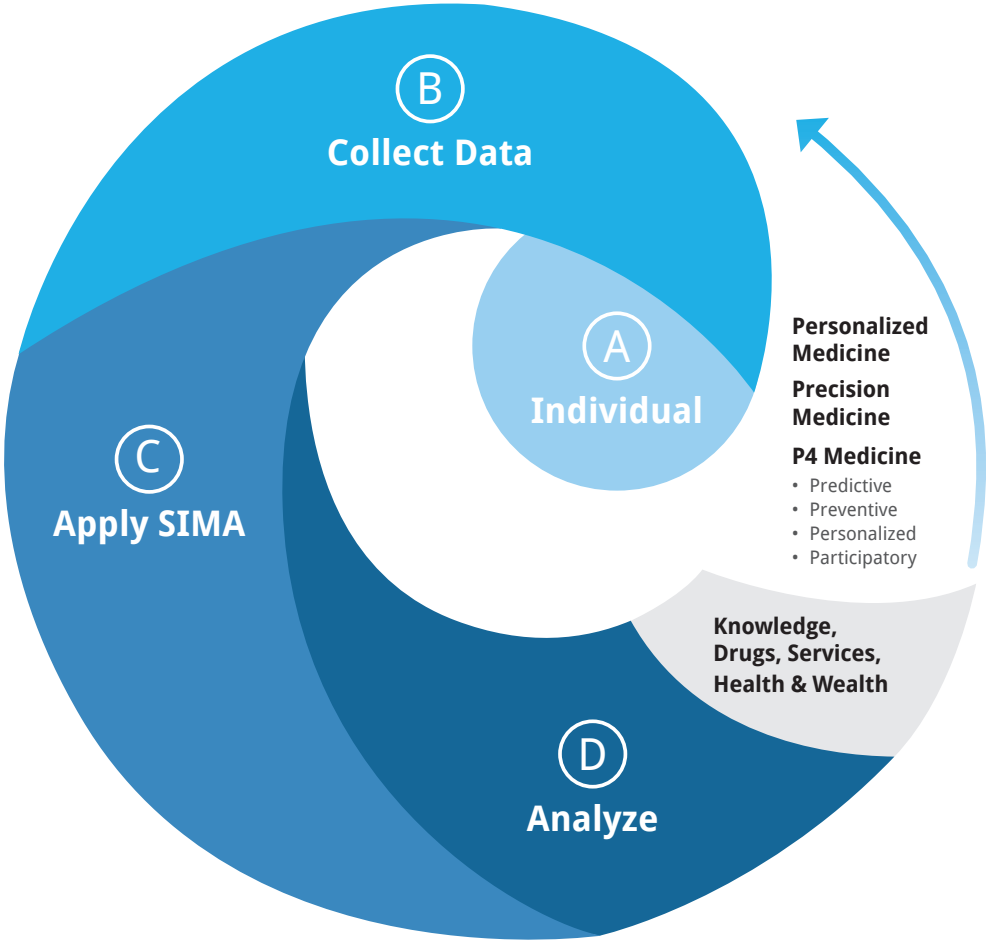
with the Science of Individuality Measurement Algorithm (SIMA)

Curtis A. Bagne, Ph.D. President and Founder, DataSpeaks, Inc.



Dr. Eric Topol wrote "how the creative destruction of medicine can and will be achieved" by "The Science of Individuality" (page 228).

SIMA (Science of Individuality Measurement Algorithm) from DataSpeaks, Inc. helps enable The Science of Individuality. This poster focuses on applications of SIMA to jump-start pharmaceutical industry productivity. SIMA often applies for patients with heterogeneous complex chronic disorders such as pain, neurological disorders such as Alzheimer's disease, mental disorders such as depression, cardiovascular disorders such as hypertension, and endocrine disorders such as diabetes. DataSpeaks, Inc. is a Michigan intellectual property development and out-licensing firm.

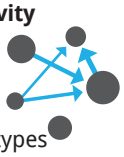


Jump-Start Pharmaceutical Industry Productivity

A **Identify Each Individual Uniquely**
- Genotype or whole genome sequence each patient, cancer, etc.
- This identifies "static nodes" in network graphs.

B **Collect Multivariate Time Series data about "action nodes."**
Such data are becoming ubiquitous.

C **Apply SIMA to measure functional network connectivity** (edges in network graphs).



Compute:
- Diagnostic phenotypes
- Treatment response phenotypes
- Agent phenotypes

D **Analyze group results statistically.**
Model individual and group results mathematically.

For more information, see the key on page 4.

Identify Drug Development Targets

John Lechleiter, CEO of Eli Lilly and Company, said that identifying new neuroscience drug development targets topped his innovation wish list. John spoke at the MichBio CEO Leaders' Club event in July 2015.

