

*Is There a Role of Big Data in Drug Development Decisions?
ACoP6 Oct. 5, 2015 Crystal City, VA*

A Pharmacometrician's Perspective for Utilization of Big Data

Marc R. Gastonguay, Ph.D.

President & CEO
Metrum Research Group
metrumrg.com

Scientific Director
Metrum Institute
metruminstitute.org

Disclaimer

The views and opinions expressed in this presentation are those of the individual presenter and should not be attributed to or considered binding on all pharmacometricians.

Big Data: What's New?

Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland
Statistics Research, Bell Labs
wsc@bell-labs.com

“Exploratory data analysis” is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there.

- John Tukey

ISI Review, , 69, 21-26. W. S. Cleveland, 2001.

<http://www.stat.purdue.edu/~wsc/papers/datascience.pdf>

J.W. Tukey. Exploratory Data Analysis: Past, Present, and Future. Tech Report No. 302. Princeton University, 1993.

Perspective: 600 Execs, 20 Industries (MIT, 2011)

PROVEN AND PERCEIVED EFFECTS ON BUSINESS PERFORMANCE



IMPROVEMENT IN PERFORMANCE*
for firms that emphasize decision-making based on data and analytics.



PERCEIVED IMPROVEMENT
in performance through Big Data over the last 3 years.



EXPECTED IMPROVEMENT
in performance due to Big Data in the next 3 years.

Sources
Over 600 senior-level executives worldwide from across 20 industries.

*MIT study, 2011

<http://www.nyi.net/blog/2012/07/big-data-and-decision-making/>

Should Pharmacometrics Follow the Trend?

BIG DATA HELPING ENTERPRISE DECISION-MAKING

Big Data Is Improving Decision-making

79% of businesses say improved uses of big data will lead to better decisions.



79%

CIO INSIGHT 2013

<http://www.cioinsight.com/it-strategy/big-data/slideshows/big-data-helping-enterprise-decision-making.html>

Why Not?

BIG DATA HELPING ENTERPRISE DECISION-MAKING



29%

Not Everyone is Banking on Big Data

29% of companies surveyed have no plans to implement big data technology now.

CIO INSIGHT 2013

BIOMEDICAL BIG DATA EXPLOSION

NIH National Center for Biotechnology Information DATA STORAGE

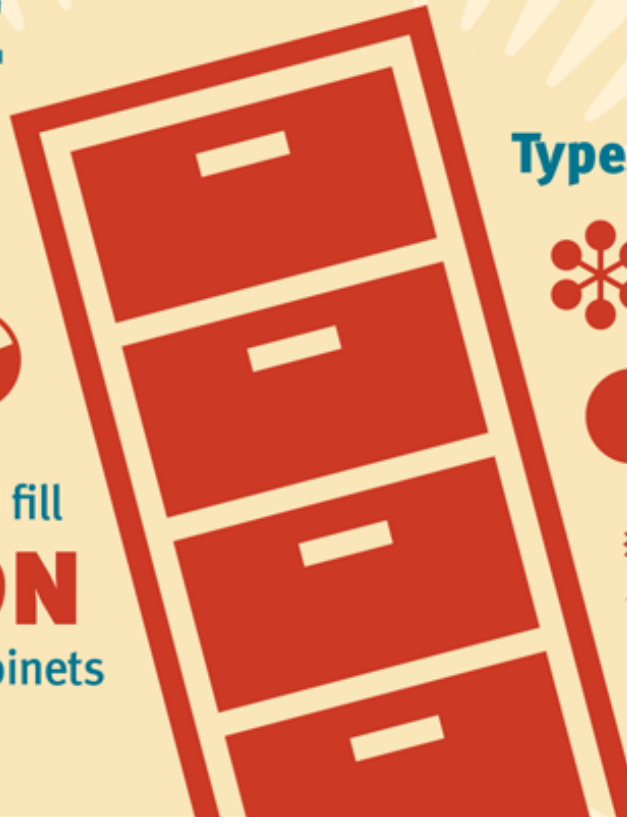
In **1990** fit on
3 floppy disks



In **1993** fit on
1 CD-ROM



In **2014** could fill
400 MILLION
4-drawer filing cabinets



Types of BD2K Awards



Enabling Data
Utilization



Analysis Methods
and Software



Enhancing
Training



Centers of
Excellence



Big Data to Knowledge (BD2K) is an initiative of the National Institutes of Health

Buzzwords, Hype, & Jargon



“In general, hype masks reality and increases the noise-to-signal ratio. The longer the hype goes on, the more many of us will get turned off by it, and the harder it will be to see what’s good underneath it all, if anything.”

- R. Schutt & C. O’Neil. Doing Data Science. O’Reilly Cambridge, 2013.

Do We Really Know What Lies Ahead?



<http://timoelliott.com/blog/2013/06/the-ethics-of-big-data-vendors-should-take-a-stand.html>

Do We Really Know What Lies Ahead?

An iceberg floating in the ocean. The tip of the iceberg, which is visible above the water, is relatively small and jagged. The much larger part of the iceberg is submerged underwater, representing hidden or unknown aspects of a problem. The words "BIG DATA" are faintly visible on the submerged part of the iceberg.

New Insights & Better Decisions

Data Quality

Contracts, IP

New Skills Development

Accuracy of Conclusions

Infrastructure Investment

Opportunity Cost

Data Cleaning



“The issue isn’t Big Data, it’s Clean Data...Now, the issue is how to make sure the data is in a usable state to glean information from it. It’s a much harder problem, and one I hope gets as much buzz and press as Big Data.”

- Andres Alvarez

Data Cleaning

e.g. Electronic Medical Records

- Imputing gaps in data
- De-identifying data
- Resolving discrepancies / duplications
- Aligning terminology
- Converting to useful formats

Are we even competently managing small data?

“The issue isn’t Big Data, it’s Clean Data...Now, the issue is how to make sure the data is in a usable state to glean information from it. It’s a much harder problem, and one I hope gets as much buzz and press as Big Data.”

- Andres Alvarez

Pooled Data for MBMA in Alzheimer's Disease

<http://www.adni-info.org/>



- Natural History
- Interpatient Variability
- Patient Specific Factors
- Imaging and CSF Biomarkers

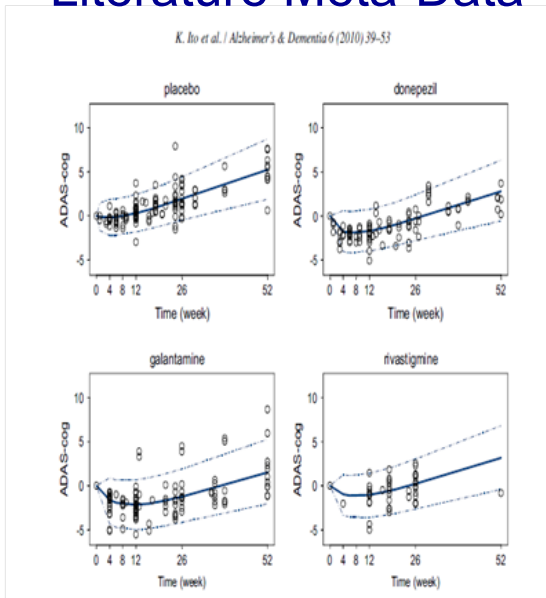
Sub-populations

Normal (N=200)

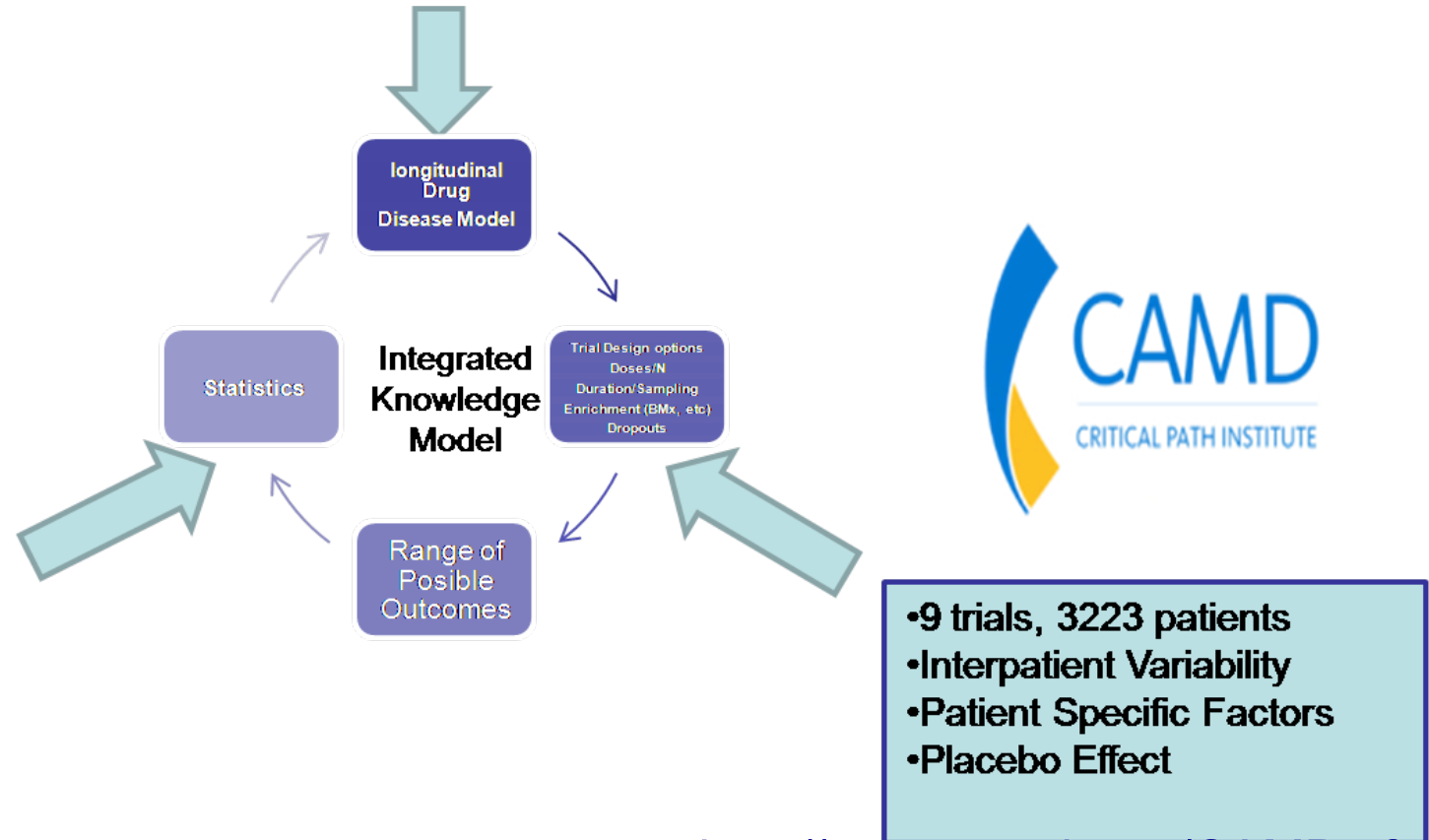
MCI (N=400)

Mild AD (N=200)

Literature Meta-Data



- 73 Trials (1990 to Present)
- Interstudy variability
- Estimate of drug treatment effects (magnitude, onset, offset)



- 9 trials, 3223 patients
- Interpatient Variability
- Patient Specific Factors
- Placebo Effect

<http://www.c-path.org/CAMD.cfm>

Data Cleaned – What Next?



"After careful consideration of all 437 charts, graphs, and metrics, I've decided to throw up my hands, hit the liquor store, and get snocked. Who's with me?!"

<http://whatsthebigdataidea.com/2015/02/27/big-data-friday-funny/>

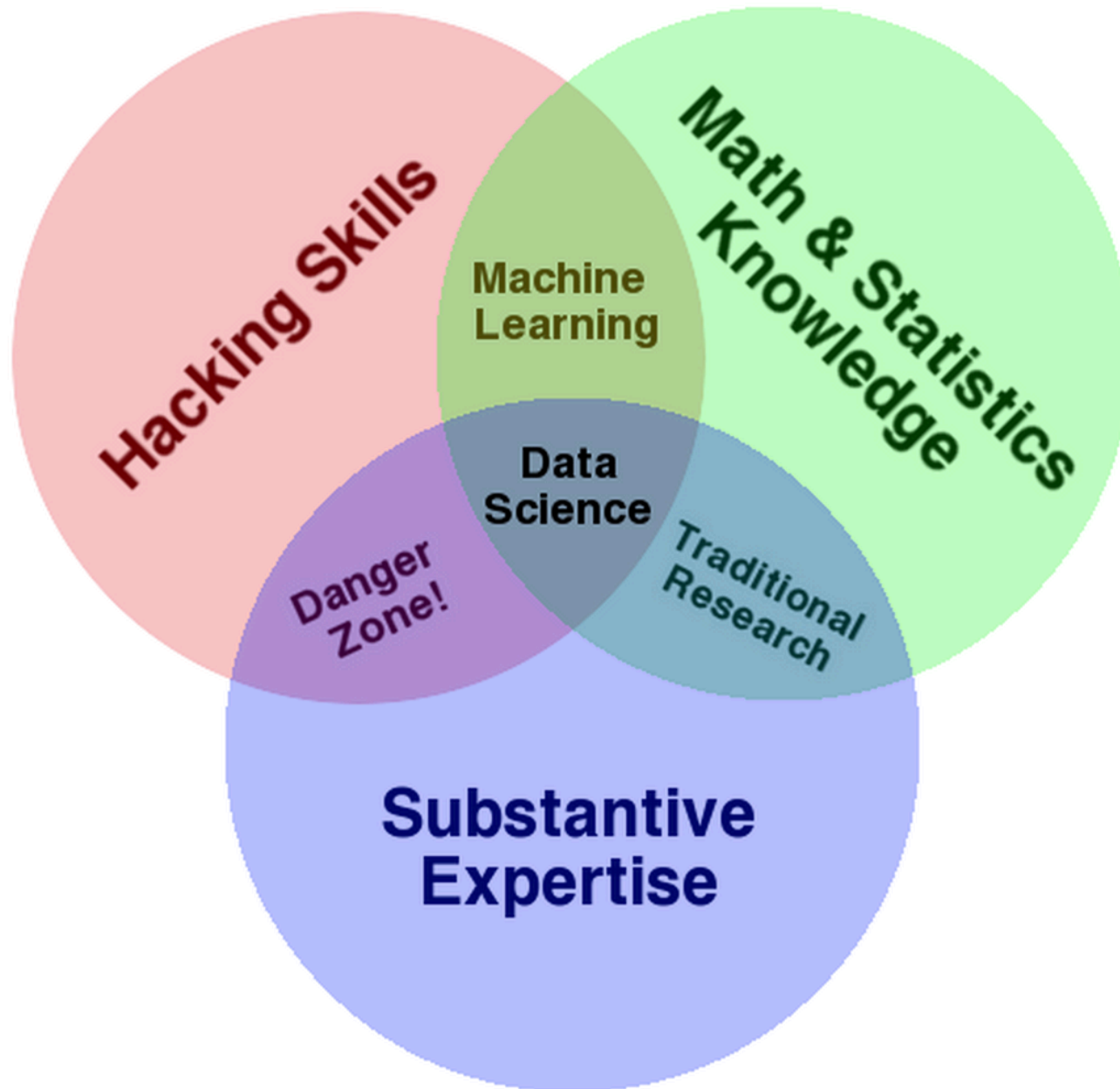
Data Cleaned – What Next?



“Long before worrying about how to convince others, you first have to understand what’s happening yourself.”

- Andrew Gelman

Skills: Converting Big Data to Big Knowledge



Big Data for Hiring Decisions



“Let’s put everything in and let the data speak for itself.”

- Dr. V. Ming

What if algorithm predicts that males will be better employees?

“Models that ignore causation can add to historical problems instead of addressing them.”

- R. Schutt & C. O’Neil.
Doing Data Science.
O’Reilly Cambridge, 2013.

