Association between Breast Cancer and Hepatitis C Virus with Age Controlling

Amy Traianou

Department of Statistics

University of Connecticut

December 5, 2022

Abstract

In Egypt, Hepatitus 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

Breast cancer is the world's most prevalent cancer and occurs in every country in women at any age after puberty. Therefore, it's extremely important to determine certain risk factors that can help detect breast cancer early and save lives. Research is constantly being performed to identify risk factors from lifestyle factors to genetics (**?**). Family history is shown to be a strong predictor for women developing breast cancer.

However, recent research is looking into how certain diseases can be a risk factor for breast cancer or increase the agression of the cancer. Research by the Department of Endocrine and Breast Surgery in China focused on Hepatitis B and it's association with breast cancer and found HBV is potentially affecting estrogen levels, potentially leading to a new way of screening for women (**?**). Additionally, some work has been done with Hepatitis C seropositivity in France, but the control group was women with other liver diseases (**?**). Thus, the results can only compare women with certain liver diseases, such as hepatitis B, alcoholic liver disease, and hemochromatosis to women with Hepatitis C.

Egypt in particular has the highest prevalence of Hepatitis C, and breast cancer is the most common cancer. Thus, an association between the two would be extremely benficial. So, this paper will examine if there is an association between testing positive for hepatitis C virus and breast cancer diagnosis while controlling for age in Egyptian populations. Age is a knwon and accepted risk factor of breast cancer; older women are more likely to develop tumors (**?**). Therefore, controlling for age is necessary.

Inconsistency of Hepatitis C prevalence in different regions of Egypt makes it difficult to generalize research to the entire country. This paper will focus on a small region within the country with high prevalence (**?**) and make conclusions relevent to a specific subgroup.

Recently on October 6th of 2022, Egyptian Minister of Health Abdel-Ghaffar and Qatari Minister of Health Hanan Al-Kuwari met in Doha and agreed to use each other's expertise in the health field. The health ministers specifically mentioned research in Hepatitis C and breast cancer as an area of interest (**?**). This meeting emphasizes the importance of diving

into Hepatitus C Virus and it's association with breast cancer. The Egyptian government has defined breast cancer as an area of concern and desires to conduct research within their population, which is very minimal at the moment.

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 stratafied 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 (**?**), sponsered by Mansoura University in Egypt. The study was conducted with several goals. The main goals we are interested in are 1. to test if seropositivity is associated with breat cancer diagnosis in Egyptian population and 2. to test if HCV-positive patients harbour more agressive breast tumors than HCV-seronegative patients.

The study group consists of 405 patients treated at the Oncology Center - Mansoura University in the past 10 years. The data was retrieved from the hosipital filing system. In order to be included, the patients' invasive breast cancer must be biopsy-proven. Patients with unknown viral marker status, multiple cancer diagnoses and virus-unrelated hepatic pathology were all excluded. Additionally, the study measured two outcome measures. The primary measure was the prevalence of HCV seropisitivity, which is defined as being Seropositive for the duration of the study, an average of 6 months. The secondary outcome measure was the number of disease related events related to breast cancer diagnosis. Disease-free survival was calculated from the time of diagnosis to the last date of follow up for the study.

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. The group consists of healthy volunteers or hosipital patients with no cancer diagnosis.

In both groups, all patients are above the age of 18. The ratio of study-to-control was intended to be 1:3. 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 Table **??**.

Table 1: 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

Of the 550 women who were involved in the study, 220 were below the age of 45 and 330 were of age 45 or above. Table **??** and Table **??** show the break down of seropositivity and seronegativity for the stratified age groups.

Table 2: 2x2 Contingency Table for pateints younger than 45

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

Table 3: 2x2 Contingency Table for pateints 45 and older

	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 of a risk of breast cancer for those with seropositive Hep C.

The assumptions of Fisher's Exact Test are that all margins of the table are fixed. The total number of subjects with and without the risk factor, (A + C) and (B + D) must be fixed. The total number of subjects with and without the event (A + B) and (C + D) must also be fixed. This is equivalent to conditioning on the marginal values.

In the data section, the data was summarized into 2x2 contingency tables. The generalzied table is shown in Figure **??**. The first step is calculate the probability of the original table occuring, using the hypergeometric probablity mass function:

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

The hypergeometric distribution is a discrete distribution, so we use the probability mass function for calculations where A is the random variable representing the count in the (1,1) cell.