- Drug development targets typically are molecular and represent nodes in network graphs.
- The number of molecular targets for approved drugs for all diseases is in the hundreds.
- Consider Interaction-over-Time scores from SIMA that quantify edges as a new category of drug development targets. SIMA has potential to increase the number of targets by orders of magnitude.
- SIMA's edge measures would include measures of brain functional connectivity that might become dysregulated in disorders served by neuroscience as well as all the other types of action nodes identified in the productivity figure key.
- Use SIMA to investigate how drugs up- or down-regulate functional network connectivity beginning at the level of each individual. This elucidates mechanisms.
- DataSpeaks, Inc. seeks leaders to try this approach to drug development.

Re-engineer RCT Design for Precision Drug Development & Medicine

SIMA enables three new ordered within-patient Randomized Controlled Trial (RCT) designs for drugs used for complex chronic disorders:

Advanced design, single-patient (N of 1) RCT designs:

- Randomize doses, including any placebo as zero-dose, to different periods of time for the single patient.
- Use a multitude of DVs for more comprehensive, integrated, and scientific evaluations of safety and efficacy for the single patient.
- Investigate overall benefit and harm; DV-specific, dose specific, delay of response specific, persistence of response specific, etc. benefit and harm for the single patient.
- Consider these N of 1 RCT designs as a potential new gold standard for precision medicine.

Single-group, multiple N of 1 RCT designs:

- A coordinated set of advanced design N of 1 RCTs
- Use a two-tailed, single-group t-test on mean overall benefit and harm in an attempt to reject the null hypothesis of no treatment effect across a multitude of safety and efficacy response variables before drilling down into detailed results.
- Consider these as a potential new gold standard for drug development and approval.

Parallel-group, multiple single-group N of 1 RCT designs:

- Randomize patients to different groups defined by type of treatment before conducting an N of 1 RCT on each patient.
- Use statistical tests for independent groups in an attempt to reject the null hypothesis of no difference in overall benefit and harm by type of treatment before drilling down into detailed results.
- Consider these as a potential new gold standard for comparative safety and efficacy evaluations.

None of these designs confound individuality with treatment effects.

Data Snapshots, Movies, & Mining

The Science of Individuality for the creative destruction of medicine and drug development calls for copious multivariate time series – by analogy, **data movies** with many time series nodes (pixels) and many repeated measurements (frames).

Instead, prevailing methods to test primary hypotheses in RCTs use **data snapshots** and change scores for one primary response variable. Change scores are like differences between only two data snapshots taken at baseline and endpoint.

Compared to data snapshots, data movies can provide orders of magnitude more information to understand individuals scientifically.

Many RCTs collect data that approximate data movies – multivariate time series as defined in the key to the productivity figure. However, much of these data is going to waste for testing primary hypotheses as illustrated in the following figure.

	REPEATED MEASUREMENTS										
	Base-line	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	End-point
IV											
DV 1											
DV 2											
DV 3											
DV 4											
DV 5											
DV 6											
DV 7											
DV 8											
DV 9											
DV 10											

IV = Independent Variable **DV** = Dependent Variable or Response Variable **DV 1** = Primary DV

1. This RCT collected data at baseline and weekly for 10 weeks for one primary response variable or DV and nine other DVs used to help assess safety and efficacy.
2. Treatment, the IV, was absent at baseline and then present for 10 weeks. (This distribution of time on and off drug is not ideal for SIMA. There is no within-patient randomization of dose to help assure that results are valid for individuals. This is an exercise in **data mining**.)
3. The area shown in blue illustrates the amount of data used in a conventional group statistical analysis with change scores for a primary DV.
4. In contrast and with SIMA we could use all the data shown in blue and gray – about 30 times as much – to test **overall benefit and harm** (see productivity figure key) across all 10 DVs. Use all repeated measurements in one analysis to add power for new insight.
5. Next, investigate each individual patient and DV separately to assess and evaluate differential safety and efficacy, delay of response, etc.
6. If data includes blood levels of drug, investigate all treatment effects as functions of blood level of drug.
7. Use the new insights to help identify new indications and contraindications, repurpose drugs, and to target the right drug to the right patient at the right dose.
8. Use these insights to improve conventional RCT designs.
9. Alternatively, re-engineer RCT design for precision drug development and medicine.

