

Richard A. **Berk** (Ph.D., Sociology, Johns Hopkins, 1970) is Professor of Criminology and Statistics. He is also Distinguished Professor of Statistics and Sociology, emeritus, UCLA. He is an elected Fellow of the Am. Stat. Assn., the Am. Assoc. for the Advancement of Science, and the Academy of Experimental Criminology; was elected to the Sociological Research Association; and won the Lazarsfeld Award for contributions to methodology from the American Sociological Assn. He has been Editor of *Evaluation Review* for 31 yrs.

Scientific Accomplishment. Berk is the author or co-author of 13 books, over 100 journal articles, and over 50 book chapters. He has published on a wide range of scientific and social scientific topics, with an emphasis on methodological advances and rigorous but practical evaluation of public policy. He was foremost in introducing Heckman's ideas on sample selection bias into sociology and related fields and pioneered social applications of experimentation, e.g., randomized trials for police intervention in domestic disputes (with L. Sherman). A great deal of this thinking is synthesized in a 2003 book on the uses of the regression model. In recent papers: (1) A review of theoretical principles and experience that suggests that randomized field experiments are still the best way to estimate causal effects, but are a considerable distance from perfection (in *J Exper Crim*). (2) An introduction to ensemble methods, sometimes better known as "data mining," (in *Soc Meth Res*); these include classification and regression trees, cross-validation, bagging, random forests, and boosting. The methods are illustrated with reference to the prediction of inmate misconduct within prisons (see also *J Quant Crim*). Berk makes the point that whereas these methods are generally touted as atheoretic, at least from the standpoint of social scientific theory, in fact theory plays approximately the same role as it does in "standard" statistical analysis. (3) A real-world policy application of these methods, in the development of a short screening tool that deputies in the LA Sheriff's Department could use in the field to help forecast domestic violence incidents in particular households. Data came from 500+ households to which sheriff's deputies were dispatched. Information on potential predictors was collected at the scene. Outcomes were measured during a 3-month follow-up. A screening instrument was developed based on a small fraction of the information collected. Making the screening instrument more complicated did not improve forecasting skill. Taking the relative costs of false positives and false negatives into account, the instrument correctly forecasted future calls for service ~60% of the time and future calls involving domestic violence misdemeanors and felonies ~50% of the time. (4) Another example, also in *J Quant Crim*, re-analyzes data used to study the role of race in capital cases in the state of Maryland. When ensemble methods are applied, the racial effects reported in early work turn out to be very fragile. Conventional causal modeling with observational data is not likely to produce robust results for a variety of criminal justice applications.

Present and Future. Berk is working on many projects, several related to population studies. A current working paper focuses on individuals sentenced to probation or parole using data on 60,000+ cases beginning supervision probation and parole supervision in Philadelphia. Statistical learning procedures that take the relative costs of false negatives and false positives into account are used forecast whether a homicide or attempted homicide will be committed. Homicide and attempted homicide are relatively rare crimes, but are among the most serious. Insofar as prospective murderers can be usefully identified, there is the possibility of shifting supervisory and rehabilitation resources to a subset of offenders who may be in greatest need. Other working papers deal with (1) forecasting methods in crime and justice; (2) evaluating the validity of simulation modeling in quantitative criminology; and (3) on the assumptions required for "solving" over-dispersion in the analysis of count data via negative binomial regression. He has a contract with the County of Los Angeles to design a sampling procedure for counting the homeless; and another with the School of Criminology at Florida State to help them do a study of the impact of schooling on later crime. He has begun collaborating (with **Todd** and **Behrman**) on a randomized controlled trial they are proposing in Mexico (with **Wolpin**) on multi-tiered incentives for better school performance. He will start work soon with the ILO to develop research designs for studying human trafficking and the impact of some of the ILO's interventions.

PSC and Research. Berk is the PI of an NSF grant on ensemble methods, He joined Penn from UCLA, where he was a member of the Calif Ctr for Pop Research and the Co-PI of an NSF grant (R. Mare, PI) modeling shifts in residential patterns and changes in residential segregation in the US. His recruitment to Penn was facilitated by the offer of PSC computing funds; these were funds explicitly delegated to the PSC by SAS for maintaining research computing at frontier levels, and demonstrate the utility of this support arrangement. Berk will serve on the Statistical Methods unit of the Methods section of the Development Core.