Problem Set 7

Inference for proportions. According to Fergusson *et al.* (2012), acutely ill patients, including neonatal infants, often receive red blood cell transfusions. However, the consequences of the prolonged storage of red blood cells on health outcomes in premature infants are not well understood. In a double-blinded, randomized controlled trial, the authors looked at health outcomes in neonatal infants who underwent red blood cell transfusions, comparing the standard protocol (transfusions of blood stored for prolonged periods) with fresh blood transfusions (transfusions of blood store for less than seven days).

Specifically, the authors examined five outcomes listed below; as well as a composite outcome, defined as at least one of the five outcomes. In this question, we focus primarily on the composite outcome. The results of the study are shown in the following table:

	Standard	**Fresh**
Necrotizing enterocolitis	15	15
Intraventricular hemorrhage	11	18
Retinopathy of prematurity	26	23
Bronchopulmonary dysplasia	63	60
Death	31	30
Composite Outcome	100	99
Sample Size	189	188

A dataset `hw7.dta` is also available on the course website, if you would rather not use the "immediate" commands in Stata.

Source: Dean A. Fergusson, MHA, PhD; Paul Hébert, MD, MHSc(Epid); Debora L. Hogan, BScN, BA, MScN; Louise LeBel, BScN; Nicole Rouvinez-Bouali, MD; John A. Smyth, LRCPSI; Koravangattu Sankaran, MBBS; Alan Tinmouth, MD, MSc(Clin Epi); Morris A. Blajchman, MD; Lajos Kovacs, MD; Christian Lachance, MD; Shoo Lee, MBBS, PhD; C. Robin Walker, MB,ChB; Brian Hutton, PhD; Robin Ducharme, HBSc; Katelyn Balchin, MSc; Tim Ramsay, PhD; Jason C. Ford, MD; Ashok Kakadekar, MD; Kuppuchipalayam Ramesh, MD; Stan Shapiro, PhD. (2012). Effect of Fresh Red Blood Cell Transfusions on Clinical Outcomes in Premature, Very Low-Birth-Weight InfantsThe ARIPI Randomized Trial. JAMA.

1. Construct a 95% confidence interval for the proportion of infants experiencing the composite outcome in the fresh red blood cell group, using the following methods:

   a) the exact binomial confidence interval

2. Is the normal approximation to the binomial appropriate in this setting?

3. Suppose you wanted to calculate a 95% confidence interval for infants experiencing intraventricular hemorrhage after receiving a fresh blood transfusion as well. Is the Wilson confidence interval still appropriate?

4. Estimate and construct a large-sample 95% confidence interval for the risk difference for experiencing the composite outcome for those with fresh blood versus the standard protocol blood. Calculate the risk difference as estimated proportion in fresh blood group minus estimated proportion in the standard blood group.

5. Use a two-sample test of proportions to determine whether there is a difference between fresh and standard groups at the $\alpha=0.05$ level of significance. What is the test statistic? Null distribution? P-value? Conclusion?

Contingency Tables. Continue using the Fergusson et al. (2012) clinical trial data from the previous problem to complete the following questions.

1. Estimate the odds ratio and a 95% confidence interval for experiencing the composite outcome for those with fresh blood versus standard protocol blood. Is there evidence of an association between blood group and the composite outcome (at the 0.05 level of significance).

2. Construct a 2x2 table for the composite outcome versus blood group. Are the expected cell counts large enough to conduct a Pearson Chi-square test?

3. Using the Pearson chi-square test, determine if there is an association between fresh versus standard blood and the composite outcome at the $\alpha=0.05$ level of significance. What is your test statistic? Null distribution? P-value? Conclusion?

Final Thoughts.

1. In the previous questions, we looked at three different tests of association: the Pearson Chi-square test, an odds ratio test, and a risk difference test. Are the results of these three tests consistent? Would you expect them to be?

2. Is there evidence of an association between blood group assignment and the composite outcome?

3. Think back to the Bonferroni correction from last week. If you were tasked with conducting hypothesis tests comparing the two blood groups for **each** of the 5 different outcomes, would you need to correct for multiple comparisons?

4. In this study, the authors state that they powered the study to detect an absolute difference of 15% in the two groups with 80% power, used a 2-sided test with $\alpha = 0.05$. After a few more adjustments, their final sample size calculation was 450.

   Now suppose you want to replicate the study using a different population. Given that the authors did not find an association in their data, you decide to increase the power and decrease the difference detected between standard and fresh groups. Using an equal number of infants in both groups, what is the total sample size needed in order to achieve 90% power, assuming that the proportion of infants experiencing the composite outcome in the standard group was 55% and 45% in the fresh blood group (again using a 2-sided test with $\alpha = 0.05$).

5. Consider a covariate, the clinical risk index for babies (CRIB), which was measured in the infants enrolled in the clinical trial. CRIB is usually associated with the composite outcome. From the baseline characteristics table in the Fergusson et al paper, we find that the median and IQR for CRIB is similar between the standard and fresh blood groups. This suggests that the distribution of CRIB is similar in both groups.

   True or False: Because the distribution of CRIB is similar betwen groups, the study investigators would not have gained any power to detect an effect by matching on CRIB score.

Randomized Clinical Trial versus Cohort Study

1. The benefit of a randomized clinical trial over an observational cohort study is that, in large enough samples, the groups are identical with respect to
    A. other extraneous factors that are associated with the outcome of interest
    B. factors that would make the results more generalizable to the larger population
    C. other factors related to the likelihood of participating in a study

Toxins and Parkinson's Disease Cohort Study

An investigator, Dr. Park, is interested in evaluating whether there is an association between exposures to toxins and risk of developing Parkinson's disease. She constructs a cohort of men and women that are living in her state in 1985 and every year, they are asked to complete a questionnaire about exposures at work and at home and whether they have been diagnosed with Parkinson's disease. The participants in Dr. Park's study contribute information on their changing toxin exposures over time for as long as they are residents in that state and for any year that they complete the questionnaire. As new people move into the state, they are enrolled in her cohort and remain in the cohort for as long as they are residents in the state and complete the questionnaire.

1. Should Dr. Park include people who reported that they had Parkinson's at the time that they were recruited?
    A. Yes
    B. No

2. Does Dr. Park need to be concerned about selection bias?
    A. Yes, if people who were at greater risk of developing Parkinson's were more likely to be exposed and were also more likely to participate in the study
    B. No, because it is a prospective cohort study so selection biases are not a concern

3. If those who are exposed to toxins are more likely to drop out of the study and more likely to develop Parkinson's disease, what effect with this have on the estimated relative risk compared to the true relative risk?
    A. No effect
    B. Biased towards the null
    C. Biased away from the null

4. Is this an open cohort or closed cohort? Why?
    A. Closed, because exposure at baseline is used for all follow-up
    B. Open, because loss to follow-up and competing risks are still a problem in this population-based study
    C. Closed, because risk ratios are the most appropriate measure to compare toxin levels and Parkinson's risk
    D. Open, because people can enter and exit the cohort over the follow up time of the study

5. True or False: Based on the data collected in her study, Dr. Park will be able to calculate absolute measures of Parkinson's disease incidence.

6. Because Dr. Park conducted a prospective cohort study instead of a cross-sectional study, she can be less concerned about
    A. Confounding
    B. Bias
    C. Reverse causation
    D. Chance