Summary

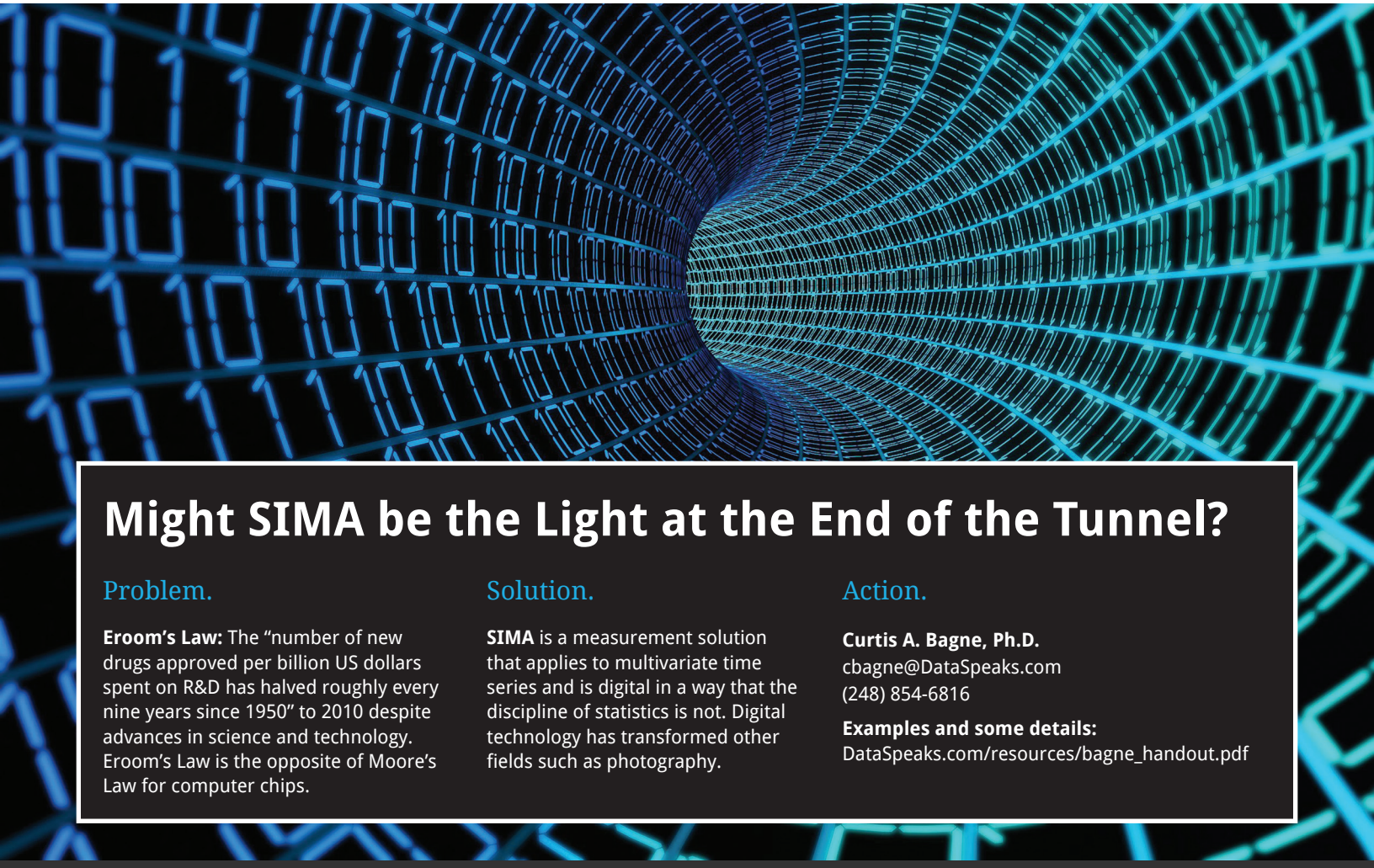
The Science of Individuality Measurement Algorithm (SIMA) from DataSpeaks, Inc. helps enable data-driven scientific understanding of individual patients and other complex adaptive systems.

SIMA can help integrate much of clinical research and patient care, making them better, faster, and less expensive.

Conclusion

SIMA offers leading pharmaceutical companies sustainable competitive advantages.

Universities should help lead the way.



Might SIMA be the Light at the End of the Tunnel?

Problem.

Eroom's Law: The "number of new drugs approved per billion US dollars spent on R&D has halved roughly every nine years since 1950" to 2010 despite advances in science and technology. Eroom's Law is the opposite of Moore's Law for computer chips.

Solution.

SIMA is a measurement solution that applies to multivariate time series and is digital in a way that the discipline of statistics is not. Digital technology has transformed other fields such as photography.

Action.

Curtis A. Bagne, Ph.D.
cbagne@DataSpeaks.com
(248) 854-6816

Examples and some details:
DataSpeaks.com/resources/bagne_handout.pdf

FIGURE KEY

A. Identify Each **Individual** Uniquely

SIMA applies to data about individual Complex Adaptive Systems (CAS). Examples of CAS for this poster include individual cells, cancers, organs, organ systems, brains, patients and other people, and other species. Static nodes are good for forensic identification because they do not change over time.

