IV. Controlling for the presence of additional variables

A. Notation:
1. Add superscript $i$ to tell which table

B. Additional variable provides an alternative explanation for association between disease and exposure: confounding
1. Definition: distortion of disease/exposure association by other factor
   - Other factor related to exposure
   - Other factor causally related to disease
2. Can change direction of relationship: Simpson’s Paradox
   (See example)

E. When do you need to stratify?
1. Heruristically: when stratifier is a confounder
2. Add superscript $i$ to tell which table
   - Corresponding logit will be $\infty$
   - Acceptable since $\lim_{x \to -\infty} x \log(x) = 0$
3. Definition
   - SE of log($\hat{\theta}$) is $1/\sqrt{\sum w_i}$
   - That is, it is related to both exposure and disease
   - Alternative explanation: stress causes
     - stomach upset
     - diseases like headaches for which aspirin is likely treatment.
4. Check for confounding via hypothesis test
   a. Procedure
     i. test for association betw. $C$ and $D$ and betw. $C$ and $E$.
     ii. adjust if these are significant
   b. Use as standard error sum of exact variances.
   c. Intuition might suggest
5. Inference
   a. Aspirin is associated with stomach upset
   b. Does aspirin cause stomach upset?
   c. Alternative explanation: stress causes
      i. stomach upset
      ii. diseases like headaches for which aspirin is likely treatment.
   d. Direction of causation not indicated in an observational study

C. Testing whether common odds ratio is 1
1. Use $T = \sum w_i (X_{1i} - E_{1i})$
   a. Intuition might suggest
   b. We will use $w_i = 1$
   c. Use as standard error sum of exact variances.
   i. Implies assumption that tables are independent.
3. Checking for confounding via hypothesis test
   a. Procedure
     i. test for association betw. $C$ and $D$ and betw. $C$ and $E$.
     ii. adjust if these are significant
   b. Uses significance as a proxy for strength of effect
   c. To make it work at all, typically make very loose criteria for significance confounders
   d. Fails to control Type 1 error

D. Estimation of the common odds ratio
1. Mantel–Haenszel estimator
   - $\hat{\theta} = \exp \left( \frac{\sum w_i \log(X_{0i}X_{1i}/[X_{10}X_{01}])}{\sum w_i} \right)$
2. If $\theta$ for the various strata are different, there is an interaction between the confounder and exposure.
   a. Use Breslow and Day statistic to test homogeneity of odds ratio in a series of $I$ $2 \times 2$ tables:
   b. $C = \sum_i (X_{0i} - \hat{E}_{0i})^2 / \sum_i (1/\hat{E}_{0i} + 1/\hat{E}_{10} + 1/\hat{E}_{11} + 1/\hat{E}_{11} + 1/\hat{E}_{11})^{-1}$
   ii. Agresti says that that generally $C$ is small
   iv. SAS appears to ignore $C$. 

E. When do you need to stratify?
1. Heruristically: when stratifier is a confounder
   a. That is, it is related to both exposure and disease
   b. Empirically, the odds ratio will change if both row and column proportions differ according to stratifier.
2. If $\theta$ for the various strata are different, there is an interaction between the confounder and exposure.
   a. Use Breslow and Day statistic to test homogeneity of odds ratio in a series of $I$ $2 \times 2$ tables: