Propensity scores

- Sometimes properly randomized trials cannot be or are not used.
- In observational studies, there may be a propensity for some subjects to get the treatment.
- This introduces a bias in the estimation of the treatment effect.
- Covariates may be quite different in the treatment and control group.
- (Covariates should be roughly the same in both groups when randomization is used.)
- The propensity score is the conditional probability of getting the treatment given the covariates.
- It is estimated for each subject using a logistic regression, with the covariates as predictors and a binary response indicating whether or not the subject got the treatment.
- The treatment effect can then be estimated for subjects with similar propensity scores, because their covariates will be balanced.
- Results can be combined over different propensity scores using matching, regression adjustment or stratification.
- This assumes that the treatment assignment and the response are conditionally independent given the covariates.
- Whereas randomization balances all covariates, observed and unobserved, propensity score methods only adjust for measured covariates.
- The fact that the propensity score is a single number (between 0 and 1) simplifies matching or stratification, compared with treating each of the many covariates separately.
- Results may still be biased due to unmeasured covariates.

- A good reference is R.B. D'Agostino, Jr. *Statistics in Medicine*, 17,(1998), "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group"
- We will look at his example on epidural use and Cesarean section.
- The ACT (active management of labour trial) was a randomized study to determine the effect of active management of labour on the rate of Cesarean sections.
- There was a baseline and randomized component to the study.
- They were also interested in whether the use of epidural is associated with C-Sections in nulliparous women.
- Subjects were not randomized as to epidural use.

- They studied 1778 women, of whom 1003 had an epidural.
- Table 4 shows 15 variables of interest.
- Eight of these show significant imbalance between the two groups, using a $F = t^2$ test.
- Propensity scores were calculated for all women using these 15 variables.
- The subjects were then separated into quintiles using the propensity scores.
- Table 4 shows that the variables are now balanced within the quintiles, using two-way analysis of variance.
- To estimate the effect of epidural use on the C-Section rate, they could have estimated the rate within each quintile and then combined (averaged) them.

- Instead they used a logistic regression, with C-Section as the binary response, epidural use as the major predictor, and propensity score and a subset of other covariates as other predictors.
- They conclude the rate of C-Section use is still significantly higher in the epidural group.