k. Power is conditional
   i. Unconditionally, is average over values of $O_{ij}$
   ii. Slightly lower than power evaluated at average.
   • Let $\eta = E[\sqrt{O_{ij}}]$
   • Suppose conditional power of form
     $$\Phi(A + B \sqrt{O_{ij}})$$
   • $A + B \sqrt{O_{ij}}$ is well above 0 if power well above 50%
   • Hence expect $A + B\eta > 0$
   • Then unconditional power
     $$E[\Phi(A + B \sqrt{O_{ij}})] = \Phi(A + B\eta) + B\phi(A + B\eta)E[\sqrt{O_{ij}} - \eta - B^2(A + B\eta)\phi(A + B\eta)E[(\sqrt{O_{ij}} - \eta)^2]/2 = \Phi(A + B\eta) - B^2(A + B\eta)\phi(A + B\eta)\text{Var}[\sqrt{O_{ij}}]$$
   • Hence power you get putting in expectation in place of $\sqrt{O_{ij}}$ is lower than unconditional power
   iii. Hence sample size must be slightly higher

B&D: 7.4

3. Scored exposures
   a. Hypotheses
     i. $H_0$: $E[O_{ij}] \propto Q_k$, reflecting differences due to PYAR, but NOT exposure increase.
     • Let $\pi_0^k = Q_k/Q+$.  
     ii. $H_A$: $E[O_{ij}] \propto \pi_1^k$, $\sum_k \pi_1^k = 1$
     b. $T = \sum_{k=0}^{K-1} \pi_k (O_{ik} - \pi_0^k O_{ij})/O_{ij}$
     c. $\mu_0 = 0$, $\sigma_0 = \sqrt{\sum_k \pi_0^k (\pi_k - \pi_0^k)^2}/O_{ij}$
     d. $\mu_A = \sum_k \pi_k (\pi_1^k - \pi_0^k)$

Lecture 10

2. Case-Control Studies
   a. Same as for cohort study, except probabilities are for exposure rather than case
   b. Suppose
      i. $\rho_0 = P[\text{Exposed}|\text{Control}]$
      ii. $\rho_1 = P[\text{Exposed}|\text{Case}]$
      iii. $\zeta = P[\text{Exposed}]$
   c. By Bayes theorem
      i. $\rho_0 = \frac{\pi_0(1-\pi_1)}{\pi_0(1-\pi_0)\pi(1-\zeta)}$
      ii. $\rho_1 = \frac{\pi_1\zeta}{\pi_1\pi_0(1-\zeta)}$
   d. Alternatively, could build test around differences in arcsine transform.
      i. $\mu_0 = 0$, $\sigma_0 = \sqrt{1/(4O_{ij} + 1)}/(4O_{ij})$
      ii. $\mu_A = \arcsin(\sqrt{\pi_1}) - \arcsin(\sqrt{\pi_0})$, $\sigma_A = \sigma_0$.
      iii. Power $\Phi(-z_\alpha + \arcsin(\sqrt{\pi_1}) - \arcsin(\sqrt{\pi_0})/\sqrt{1/(4O_{ij} + 1)}/(4O_{ij})}$

3. Hypergeometric model, conditioning on all marginals
   a. Unfortunately $\mu_A$ and $\sigma_A$ do not have easy expressions.
   b. Recall Cornfield's Approximation:
      i. Solve $E_i = O_i + E_{ij} = O_{ij}$, $E_{i1}E_{i0}/E_{11}E_{10} = \psi$
      ii. $E_{00} = (O_{0+} + O_{00})/2 + \frac{O_{00}}{2(\psi - 1)} - \sqrt{\frac{O_{0+} O_{00} (\psi - 1) + (O_{0+} + O_{00} - O_{00}) (\psi - 1)^2}{(\psi - 1)^3}}$
      iii. $\text{Var}[O_{11}] \approx (\sum_i O_{ij})^{-1}$
      iv. For $\psi = 1$, this is underestimate. See Figures 9 and 10.
   c. Example: Prostate cancer for stage 3 patients
      a. Setup
         i. 217 assigned high dose, 75 assigned low dose, 195

$$\sigma_A = \sqrt{(\sum_k x_k^2 \pi_k^A - (\sum_k x_k \pi_k^A)^2)/O_{ij}}$$
B&D: 7.6

E. Two-Way Studies ($2 \times 2$ tables)

1. Cohort Studies
   a. Suppose
      i. $\pi_0 = P[\text{Exposed}|\text{Unexposed}]$
      ii. $\pi_1 = P[\text{Exposed}|\text{Exposed}]$
   b. Hypotheses
      i. $H_0: \pi_0 = \pi_1 (= \pi_0^0)$
      ii. $H_A: \psi = \pi_1(1 - \pi_0)/[\pi_0(1 - \pi_1)]$
      • Denote alternative probabilities by $\pi_1^A$ and $\pi_0^A$
      iii. Typically write $\pi_0^A$ and $\pi_1^A$ as functions of parameter measuring distance between them and $\pi_0^0$.
      • In our case, function of $\psi$ and $\pi_0$
      • For instance, $\pi_0^A = \pi_0^0$ and $\pi_1^A = \frac{\pi_0^0 \psi}{1 - \pi_0^0 + \pi_0^0 \psi}$
   iv. Power depends on $\pi_0$
      • Typically $\pi_0^0$ between $\pi_0^A$ and $\pi_1^A$
   c. Use standard two–sample binomial test.
      i. $\mu_0 = 0$, $\mu_A = \pi_1^A - \pi_0^A$
      ii. $\sigma_0 = \sqrt{\pi_1^0(1 - \pi_1^0)/O_{0+} + \pi_0^0(1 - \pi_0^0)/O_{1+}}$
      iii. $\sigma_A = \sqrt{\pi_1^0(1 - \pi_1^0)/O_{0+} + \pi_0^1(1 - \pi_0^1)/O_{1+}}$
   d. Alternatively, could build test around differences in arcsine transform.
      i. $\mu_0 = 0$, $\sigma_0 = \sqrt{1/(4O_{ij} + 1)}/(4O_{ij})$
      ii. $\mu_A = \arcsin(\sqrt{\pi_1}) - \arcsin(\sqrt{\pi_0})$, $\sigma_A = \sigma_0$
      iii. Power $\Phi(-z_\alpha + \arcsin(\sqrt{\pi_1}) - \arcsin(\sqrt{\pi_0})/\sqrt{1/(4O_{ij} + 1)}/(4O_{ij})}$