How Big Data Is Playing Recruiter for Specialized Workers. NY Times April 27, 2013

<http://www.nytimes.com/2013/04/28/technology/how-big-data-is-playing-recruiter-for-specialized-workers.html>

<http://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html>

Assumptions & Biases in Big Data

BigData != AllData

> 20 million tweets between October 27 and November 1
study combined Sandy-related Twitter and Foursquare data
Results:
grocery shopping peaked the night before the storm
nightlife picked up the day after (most tweets about Sandy
came from Manhattan, very few messages originated from
Jersey shore)



http://photos.nj.com/star-ledger/2012/11/hurricane_sandy_before_and_aft_7.html
The Hidden Biases in Big Data. Kate Crawford. Harvard Business Review
<https://hbr.org/2013/04/the-hidden-biases-in-big-data/>

Google Flu Trends Tracker

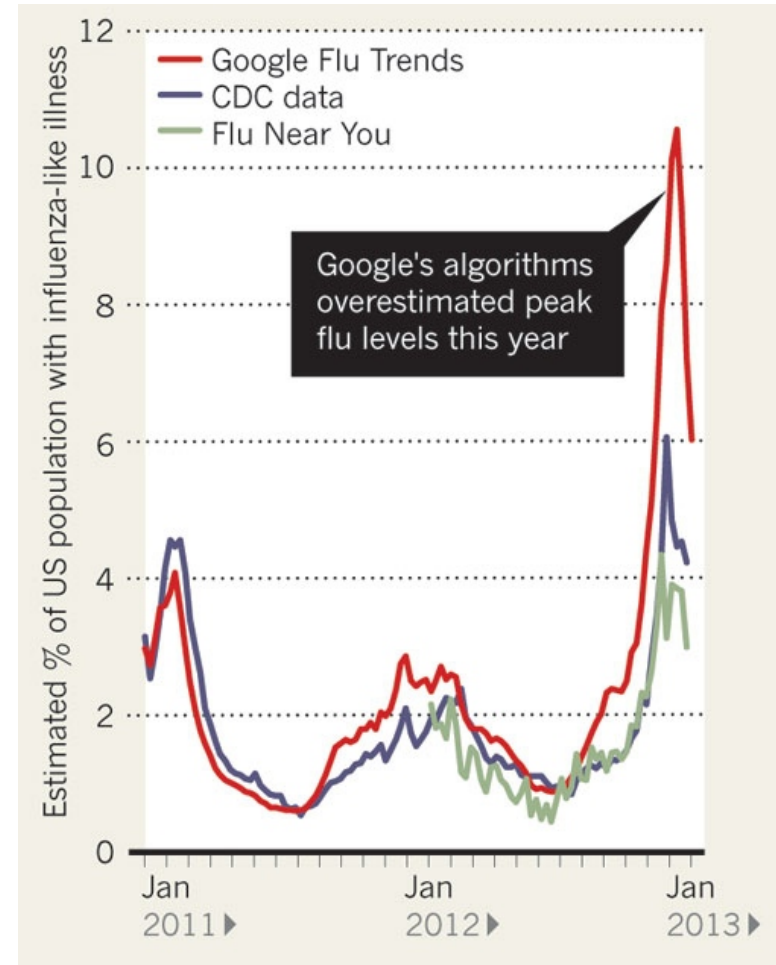
Nature. 2009 Feb 19;457(7232):1012-4. doi: 10.1038/nature07634.

Detecting influenza epidemics using search engine query data.

Ginsberg J¹, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L.

Big data can be... “complete bollocks. Absolute nonsense... There are a lot of small data problems that occur in big data... They don’t disappear because you’ve got lots of the stuff. They get worse.”

David Spiegelhalter
Winton Professor of the Public
Understanding of Risk,
Cambridge University



Harford. Big Data: Are We Making a Big Mistake? <http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html>
D. Butler. Nature. <http://www.nature.com/news/when-google-got-flu-wrong-1.12413>

Accurate Inference about Covariate Effects?

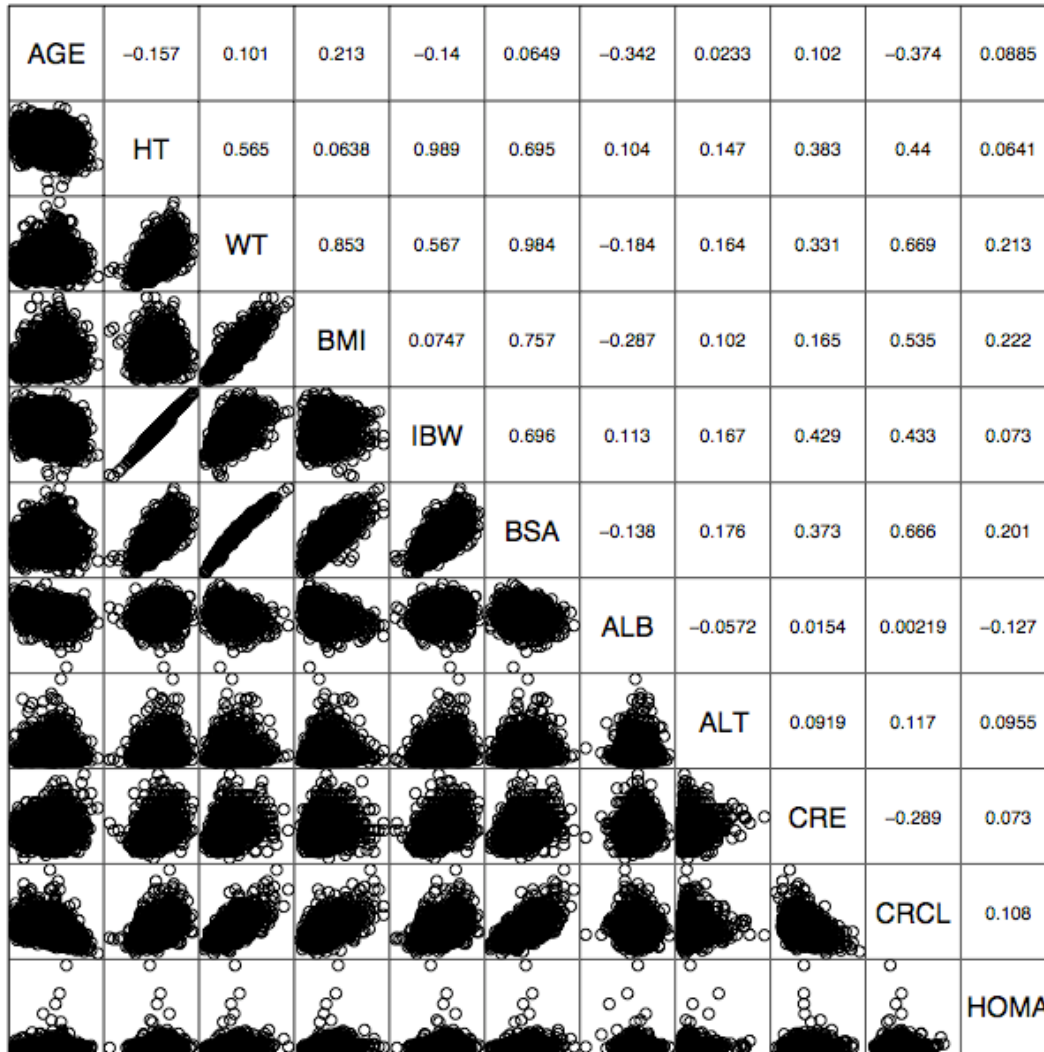
Covariate Effects:

Creatinine clearance was the most significant covariate on [REDACTED] clearance. The clinical importance of this finding is unknown, since less than 2% of [REDACTED] dose is excreted in the urine. This may be an artifact of the data as the current analysis data set did not include patients with moderate or severe renal impairment. Weight, age and sex were not significant covariates and, therefore, require no dose adjustment.

Making accurate inferences about predictors themselves is challenging:

- Studies not prospectively designed, randomized, stratified, powered
- Correlation between predictors
- Particularly problematic with stepwise and similar approaches

