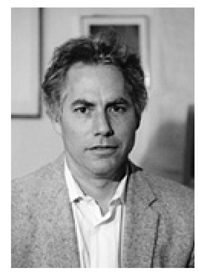
STATISTICS-BIOSTATISTICS SEMINAR

TITLE: **MINIMAX POINT AND INTERVAL ESTIMATORS IN CAUSAL INFERENCE AND MISSING DATA MODELS: APPLICATION OF A UNIFIED THEORY OF PARAMETRIC, SEMI AND NONPARAMETRIC STATISTICS BASED ON HIGHER DIMENSIONAL INFLUENCE FUNCTIONS**

SPEAKER: James Robins
Harvard School of Public Health



TIME: 4:00 P.M.

DATE: Wednesday, November 1, 2006

ROOM: 140 BARDEEN

ABSTRACT:

Suppose given data on a dichotomous treatment T, a response Y, and a high (say 50) dimensional vector of pre-treatment covariates X, one wishes to estimate the average causal effect of T on Y and is willing to assume no unmeasured confounding (ignorability given X). A recent advance is the development of doubly robust (DR) estimators that are n1/2 consistent (the usual parametric rate) if either (but not necessarily both) (i) a working outcome regression (OR) model for the regression of Y on T and X , or (ii) a working propensity model for the regression of T given X are correct . However, DR estimators are inconsistent if, as is inevitable, both working models are misspecified. Further, with high dimensional X, due to lack of power, it is not possible to effectively test whether the working models are sufficiently close to being correct to guarantee small bias. Thus it seems a more honest assessment of uncertainty to use confidence intervals that (i) will include the true treatment effect at their nominal coverage rate under weaker assumptions than for the DR estimators even at the price of shrinking to zero (with increasing sample size) at a rates less than the usual n-1/2 parametric rate. We accomplish this goal by introducing novel rate-optimal estimators and confidence intervals for the treatment effect based on higher dimensional U-statistics. These estimators are derived using a new unified theory of parametric, semi, and nonparametric statistics based on higher order scores (i.e., derivatives of the likelihood) and influence functions that applies equally to both the square-root-n and non-square-root n problems, reproduces the results previously obtained by the modern theory of non-parametric inference, produces many new non-root- n results, and most importantly opens up the ability to perform optimal non-root n inference in complex high dimensional models.

Coffee and Cookies at 3:30 p.m. in Room 1210 MSC