B. Collect **Multivariate Time Series** Data About Action Nodes

Here Multivariate Time Series refers to two or more time series variables with two or more repeated measurements made at the same times. Temporal resolution can be from fractions of a second to days, weeks, months, or more. Time series can be nested.

Action nodes in network graphs can vary and fluctuate in level over time. Action nodes bring individuals to life. Action nodes can be biological, psychological, or social – internal or external.

Examples of action nodes **internal** to or characteristic of an individual (e.g., being depressed at times) include:

- Gene expression levels
- Laboratory variables such as proteins, lipids, carbohydrates, metabolites, minerals
- Blood levels of drugs and drug metabolites
- Signs and symptoms of disorders
- Electrophysiological measures
- Brain region or voxel activity levels
- Self reports as of pain
- Physical and mental performance, cognition
- Behaviors and exercise
- Quality of life

Examples of **external** or environmental action nodes include intended and actual drug doses, diet, and environmental exposures (allergens, pollutants, etc.). Treatments are environmental variables delivered with therapeutic intent. Time series are becoming **ubiquitous** with hospital, home, wearable, implanted, and environmental monitoring devices; apps; electronic diaries; pill monitors; functional Magnetic Resonance Imaging (fMRI) of brains; etc.

SIMA must have at least one time series for an internal or external **independent or predictor variable (IV)** AND one or more time series for a **dependent, predicted, or response variable (DV).** Each time series is a node in a network graph.

C. Apply SIMA to measure functional network connectivity (edges in network graphs).

SIMA measures network connectivity with Interaction over Time (IoT) scores. IoT scores are internally standardized (mean = 0, standard deviation = 1 unless 0 is the only potential score).

SIMA uses probabilities to help separate signals from noise. For simplicity, SIMA will be illustrated with two time series nodes, X and Y, such as drug dose and pain level, protein X and protein Y, or brain region X and brain region Y.

The current version of SIMA software computes IoT scores as functions of up to 8 analysis parameters simultaneously:

1. Level of X
2. Level of Y
3. Delay of association or effect of X on Y
4. Persistence of association or effect of X on Y
5. Episode length for X
6. Episode criterion for X
7. Episode length for Y
8. Episode criterion for Y

SIMA does not assume that interactions over time are linear. Use SIMA to assess the temporal criterion of casual relationships for individuals. Use SIMA to assess Boolean events such as:

- Effects of drug X AND drug Y on symptom Z
- Effects of protein X OR protein Y on protein Z

Use more repeated measurements to increase IoT score reliability and increase power. Process streaming time series iteratively as they become available for real-time decision support. Compute IoT scores to help quantify emergent properties of CAS (e.g., coordinated action). Investigate brain to behavior and behavior to brain effects.

Benefit and harm scores are a variation of IoT scores for **evaluation** as in **Randomized Controlled Trial (RCT)** designs. Benefit and harm scores:

- Adjust signs. For example, set both higher levels of HDL ("good" cholesterol) and lower levels of LDL ("bad" cholesterol) while on drug to be beneficial.
- Consider using differential weights for multiple DVs to account for differences in clinical significance and patient preferences before averaging DV-specific **benefit and harm scores** to compute **overall benefit and harm scores** across dozens of time-varying response variables. Use these to help integrate and balance safety and efficacy evaluations scientifically starting at the level of each individual patient.

Compute phenotypes:

Diagnostic phenotypes:

- All action nodes are internal to or characteristic of the individual.
- Compute taxonomies of functional disorders and diseases that are more objective, reliable, specific, mechanistic, and actionable.

Treatment response phenotypes:

- At least one action node time series such as drug dose operates in SIMA as an IV.
- At least one action node time series such as signs, symptoms, or biomarkers operate as DVs.
- Randomize and blind doses, including placebo as zero dose, over time to help assure that treatment response phenotypes are valid for individuals.

Agent phenotypes:

- Quantify how individual CAS affect their environments over time. Examples: How does one's pancreas or brain affect one's behavior or body as a whole? How does one's behavior affect other people and one's environment?
- At least one action node time series is about the individual or its behavior. At least one action node time series is about one's environment. Everyone is an agent. **You can be an agent for change.**

SIMA can help advance **computational phenomics**.

D. Analyze Group Results Statistically

Describe groups. Make inferences from samples of individuals to populations.
Identify treatment effect factors – e.g., anticholinergic drug effects.
Help assess clinical significance of treatment effects – e.g., how do drug effect IoT scores on a protein or lipid fraction affect survival?

Identify genetic and other predictors of:

- Susceptibility to disorder and disease
- Differential response
- Differential dose requirements.

Model individual and group results mathematically.
SIMA comes between data collection and statistical analyses to help connect the power of mathematics and statistics to the real world.