# Adjusting for clustering of the Mantel-Haenszel risk ratio

Student Name: Yehia Jadayel (yjada048@uottawa.ca)

Supervisor: Dr. Gilles Lamothe

University of Ottawa, Faculty of Science

## Introduction

In the basic science for public health that is epidemiology, there exists a measure of how likely people are prone to having a certain outcome/disease of interest given that they have been exposed to a specific factor; the common example being a patient with lung cancer that has been exposed to smoking. This is known as the risk ratio. In this research project, we shall be calculating the Mantel-Haenszel risk ratio: a risk ratio adjusted for similarities between groups of people in a population. It is a very popular form of the risk ratio because it properly average's the risk ratio in a way so that all the data is properly weighted. It is a stratification method that accounts for possible confounding of other factors.

Aside from these confounding variables, there can exist variables, e.g. communities, which can have random effects on the association between the outcome and the exposure. This is known as clustering. It usually translates into less precise estimates of the risk ratio.

The purpose of this project is to write a program to compute the Mantel-Haenszel risk ratio to account for confounding variables and to adjust the estimate for the effects of the clustering variables.
The idea of adjusting for clustering of the Mantel-Haenszel ratios has already been done with another type of epidemiological ratio, known as the Odds Ratio. This has been previously accomplished by Dr. Gilles Lamothe, and essentially, executing this adjustment on the risk ratio is the goal of this project.

## Results

After running the final program for multiple exposures, we found that all of the confidence intervals gave a slightly larger confidence interval than the original risk ratio. Some confounding variables had a bigger impact on the increase of the interval more than others. The variation in terms of how much bigger the confidence interval became was dependent on whether the clustering variables were relevant to the certain type of exposure. The ratio of how much bigger the interval became is known as the cluster effect. The increase of variability also signifies the differences between communities within the Bosnian population.

Also, this new adjustment shows that some conclusions that came from the unadjusted ratios that were originally significant are actually not significant. This is because after the adjustment, some confidence intervals straddle 1. This signifies that a conclusion cannot be made as to whether an exposure to a factor increase or decreases the chances of having a certain outcome.