Inferences about Covariates



Correlation/Collinearity

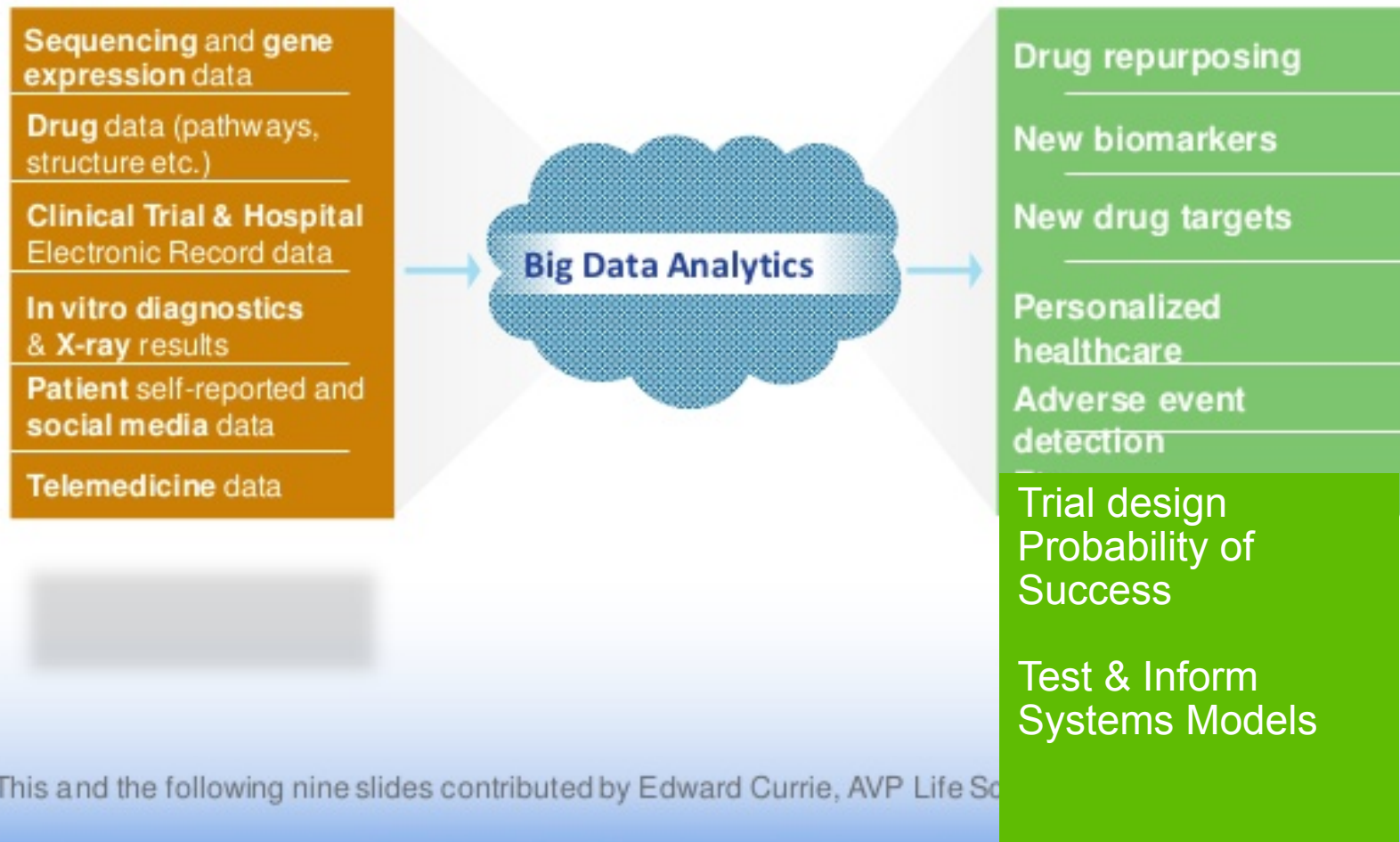
When the goal is to make inferences about the covariates themselves...

Covariate effects to be included in model should be independent, e.g. they carry unique information.

See: Bonate, P. L.. The effect of collinearity on parameter estimates in nonlinear mixed effect models (article). Pharmaceutical Research. 1999 Volume 16 Number 5 Pages 709-717.

Opportunities to Link Big Data and Pharmacometrics

Big Data in Life Sciences*



*This and the following nine slides contributed by Edward Currie, AVP Life Sciences

How?



© marketoonist.com

<http://bigdataisafad.com/the-slightest-idea-about-big-data-funny/>

Learning from other Disciplines

Springer Series in Statistics

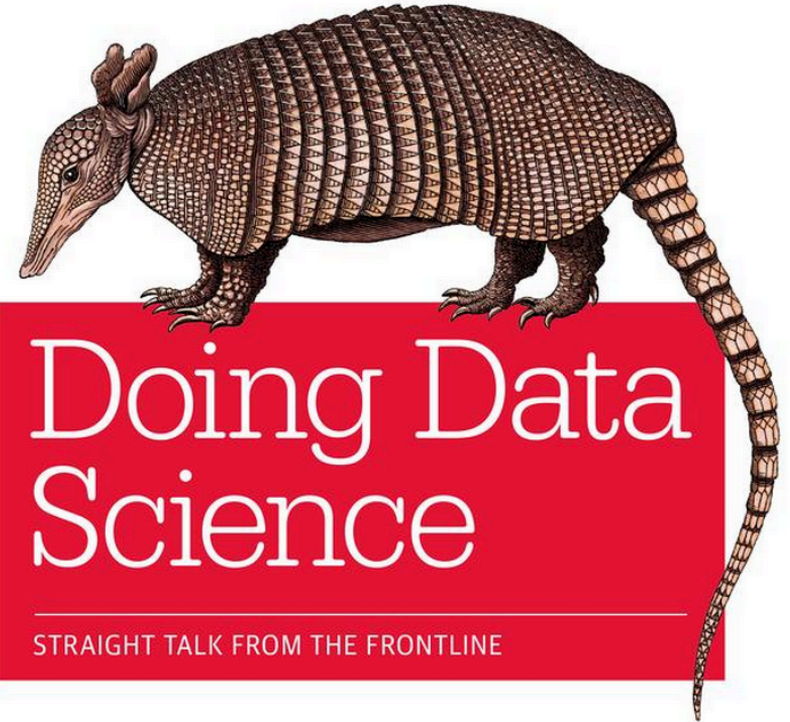
Trevor Hastie
Robert Tibshirani
Jerome Friedman

The Elements of Statistical Learning

Data Mining, Inference, and Prediction

Second Edition

O'REILLY®



Rachel Schutt & Cathy O'Neil

Genomics Meets Pharmacometrics

[Pharmacogenet Genomics](#), 2013 Jan;23(1):9-18. doi: 10.1097/FPC.0b013e32835ade82.

Pharmacogenetics-based population pharmacokinetic analysis of etravirine in HIV-1 infected individuals.

[Lubomirov R¹](#), [Arab-Alameddine M](#), [Rotger M](#), [Fayet-Mello A](#), [Martinez R](#), [Guidi M](#), [di Iulio J](#), [Cavassini M](#), [Günthard HF](#), [Furrer H](#), [Marzolini C](#), [Bernasconi E](#), [Calmy A](#), [Buclin T](#), [Decosterd LA](#), [Csajka C](#), [Telenti A](#); [Swiss HIV Cohort Study](#).

[Br J Clin Pharmacol](#), 2008 Apr;65(4):548-57. Epub 2007 Oct 8.

Effect of CYP2C19 polymorphism on nelfinavir to M8 biotransformation in HIV patients.

[Hirt D](#), [Mentré F](#), [Tran A](#), [Rey E](#), [Auleley S](#), [Salmon D](#), [Duval X](#), [Tréluyer JM](#); [COPHAR2- ANRS Study Group](#).

[Clin Pharmacokinet](#), 2013 Jul;52(7):543-53. doi: 10.1007/s40262-013-0057-6.

Simultaneous pharmacogenetics-based population pharmacokinetic analysis of darunavir and ritonavir in HIV-infected patients.

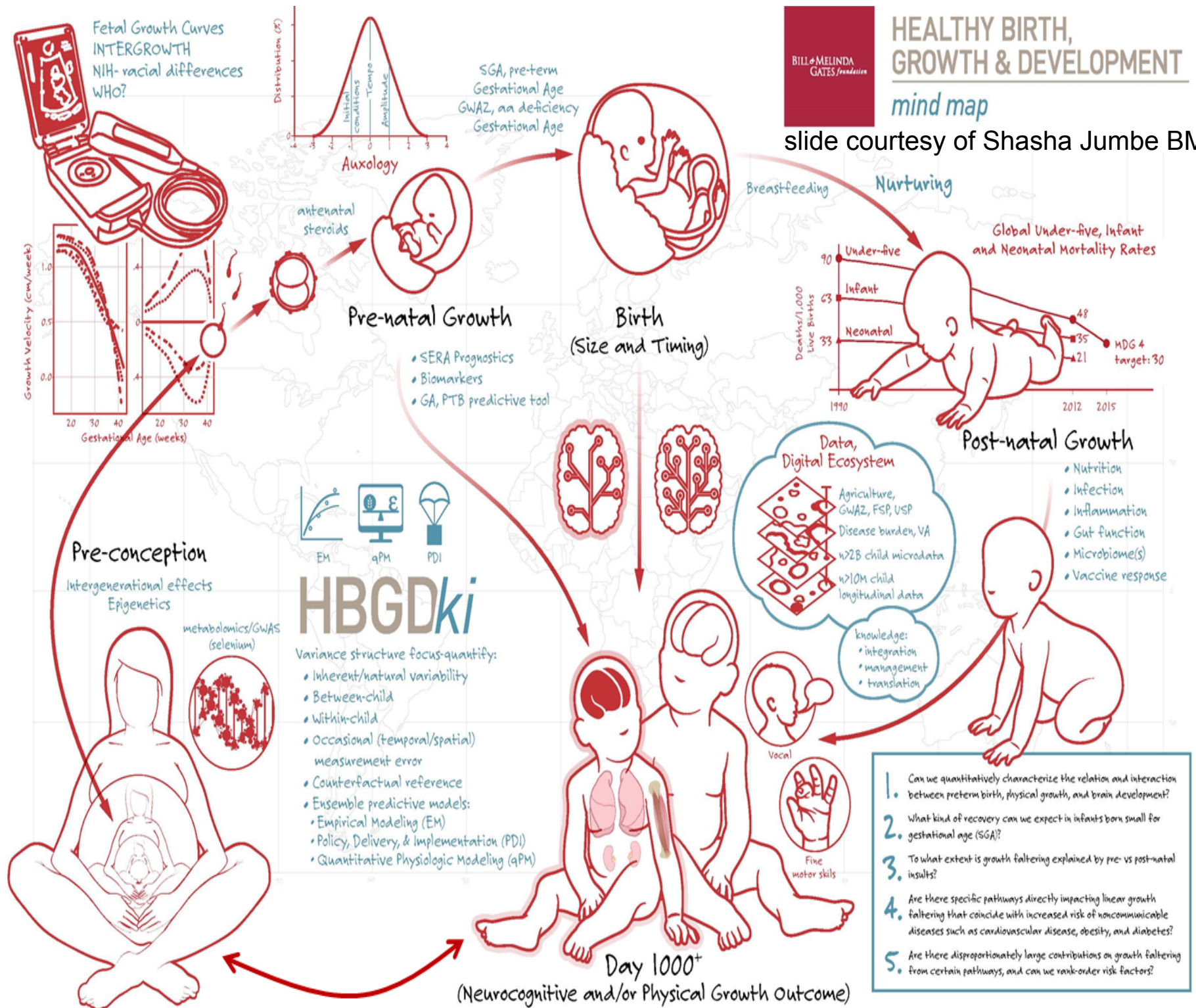
[Moltó J¹](#), [Xinarianos G](#), [Miranda C](#), [Pushpakom S](#), [Cedeño S](#), [Clotet B](#), [Owen A](#), [Valle M](#).



HEALTHY BIRTH, GROWTH & DEVELOPMENT

mind map

slide courtesy of Shasha Jumbe BMGF

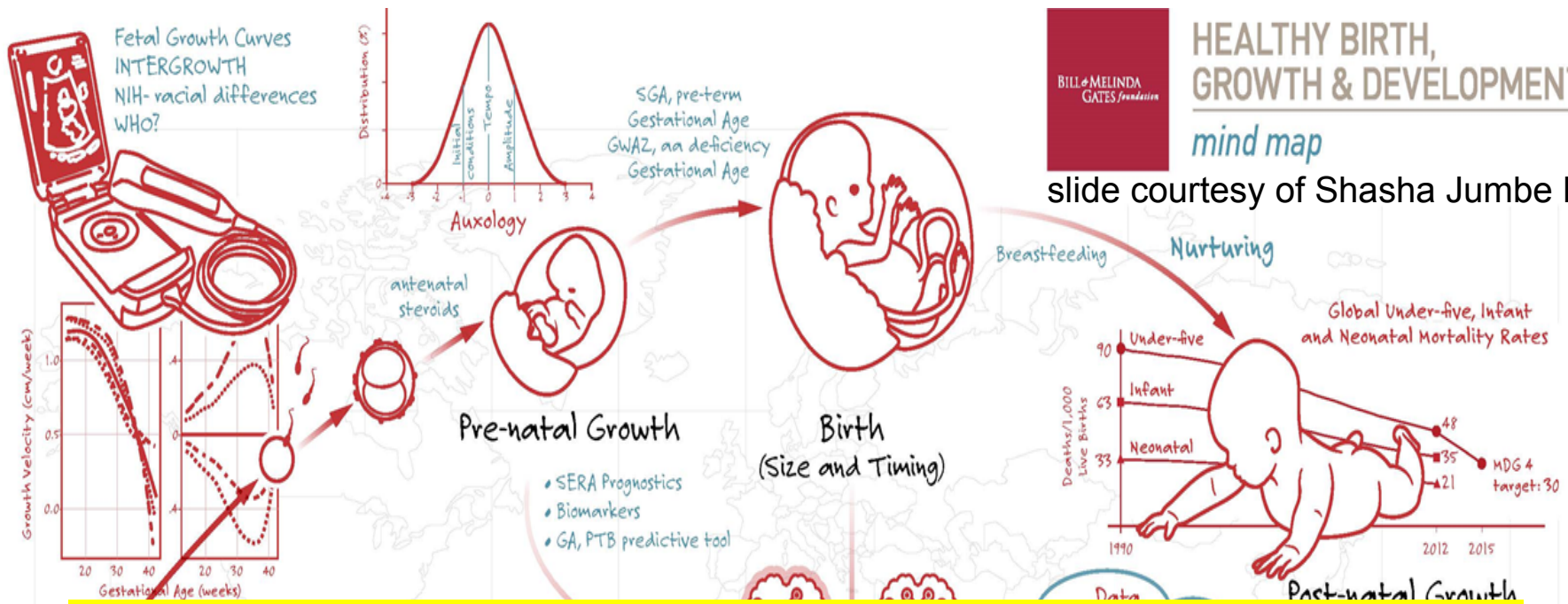




HEALTHY BIRTH, GROWTH & DEVELOPMENT

mind map

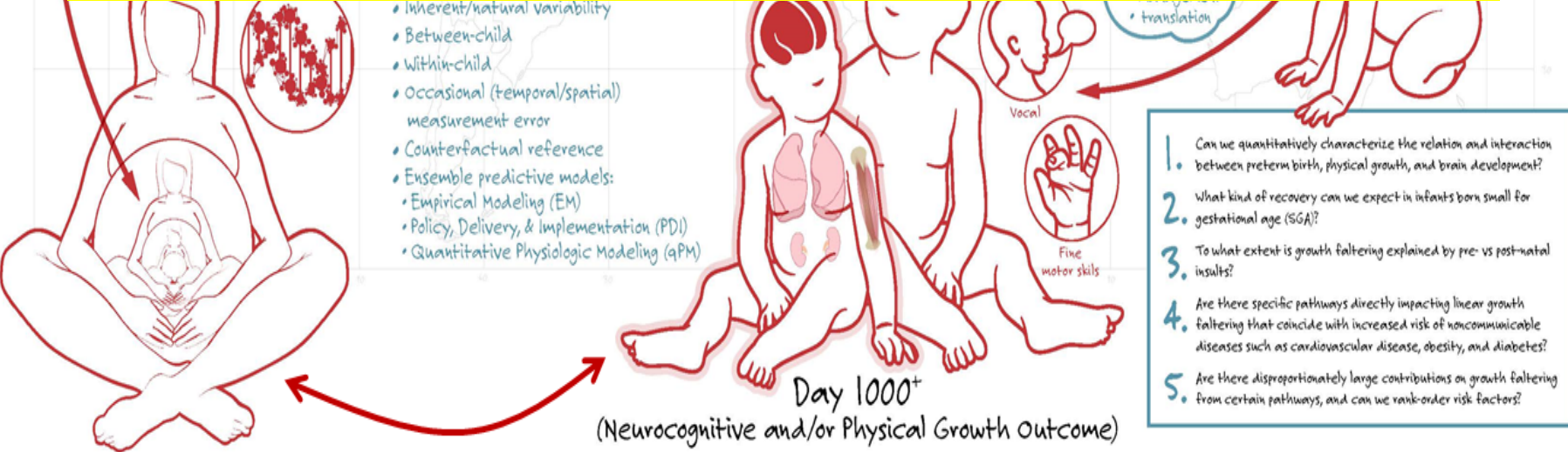
slide courtesy of Shasha Jumbe BMGF



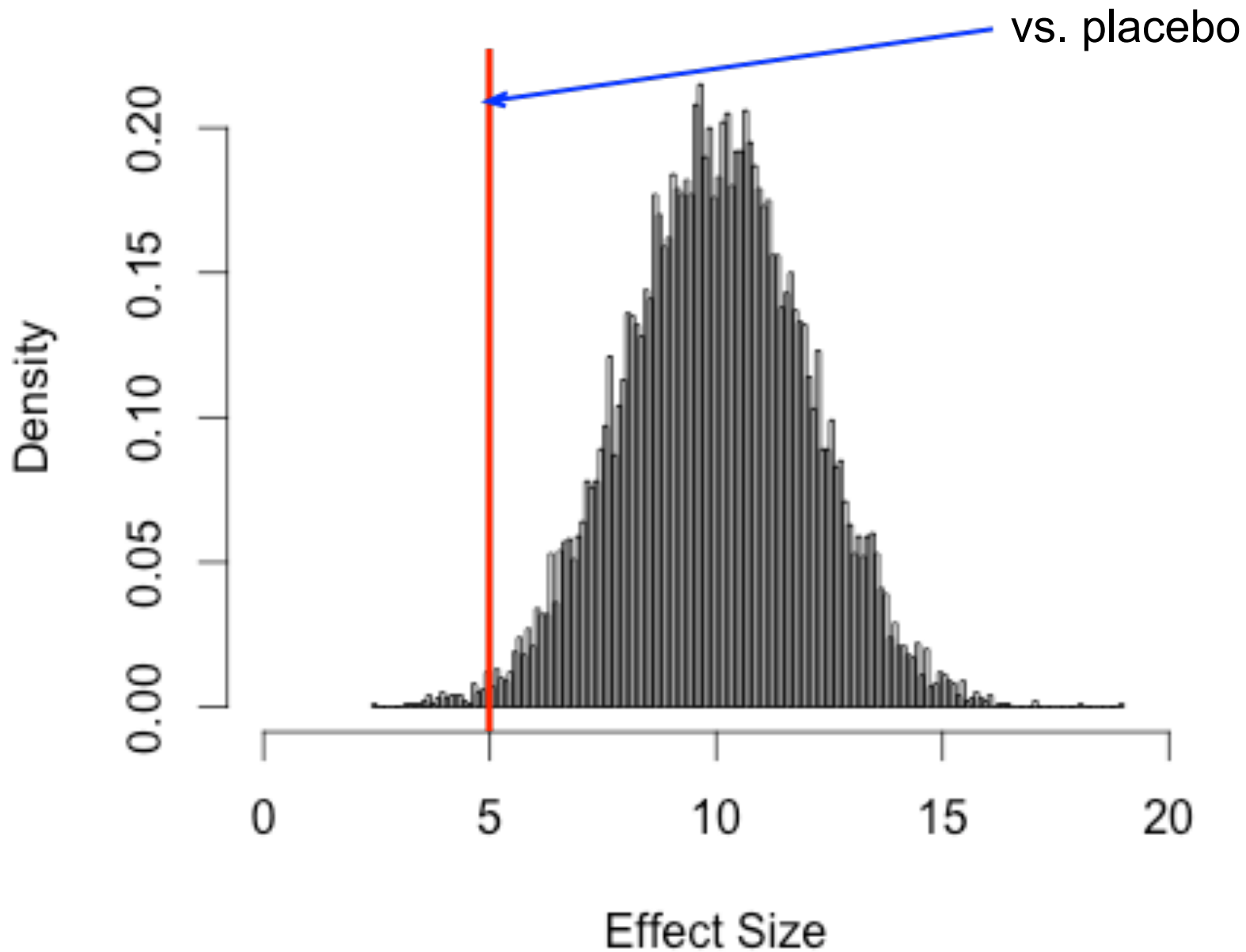
• Collected data from 1.67 MM individual children

• >10 MM by the end of the year

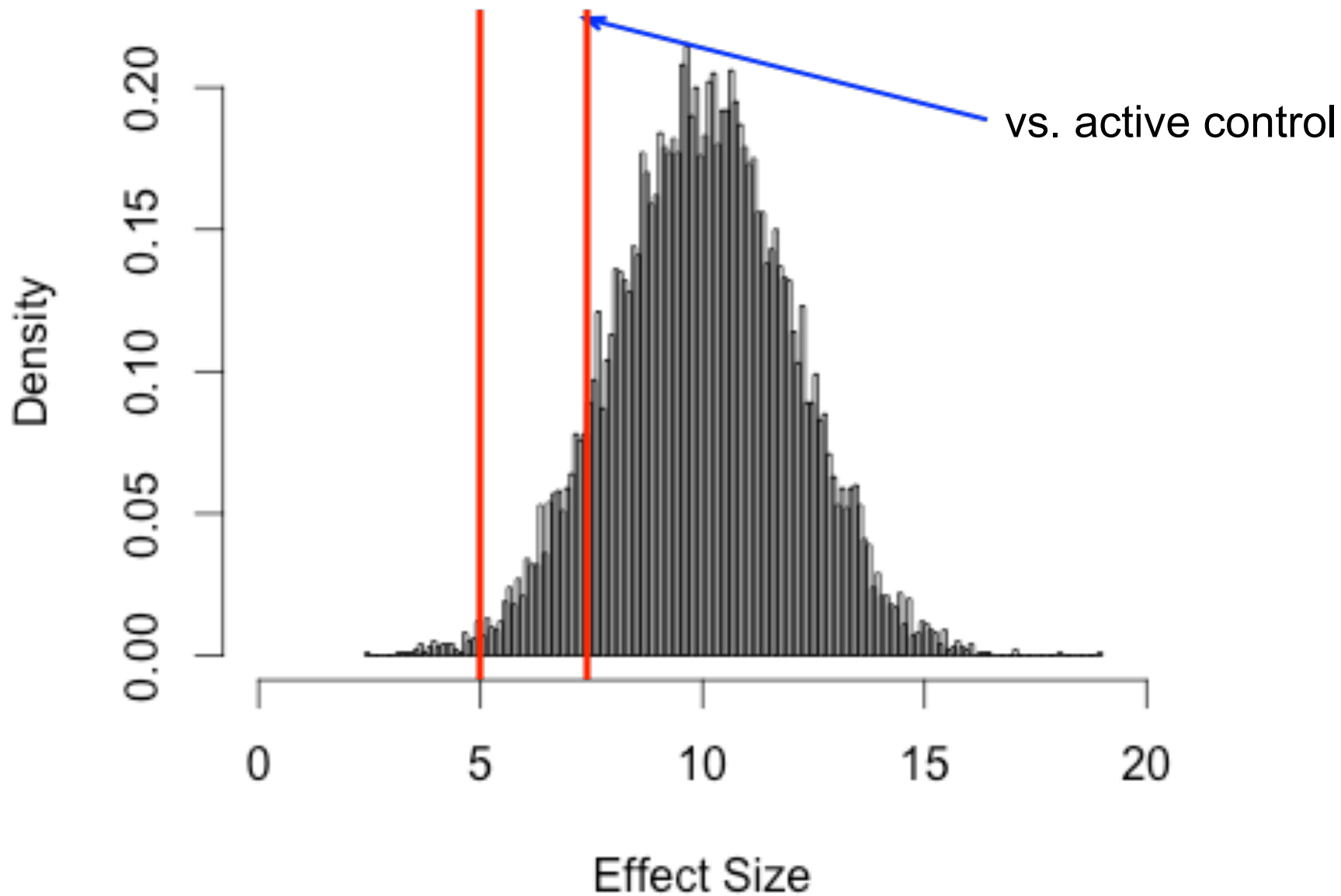
- Shasha Jumbe BMGF



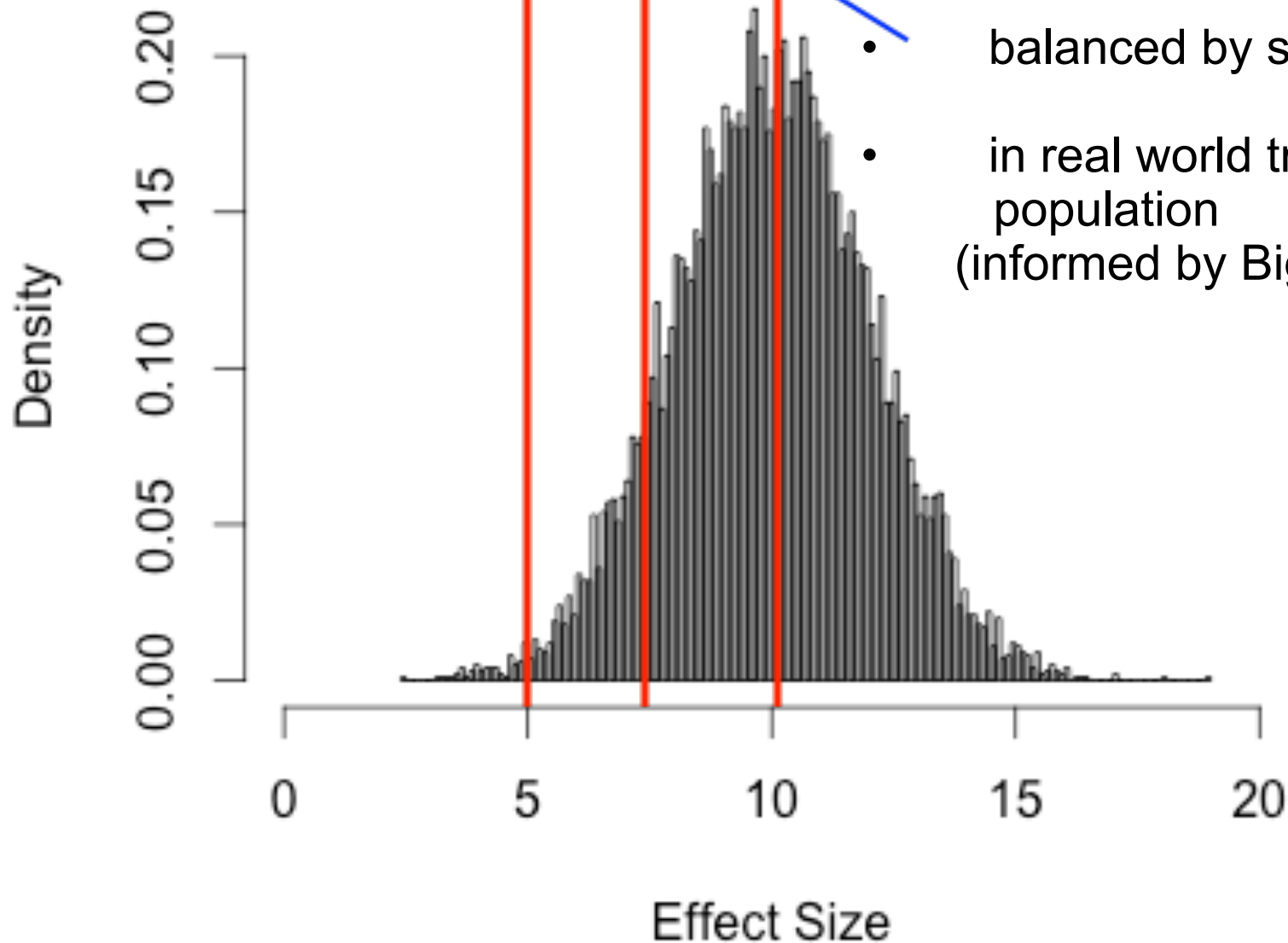
Probability of Success: Changing Targets



