Poster #25
Aaron Miller
PhD Candidate, College of Pharmacy
Pharmacy Practice & Science

Title of Research: Limitations of Case-Control Exact-Matching when Estimating Average Treatment Effects
Other Authors: Linnea A Polgreen Philip M. Polgreen

Introduction/Purpose:
A common technique for estimating treatment effects in non-randomized studies is exact-matching (EM), where controls are identically matched to cases using a set of match variables and then compared using a bivariate statistical test (e.g. Student's t-test). Such exact-matching can be problematic due to the requirement of choosing a limited set of match variables.

Experimental Design:
We demonstrate the potential dangers of EM by estimating the length of stay (LOS) attributable to nosocomial Clostridium difficile infections, using the 2011 HCUP Nationwide Inpatient Sample. We first use common multivariate regression and propensity score matching (PSM) methods to estimate attributable LOS. We then estimate LOS using EM along with three search strategies to select match variables: (1) a naive strategy, using combinations of commonly cited match variables, (2) an informed search, using regression to find the strongest predictors of LOS, and (3) a random search, by randomly selecting sets of match variables.

Results:
Regression and PSM methods produce a consistent set of estimates ranging from 1.94 to 2.89 days. EM using the naive search produced significantly higher estimates averaging 6.1 days. The informed search identified a set of match variables that allowed EM to obtain an estimate consistent with the regression and PSM methods (2.59 days). The random search produced a set of match variables that on average performed slightly better than the naive search (5.17 days).

Conclusions:
In practice EM may be constrained by the need to choose an optimal set of match variables prior to analysis. EM was only able to produce estimates similar to regression and PSM methods after first using regression to select ideal match variables. For many studies where EM is applied, such a data driven selection technique may be impractical. Moreover, the match variables most frequently used in current literature appear to be no better than randomly drawn variables.