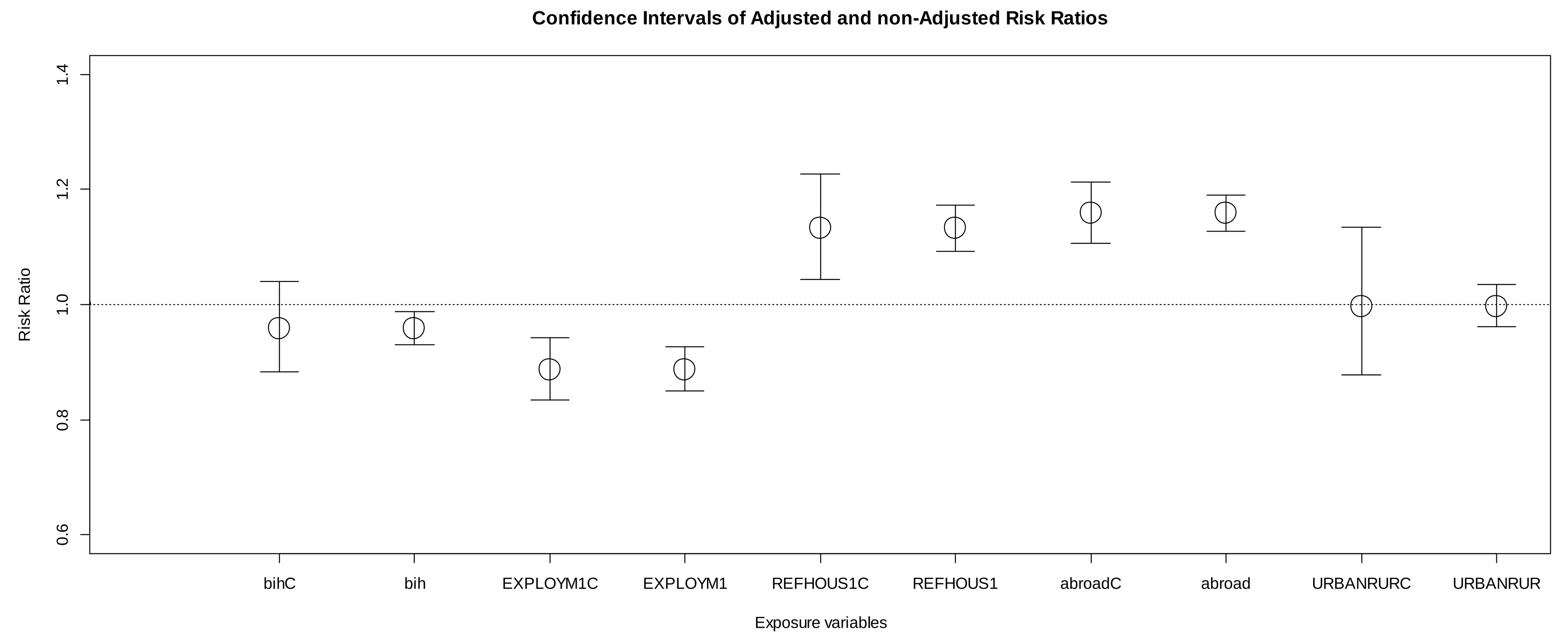


Figure 1: Confidence intervals of Risk Ratios plotted for each exposure variable. Exposures with a "C" at the end signify adjustment for clustering.

## Methodology

The bulk of this project was done with R: an open source statistical programming software used by statisticians to obtain clear and specific results. The research project was divided into numerous steps in order to reach the goal of adjusting the risk ratio.

First step was to write a script to calculate the risk ratio. Essentially, a very simple program was written to properly obtain the risk ratio for a certain population.

A script was then written to calculate the Mantel-Haenszel Risk Ratio with confidence interval by including confounding variables in the calculations. Confounding variables are variables that often distort statistical ratios due to their hindering effects on the data. The advantage of using the very popular Mantel-Haenszel method means that our results would control for these confounding effects.

The delta method was used to write a script to calculate the Mantel-Haenszel risk ratio with confidence interval on a real set of a data while adjusting for clustering. The set of data was collected by CIET in the year 1995-6 in Bosnia-Herzegovina, consisting of variables such as geographical placement, whether the subject is employed or not, whether the subject lives in a rural or urban area, etc. Also, the script allows for a varying number of confounding variables as well.

The Final program was then implemented on the Bosnian data with interest of whether certain groups within the population had the outcome of interest, which was whether they had sufficient food to satisfy their families.

| Risk (exposure) | Disease Status (outcome) | |
|---|---|---|
| | Present | Absent |
| Smoker | A | C |
| Non smoker | B | D |

Figure 2: Table showing the division of the groups for the calculation of a risk ratio.

Risk Ratio = $\dfrac{\text{Having the outcome of interest and exposed to factor of interest}}{\text{Having the outcome of interest without being exposed}}$

$$Risk\ Ratio = \frac{A/(A+C)}{B/(B+D)}$$

The Mantel-Haenszel Risk Ratio

$$RR_{MH} = \frac{\sum_{i=1}\left(\frac{b_i+d_i}{n_i}\right)a_i}{\sum_{i=1}\left(\frac{a_i+c_i}{n_i}\right)b_i}$$

* Where n is the number of subjects in a specific strata.

## Conclusion

We adjusted the Mantel-Haenszel risk ratio to account for factors with random effects such as communities. Now statisticians will have data that will truly take into account all the effects of the variables that affect the risk ratio and its confidence interval, making the data much more trustworthy and sensible. In essence, the goal of creating a more robust risk ratio has been accomplished.

## References and Acknowledgments

Author : Yehia Jadayel (yjada048@uottawa.ca)
Special thanks to Project Supervisor: Dr. Gilles Lamothe
Reference: CIET Food Security in Bosnia and Herzegovina