Name: _____ SID: _____

Collaborators: _____

# Week 7 Problem Set

Tests and measures of association
PHW142

You must put your name and SID at the top of the page.

Please include:

- explanations of your reasoning (even if we forget to ask over and over)
- the formulas and major steps when the question asks you to do this
- relevant R code and results
- interpretations in the context of the problem scenario

**This problem set is worth 20 points.** Please submit this file to bCourses as a PDF file.

There is a table on page 12 of the Week 7 Reader that lists all the R functions and the packages in which they are found.

Hint: If you are getting a syntax error (i.e. R is saying you don't have the right arguments), try calling the specific function from the package with the syntax `package::function()`. For example,

```
epitools::binom.approx()
```

## Part 1. Intensive treatment for type 1 diabetes: the DCCT study

1. **(11 points)** For a long time, the medical community has been divided about how aggressively to control the blood glucose level of type 1 diabetes. One part of the Diabetes Control and Complications Trial (DCCT) study randomly assigned volunteers with type 1 diabetes with retinopathy (damage to the retina that can lead to blindness) before the study began to either the standard of care treatment or to a more intense treatment aimed at maintaining a blood glucose level as close to normal as possible. The health of both groups was closely monitored for 6 years.

   For some participants, their retinopathy continued to get worse (progression).

| | progression | no progression | total |
|---|---|---|---|
| conventional tx | 143 | _____ | 352 |
| intensive tx | 77 | _____ | 363 |

   1.1 Start by filling in the missing values in the table above. **0 points**

   .

1.2 Calculate the proportion of patients who progressed in each group. **1 point**

1.3 Use the conditional probabilities you calculated in question 2 to estimate the risk difference and the relative risk, considering the conventional treatment as the ”exposure.” **2 points**

1.4 State the null and alternative hypotheses for the $\chi^2$ test in terms of the risk difference and in terms of the relative risk. **2 points**

1.5 Check that the conditions to use the $\chi^2$ distribution to find the P value are satisfied. **2 points**

1.6 Create a matrix of the counts in R. (Use the format that is suited to using the `epitab()` function to find a confidence interval for the relative risk.) Carry out the $\chi^2$ test.

See page 43 in the Week 7 Reader for how to set the matrix up for `epitab()` .

Paste your R functions and results here: **1 point**

1.7 Interpret the $\chi^2$ test: State the P value and the conclusion in the context of this problem, using either the relative risk or the risk difference. (Use $\alpha = .05$ ) **1 point**

1.8 Find a 95% confidence interval for the relative risk. **1 point**

Paste your R `epitab()` function results here:

You should recognize P(progression | conventional tmt) and P(progression | intensive tmt) in the output.

1.9 Write a short summary for the confidence interval for the relative risk. **1 point**

## Part 2. Blunt force pancreatic injury and complications

2. **(5 points)** In a small pilot study, researchers studied 26 children with blunt force injuries to the pancreas. 19 of the patients were classified as having minor injuries, and 7 were classified as having major injuries. In the major injury group, 6 of the 7 developed life-threatening complications, compared to 3 of the 19 in the minor injury group.

2.1 Find the expected counts under the null hypothesis of independence and use them to explain why the conditions to use the $\chi^2$ distribution to find the P value are not satisfied. **2 points**

2.2 Carry out Fisher's exact test for association. Use a two-sided alternative hypothesis. You will have to create a matrix of counts in R. Paste your R code and results here: **1 point**

2.3 Interpret the results of the Fisher test: State the P value and the conclusion in the context of this problem. (Use $\alpha = .05$) **1 point**

2.4 Use the Agresti-Caffo 'plus 4' method to find a 95% confidence interval for the difference of the population proportions developing life-threatening complications. Paste your R code and results: **1 point**

Write a summary sentence, bringing together your test conclusions and the confidence interval.

# Part 3. Association between "heart attack" (myocardial infarction, MI) and type 2 diabetes

3. **(4 points)** Investigators had a sample of 144 people who had a heart attack (MI). For each person who had a heart attack, the investigators selected a control individual of the same gender and age who was free of heart disease. So, the study has 144 matched pairs.

   All of the study participants were screened for type 2 diabetes.

   The investigators hypothesized that the individuals who had an MI were more likely to have type 2 diabetes.

   Here are the data for the 144 pairs:

| | no MI | |
|---|---|---|
| | type 2 diabetes | no type 2 diabetes |
| MI | | |
| type 2 diabetes | 9 | 37 |
| no type 2 diabetes | 16 | 82 |

3.1 What are the investigators' null and alternative hypotheses?

Be sure to phrase the hypotheses in terms of the discordant pairs in the table. **1 point**

3.2 Use the function `binom.exact()` to find the P value for the test. Paste your R code and results here: **1 point**

3.3 What do you conclude? Explain fully. **2 points**