1116-60-2887 **Alex Blocker** and **Xiao-Li Meng*** (meng@stat.harvard.edu), Department of Statistics, Science Center, 7th Floor, Harvard University, Cambridge, MA 02138. *The potential and perils of preprocessing: Building new foundations.*

Preprocessing forms an oft-neglected foundation for a wide range of statistical and scientific analyses. However, it is rife with subtleties and pitfalls. Decisions made in preprocessing constrain all later analyses and are typically irreversible. Hence, data analysis becomes a collaborative endeavor by all parties involved in data collection, preprocessing and curation, and downstream inference. Even if each party has done its best given the information and resources available to them, the final result may still fall short of the best possible in the traditional single-phase inference framework. The technologies driving "Big Data" explosion are subject to complex new forms of measurement error. Simultaneously, we are accumulating increasingly massive databases of scientific analyses. As a result, preprocessing has become more vital (and potentially more dangerous) than ever before. We propose a theoretical framework for the analysis of preprocessing under the banner of multiphase inference. We provide some initial theoretical foundations for this area, including distributed preprocessing, building upon previous work in multiple imputation. We motivate this foundation with two problems from biology and astrophysics, illustrating multiphase pitfalls and potential solutions. (Received September 22, 2015)