

**27<sup>th</sup> Annual E. Leonard Arnoff/Milton J. Schloss  
Memorial Lecture on the Practice of Business Analytics**

# **How Small Are Our Big Data: Turning the 2016 Surprise into a 2020 Vision**

**Xiao-Li Meng, Harvard University**

**Whipple V. N. Jones Professor of Statistics**

**Dean of the Graduate School of Arts and Sciences**

The term “Big Data” emphasizes data quantity, not quality. However, much of the current measures of statistical uncertainties and errors are adequate only when the data are of perfect quality. We show that once we take into account the data quality, the effective sample size of a Big Data set can be vanishingly small. Without understanding this phenomenon, Big Data can do more harm than good due to drastically inflated precision assessments that cause gross overconfidence. This overconfidence leads us to be caught by surprise when reality unfolds, as we all experienced during the 2016 US Presidential election. Data from the Cooperative Congressional Election Study (conducted by Stephen Ansolabehere, Douglas Rivers and others, and analyzed by Shiro Kuriwaki), are used to assess the data quality in 2016 US election polls, with the aim of gaining a clearer vision for the 2020 election and beyond.

**Xiao-Li Meng** is well known for his depth and breadth in research, his innovation and passion in pedagogy, and his vision and effectiveness in administration, as well as for his engaging and entertaining style as a speaker and writer. Meng has received numerous awards and honors for the more than 150 publications; he has delivered more than 400 research presentations and public speeches. His interests range from the theoretical foundations of statistical inferences (e.g., the interplay among Bayesian, frequentist, and fiducial perspectives) to statistical methods and computation (e.g., posterior predictive p-value; EM algorithm; Markov chain Monte Carlo) to applications in natural, social, and medical sciences and engineering (e.g., complex statistical modeling in astronomy and astrophysics, assessing disparity in mental health services, and quantifying statistical information in genetic studies).

**6:00 PM, Wednesday, April 18, 2018**

**Kingsgate Marriott Conference Center (Ballroom on the 2nd Floor)**

Registration: <https://www.regonline.com/arnoff-schloss-lecture>

**Sponsor: Department of Operations, Business Analytics, and Information Systems**



**Free & Open to the Public**  
(513-556-7023; [lauren.jacquot@uc.edu](mailto:lauren.jacquot@uc.edu))