Probability of Success: Changing Targets



Probability of Success: Changing Targets



- vs. future competitor
- balanced by safety and cost
- in real world treatment population (informed by Big Data)

Probability of Success in Real World Trial

Trial Uncertainty Simulator

Demographics

Test CP ▾

Placebo ▾

Ph2 --> Ph3 Adjustments

Output ▾

Advanced

Reference value for true treatment difference

4

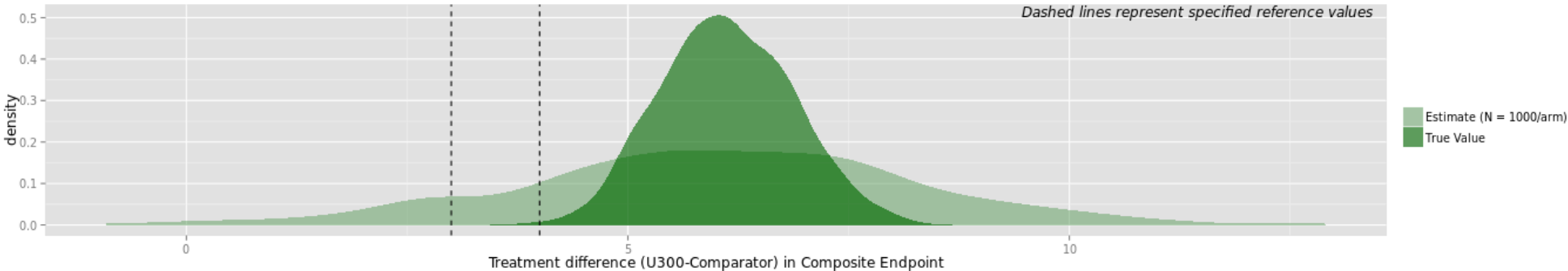
Reference value for estimate of treatment difference

3

Sample size per arm

1000

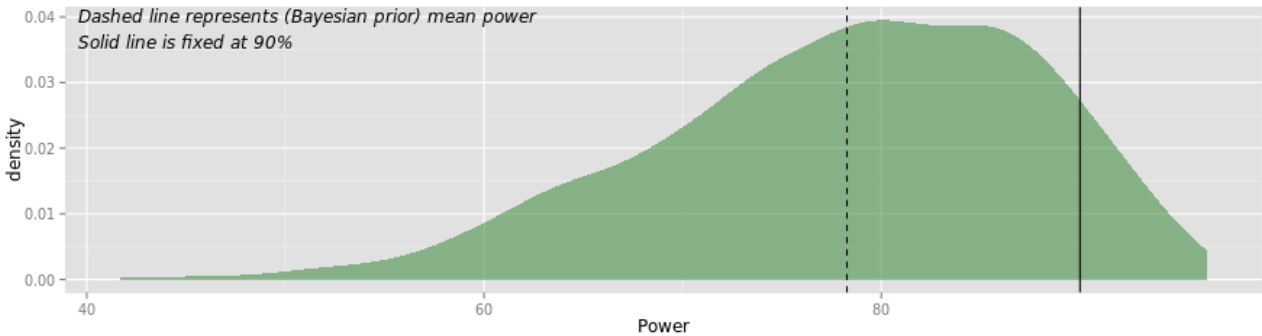
Prior Uncertainty and Estimation Variability



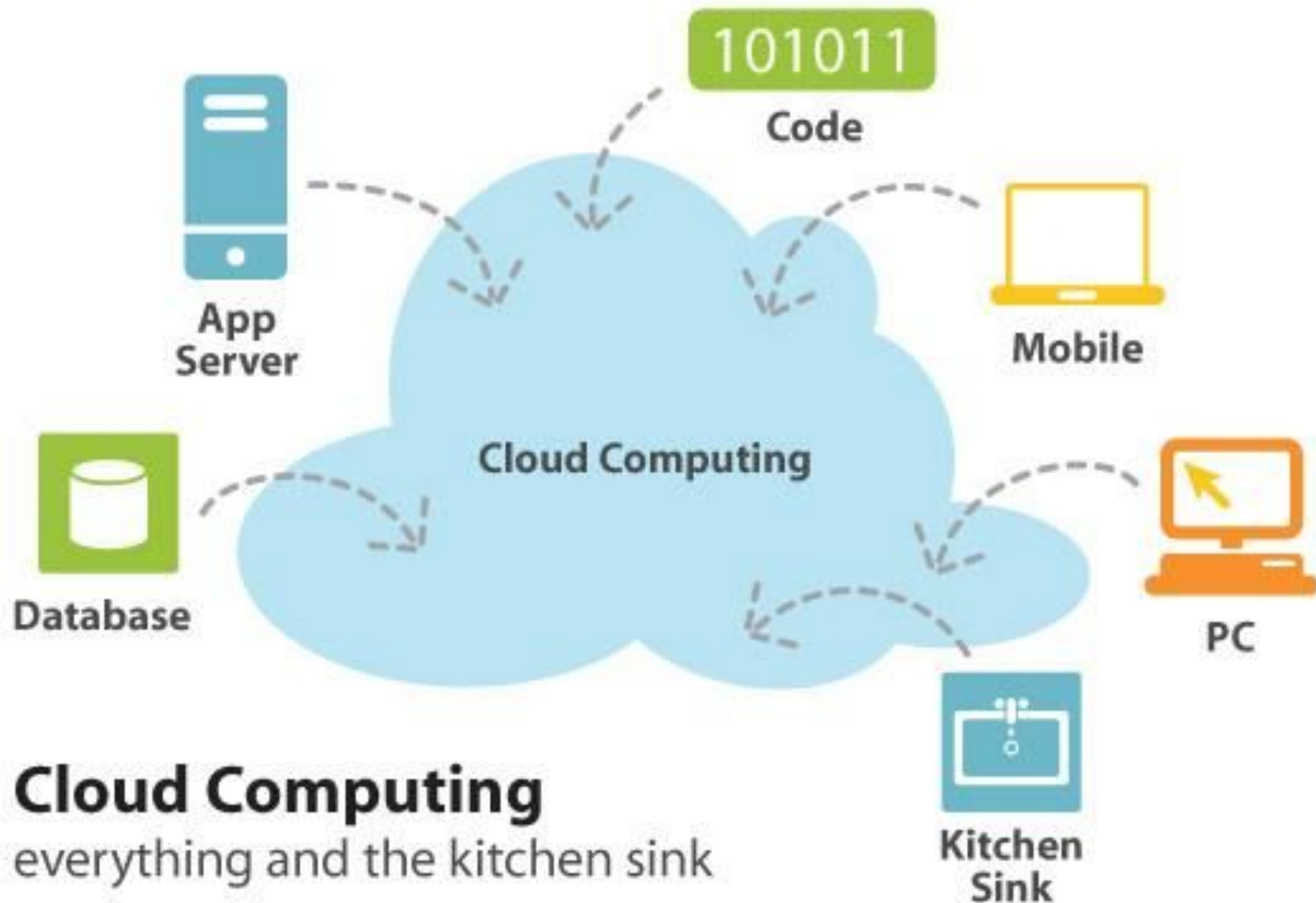
Tabular Summary

	Estimate (%)
Expected Treatment difference (U300-Comparator)	6.11
P(True value > 4)	99.90
P(Trial estimate > 3)	89.70
Average (Bayesian Predictive) Power	78.30

