

Association between Breast Cancer and Hepatitis C Virus with Age Controlling

Amy Traianou
Department of Statistics
University of Connecticut

November 14, 2022

Abstract

In Egypt Hepatitis C virus infection is a common problem that often impedes treatment of Breast Cancer in female patients. While previous work has analyze an associaiton between HCV seropositivity and breast cancer diagnosis, this paper will focus on the potential confounding of age on the association. We will use Fisher's Exact Test to estimate the crude estimate of the association and the Mantel-Haenszel method to stratify the data between ages. The possible effect of age in the association can lead to improved care for patients.

KEYWORDS: Breslow-Day; Fisher's Exact Test; Mantel-Haenszel.

1 Introduction

This paper will explore the risk factors associated with breast cancer in women. Breast cancer is extremely prevalent and researchers are constantly trying to determine risk factors to identify women for preventative exams. There are many studies attempting to determine if there is an association between certain risk factors and breast cancer including chronic hepatitis C infection (?) and hepatitis B (?).

Specifically, this paper will determine if there is an association between testing positive for hepatitis C virus and breast cancer while controlling for age in Egyptian populations. Both Hepatitis C and breast cancer are extremely prevalent in Egypt (?). Very recently on October 6th, Egypt and Qatar agreed to collaborate in the health sector and use each other's expertise. The health ministers specifically mentioned research in Hepatitis C and breast cancer as a area of interest (?). Thus, research surrounding these diseases is becoming more important in the field.

The rest of the paper is organized into several sections. Section ?? describes the study done in 2020 to test if HCV seropositivity is associated with breast cancer diagnosis in Egyptian population. Section ?? describes the following statistical methods; Fisher's Exact Test, Breslow-Day test, and Mantel-Haenszel method. Section ?? applies the methods to data from the study, including stratified between ages. Section ?? concludes the paper.

2 Data

The data used in this paper is from a 2020 retrospective case-control study based on Egyptian female populations (?), sponsored by Mansoura University in Egypt.

The study group consists of 405 patients treated at the Oncology Center - Mansoura University in the past 10 years. In order to be included, invasive breast cancer must be biopsy-proven. Patients with unknown viral marker status, multiple cancer diagnoses and virus-unrelated hepatic pathology were all excluded.

The control group consists of data from a previous study, conducted from 2015-2017. There are 145 females from the same geographic distribution, all with no previous cancer diagnosis. In both groups, all patients are above the age of 18. For this paper, patients are stratified based on age- older patients are considered 45 years or above and younger patients are below 45 years. The data is summarized in the following 2x2 contingency tables.

2x2 Contingency Table for all patients:

| | Breast Cancer | No Cancer Diagnosis | Total |
|------------------|---------------|---------------------|-------|
| HCV Seropositive | 88 | 15 | 103 |
| HCV Seronegative | 317 | 130 | 447 |
| Total | 405 | 145 | 550 |

2x2 Contingency Table for patients younger than 45:

| | $Y = 1$ | $Y = 0$ | Total |
|---------|---------|---------|-------|
| $X = 1$ | A | C | A+C |
| $X = 0$ | B | D | B+D |
| Total | A+B | C+D | N |

Figure 1: 2x2 Contingency Table

$$P_a = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{a+b+c+d}{a+c}} = \left[\frac{(a+b+c+d)!a!b!c!d!}{(a+c)!(b+d)!(a+b)!(c+d)!} \right]^{-1}$$

Figure 2: Hypergeometric pmf

| | Breast Cancer | No Cancer Diagnosis | Total |
|------------------|---------------|---------------------|-------|
| HCV Seropositive | 17 | 2 | 19 |
| HCV Seronegative | 110 | 91 | 201 |
| Total | 127 | 93 | 220 |

2x2 Contingency Table for patients older than 45:

| | Breast Cancer | No Cancer Diagnosos | Total |
|------------------|---------------|---------------------|-------|
| HCV Seropositive | 71 | 13 | 84 |
| HCV Seronegative | 207 | 39 | 246 |
| Total | 278 | 52 | 330 |

3 Methods

To test the association of breast cancer and Hepatitis C seropositivity, one can use the chi-square or Fisher's Exact test, depending on the conditions from the sample size (?). Fisher's exact test is useful for when the normality assumption is violated and the expected values of the 2x2 table are too small. The test uses the hypergeometric distribution to test if the probabilities are the same between the two groups. Thus, we can determine if there is more

| | $Y = 1$ | $Y = 0$ | Total |
|---------|---------------------------|---------------------------|---------------------------|
| $X = 1$ | A_i | C_i | $A_i + C_i \equiv M_{1i}$ |
| $X = 0$ | B_i | D_i | $B_i + D_i \equiv M_{0i}$ |
| Total | $A_i + B_i \equiv R_{1i}$ | $C_i + D_i \equiv R_{0i}$ | N_i |

Figure 3: 2x2 Contingency Table with Stratified Data

$$\chi_{BD}^2 = \sum_{i=1}^S \frac{[A_i - E(A_i|\widehat{OR}_{M-H})]^2}{\text{Var}(A_i|\widehat{OR}_{M-H})}.$$

Figure 4: Breslow Day Test Statistic

$$\chi_{M-H}^2 = \frac{\left(\left|\sum_{i=1}^S A_i - \sum_{i=1}^S E(A_i)\right| - 0.5\right)^2}{\sum_{i=1}^S V(A_i)}$$

Figure 5: Mantel-Haenszel Test Statistic

of a risk of breast cancer for those with seropositive Hep C. In the data section, the data was summarized into 2x2 contingency tables. The generalized table is shown in Figure ?? . The first step is calculate the probability of the original table occurring, using the hypergeometric pdf. Figure ?? shows the formula used.

Based on the alternative hypothesis, you then calculate the probability of all more extreme tables based on the a value. In this paper, p1 is the probability of having breast cancer given that the patient is seropositive and p2 is the probability of having breast cancer given that the patient is seronegative. Then, we want to determine if there is an association, specifically if the risk of having breast cancer is higher for those who are seropositive. The alternative hypothesis is that p1 is greater than p2- which will be tested in the application section.

After calculating the crude estimate, a Breslow-Day test will test for homogeneity of odds ratios between the age stratified data. The generalized form of stratified 2x2 contingency tables is shown in Figure ?? . The Breslow-Day test statistic is shown in Figure ?? , which will be computed using SAS.

Based on the results of the Breslow Day test for homogeneity of odds ratios, the Mantel-Haenszel test will compute a common odds ratio and test if there is an association between seropositivity and breast cancer while controlling for age. The Mantel-Haenszel test statistic is shown in Figure ?? where the expected value of A is

$$E(A_i) = \frac{M_{i1}R_{1i}}{N_i} \tag{1}$$

and the variance of A is

$$V(A_i) = \frac{R_{1i}R_{0i}M_{1i}M_{0i}}{N_i^2(N_i - 1)}. \tag{2}$$

4 Application

-

5 Discussion

-