Lecture 10
Lecture 10

5. Matched designs
   a. Inference will be done on discordant pairs
   b. Number of discordant pairs required will be given by binomial sample size described earlier
      i. Need
   c. \( \pi_1 (1 - \pi_0) + \pi_0 (1 - \pi_1) \) is proportion of pairs that will be discordant.
      i. If matching is necessary, \( \pi_1 \) and \( \pi_0 \) will vary over strata
      ii. Want average value of \( \pi_1 (1 - \pi_0) + \pi_0 (1 - \pi_1) \)
   d. Hence divide by this to get necessary number of pairs
   e. Remember that the eventual number is random.
      i. Note also that distribution depends on whether alternative hypothesis is true
      ii. If \( \pi_0 > .5 \), expect more discordant pairs under \( H_0 \)
      iii. If \( \pi_0 < .5 \), expect more discordant pairs under \( H_A \)

F. Monte Carlo Methods
   1. Power calculations are often done using computer simulation.
      a. Randomly recreate data sets under null hypothesis
      b. Test hypothesis of no effect on each data set
      c. Power is proportion of tests rejected
   2. Advantages:
      a. Aren’t tied to things with good normal approximation
      b. Can calculate unconditional power for conditional test easily
      c. Setting it up is more straightforward

3. Disadvantages
   a. Need computer to do calculations
      i. If the number of trials is large, computation time might be long
   b. Results will contain some random variation

V. Bioassay
   A. Preliminary problem:
      1. \( Y_j = \zeta + \epsilon_j \)
      2. \( W_j = \mu + \delta_j \)
      3. \((\epsilon_j, \delta_j) \sim N\left(0, \begin{pmatrix} \sigma^2 & \rho \sigma \tau \\ \rho \sigma \tau & \tau^2 \end{pmatrix}\right)\)
      4. Estimate \( \zeta = \zeta / \mu \)
      5. Consider estimator \( \bar{Y} / \bar{W} \)
   B. Problem: random variable in denominator.
      a. Unfortunately distribution is non-standard
      b. Expectation is
         \[
         n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{(w-\mu)^2}{2\pi \sigma^2} - \frac{(y-\zeta)^2}{2\pi \tau^2}\right) \frac{w}{y} dw \, dy
         = \sqrt{n} \int_{-\infty}^{\infty} \exp\left(-\frac{(y-\mu)^2}{2\pi \sigma^2} + \frac{\mu}{y}\right) dy
         \]
      i. Integral doesn’t converge absolutely.
      ii. Similar to log odds ratio case
7. Two methods for approximating the distribution of the ratio of means
   a. Using delta method, mean and variance of approximating distribution are 
      \( \frac{\xi}{\mu} = \xi \) and
      \[
      \left( \frac{1}{\mu^2} - \frac{\xi^2}{\mu^2} \right) \left( \frac{\sigma^2}{n} \right) \left( \frac{\rho \sigma^2}{n} \right) \left( \frac{\tau^2}{n} \right) = \mu^{-2} n^{-1} (\sigma^2 - 2 \rho \sigma \tau \xi + \tau^2 \xi^2). \]
   b. Exact distribution:
      i. Let \( U = \frac{\bar{W} - t\bar{Y} + \mu - \xi}{\sqrt{\sigma^2 / n + \rho \sigma^2 / n - 2 \rho \sigma \tau / n}} \) and \( V = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \)
      ii. Let \( u = \frac{\sqrt{n} (\mu - \xi)}{\sqrt{\sigma^2 + \rho^2 \sigma^2 - 2 \rho \sigma \tau}} \) and \( v = \frac{\mu}{\sigma / \sqrt{n}} \).
      iii. \( P \left[ \frac{\bar{W} - t\bar{Y} \leq t}{1} \right] = P \left[ \bar{W} - t\bar{Y} \leq 0 \right] + P \left[ \bar{W} - t\bar{Y} \geq 0 \right] \)
        \( = P \left[ U \leq u \right] + P \left[ U \\geq u \right] \)
        \( = P \left[ U \leq u \right] - P \left[ U \\geq u \right] \)
        \( = \Phi(\frac{\sqrt{n} (t - \xi)}{\sqrt{\tau^2 + \rho^2 \sigma^2 - 2 \rho \sigma \tau / \mu}}) + R \)

   for \( |R| \leq \Phi(-\sqrt{n} \mu / \sigma) \). See Figures 11
B. Real aim: confidence intervals
   1. \( \xi \bar{W} - \bar{Y} \sim N(0, \xi^2 \tau^2 / n + \sigma^2 / n - 2 \rho \sigma \xi / n) \)
   2. \( P \left[ \xi \bar{W} - \bar{Y} - (\xi \bar{W} - \bar{Y})^2 \right] \leq z_{\alpha/2} \] = 1 - \( \alpha \).
   3. Set of \( \xi \) satisfying statement inside probability is CI (sort of)
      a. Restriction is quadratic inequality.

This page intentionally left blank.