Power

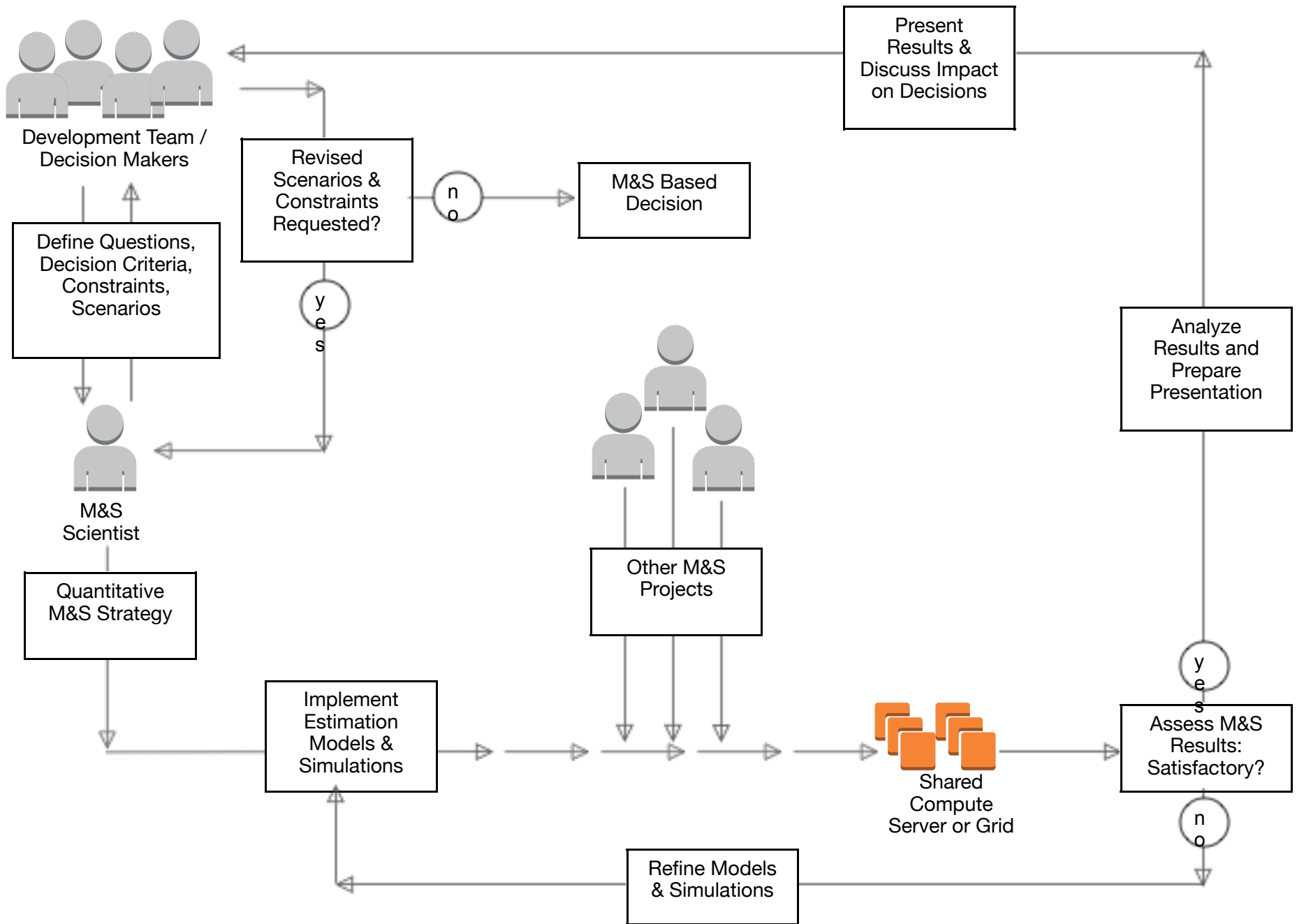


Enabling Technology: Big Computation (EC2)

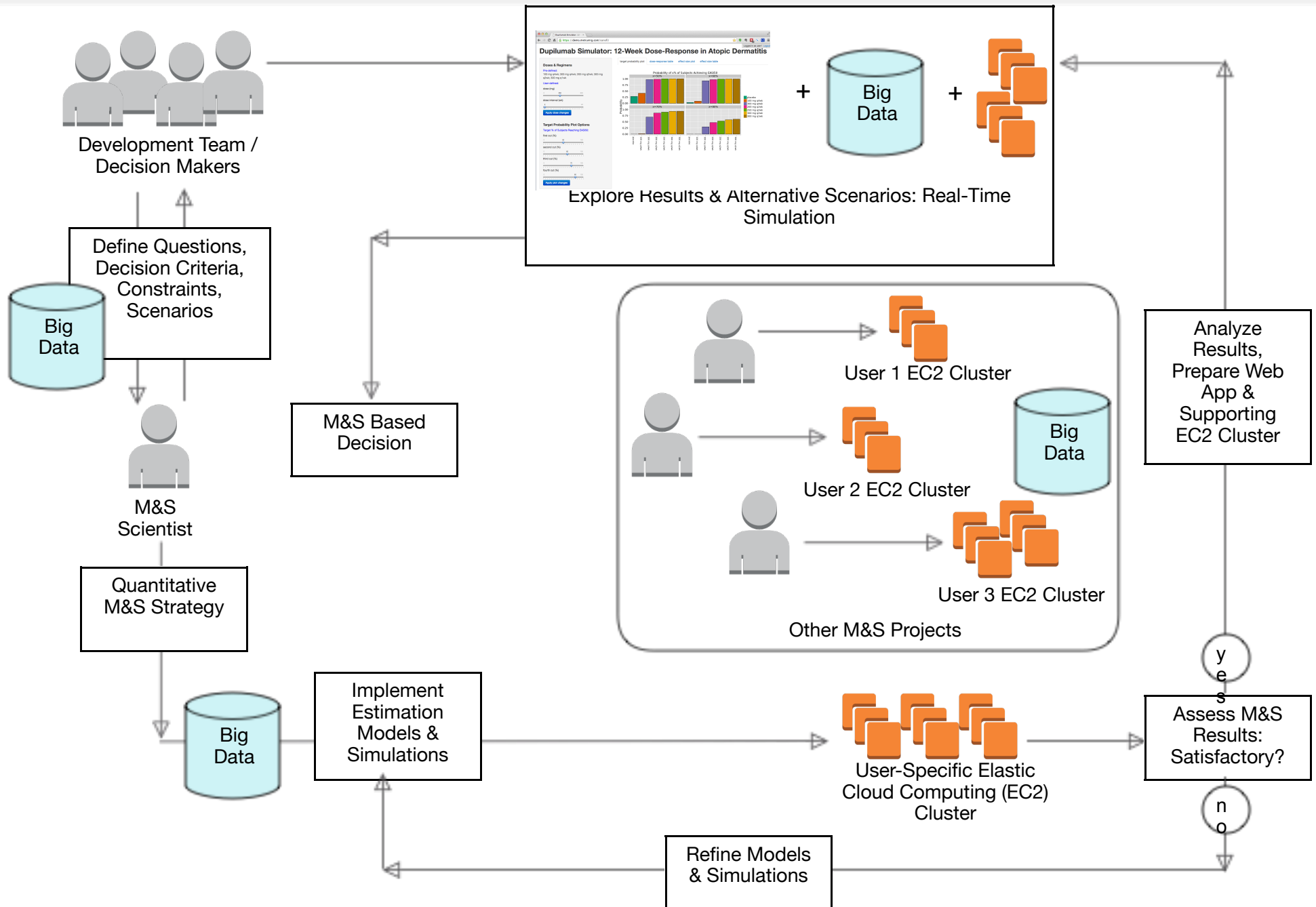


ACOP4 April 2013 Open Forum on Cloud Computing
http://www.cloudict.com/1/Cloud_Computing.html

Typical Process for M&S Based Decision Support



Big Data & EC2 Enabled M&S for Decision Support



Big Data & Pharmacometrics

Some Opportunities

- Inform and test systems models
- Develop well informed questions
- Provide real-world patient characteristics for trial simulation
- Assess probability of success in real world

Cautions

- Continue to seek understanding / causation
- Recognize the biases and limitations
- Balance effort / investment with existing needs
- Don't discount the value of well designed prospective studies

What's Next?



<https://s-media-cache-ak0.pinimg.com/236x/b3/4f/c9/b34fc9fbc5797c7210ff1d86796d9bd.jpg>