The 2010 Deming Student Scholar award recipients are:

- Michael Baiocchi, University of Pennsylvania, Wharton,
- Tuan H. Nguyen, Rutgers University, and
- Chia-Hao Wang, University of Pennsylvania, School of medicine.

They will be giving a brief presentation on their work on Monday, December 6, 2010 at 4 pm.

Abstracts:

Near/Far Matching: Quantifying the Benefit of High Level Neonatal Intensive Care Units
Michael Baiocchi

Instrumental variables (IV) is a framework for making causal inferences about the effect of a treatment based on an observational study in which there are unmeasured confounding variables. The most common form of IV estimation is two-stage least squares (2SLS), which works well when the outcome of interest is continuous.
However, in many policy settings the object is to estimate the effect of a treatment on a binary outcome. We propose a nonparametric matching technique - Near/Far Matching - which is capable of estimating population level treatment effects when the outcome is binary. We provide a test statistic with standard errors. We illustrate our method using a study of neonatal intensive care units treatment effect on premature babies born in Pennsylvania.

Modeling and nonparametric methodology for count data in drug studies
Tuan H. Nguyen, Javier Cabrera, and Jos´e Pinheiro

Objectives: To evaluate and compare the statistical operational characteristics of alternative methods of analyzing longitudinal count data in the context of clinical trial data. Typical statistical models for these data are generalized linear mixed model (GLMM), generalized estimating equations (GEE), and nonparametric methodologies (Wilcoxon, Van Elteren tests, and Lehmann-Hodges estimators).
Study Design and Settings: We compare those approaches via a simulation study in term of power, bias, rooted mean squared error, and coverage probabilities. The simulated datasets try to resemble data from a typical trial where observations are only collected during a subset of weeks during a trial. The mean structure consists of patient's characteristics (age, gender), drug-response model, and random effects (week within patient, patient within center, and center). The non-informative missing data mechanism (MCAR, and exponential dropout) is also implemented. We vary the treatment effects and the rate of dropout.
Results: In term of testing hypothesis, GLMMs provide the most statistical powerful tests among all models. In term of estimation, GLMMs and GEEs perform similarly in term of bias and accuracy, while nonparametric method (Lehmann-Hodges estimators) does worse because of missing data due to dropout.
Conclusions: The GLMM (with sandwich estimator for covariance matrix) is the model of preference in our study.

Mediation Analysis Using The G-estimation Approach Under Log-linear Structural Nested Mean Models

Chia-Hao Wang, Thomas R. Ten Have, and Marshall M. Joffe

In randomized clinical trials where the effects of post-randomization factors are of interest such as in mediation analyses, the standard regression analyses are biased due to unmeasured confounding. Causal methods such as the instrumental variables (IV) and G-estimation procedures under structural nested mean models (SNMMs; Robins, 1994, 1997) allow one to make valid inference even if unmeasured confounding is present. However, other assumptions are needed for identifiability and estimation. Under the linear SNMM, two commonly used IV approaches, namely the two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI), have been applied to adjust for confounding and shown to be valid (e.g., Dunn and Benstall 2007; Wang et al. in preparation). However, these IV approaches result in biased estimators under the log-linear SNMM in the presence of unmeasured confounding (Wang et al. in preparation). We consider the G-estimation approach as an alternative solution to remove bias. The G-estimation method was previously developed under either the exclusion restriction assumption or the sequential ignorability assumption. We propose a general framework where these two assumptions are relaxed.

Specifically, we are interested in the direct effect of the treatment adjusting for a post-randomization factor (mediation analysis) and only assume that treatment is randomized but not the mediator. In contrast to the IV log-linear regression methods, we have shown theoretically that the proposed G-estimators are unbiased in the presence of unmeasured confounding. We also present simulation studies confirming that the proposed approach is unbiased with valid confidence intervals. Finally, we illustrate our method in a lung cancer randomized trial for mediation analysis where the sequential ignorability assumption is violated, and the results are discussed and compared to those from the standard log-linear regression approach.