After finding the probability of the original table, you then calculate the probability of all more extreme tables based on the a value. Fisher's exact test is often a one-sided test based on the research being conducted. In this paper, p1 is the probability of having breast cancer given that the patient is seropositive and p2 is the probability of having breat 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

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

Figure 1: 2x2 Contingency Table

	$Y = 1$	$Y = 0$	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 2: 2x2 Contingency Table with Stratified Data

section.

After calculating the crude estimate without age stratfied, a Breslow-Day test will test for homogeneity of odds ratios between the age stratified data, younger than 45 and age 45 or older. The generalized form of stratified 2x2 contingency tables is shown in Figure **??**. The risk factor is in the rows while the event of interest is in the columns. Each strata forms its own table that is used for calculations. The Breslow-Day test statistic will be computed using SAS with the formula:

$$\chi^2_{BD} = \sum_{i=0}^{s} \frac{[A_i - E(A_i|\hat{OR}_{M-H})]^2}{Var(A_i|\hat{OR}_{M-H})} \tag{2}$$

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 can only be conducted if the homogeneity of odds ratios is not rejected. The Mantel-Haenszel test statistic is:

$$\chi^2_{M-H} = \frac{(|\sum_{i=0}^{s} A_i - \sum_{i=0}^{s} E(A_i)| - 0.5)^2}{\sum_{i=0}^{s} V(A_i)} \tag{3}$$

The expected value and variance of A for the ith strata is

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

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

This procedure only focuses on odds ratios so it can be used for prospective, retrospective, and cross-sectional studies. Additionally, it focuses on the (1,1) cell on the table just like Fisher's Exact test. It should be noted that the test statistic includes a continuity correction while SAS computes the test statistic without a continuity correction.

4 Application

As stated in the methods section, we will begin with Fisher's Exact test to determine the relationship between HCV seropositivity and breast cancer without controlling for age. Then, we will conduct a Breslow-Day test to test for homogeneity of odds ratios between the age stratified data. Finally, we will perform a Mantel-Haenszel test to compare the strata.

Fisher's exact test for the pooled data of all patients tested

$$H_0 : p_1 = p_2, H_a : p_1 > p_2$$

at a 0.05 level of confidence.

The test resulted in a p-value of 0.0027, which is less than alpha = 0.05. Thus, the null hypothesis is rejected and there is evidence suggesting patients who are seropositive have a higher breast cancer rate than those who are seronegative without controlling for age.

Next, the Breslow Day test checks for homogeneity between the stratified ages.

$$H_0 : OR_1 = OR_2, H_a : OR_1 \neq OR_2$$

at a 0.01 level of confidence. Where OR 1 is the odds ratio of women who did and did not develop breast cancer with seropositivity for women younger than 45 and OR2 is the odds ratio of women who did and did not develop breast cancer with seropositivity for women 45 years and older.

The test resulted in a p-value of 0.0133, which is greater than the alpha value of 0.05. Thus, we fail to reject the null hypothesis and conclude there is not ennough evidence to suggest the strata have different odds ratios and proceed with the Mantel-Haenszel procedure.

The Mantel-Haenszel test produced a common odds ratio of 1.6847 and a 95 percent confidence interval of (0.9283, 3.0574). The interval contains 1 so after controlling for age, the association is not supported. Addtionally, we will test the following hypothesis using the Mantel-Haenszel test statistic:

$$H_0 : OR_{common} = 1, H_a : OR_{common} \neq OR_1$$

The calculated p-value is 0.0773. Therefore, the fail to reject the null hypothesis and there is not enough evidence that there is an association between HCV seropositivity and breast cancer diagnosis after controlling for age.

However, the individual odds ratios for age groups are worth noting. For the older group (45 and older) the odds ratio estimate and interval are 1.0290 and (0.5197, 2.0374). The odds ratio estimate and interval for the younger group (less than 45) are 7.0318 and (1.5828, 31.2399). Therefore, the association between seropositivity and breast cancer is much stronger for females less than 45 years.

5 Discussion

Based on the analysis, it is evident that for women under the age of 45, seropositivity and breast cancer are strongly associated. As being HCV positive can complicate treatment of cancer, this is critical information. In Egypt, breast cancer remains the most common form or cancer and high levels of hepatitus C continue.

There are several limitations to the study that call into question the validity of the results. The small sample size is a result of lack of available data. Therefore, it is hard to find a meaningful association. The study was retrospective, which limited analysis to the odds ratio and not risk difference or relative risk. The method of data collection of the control group is also not ideal. Using historical data can cause issues as direct populations are not being compared. Thus, future studies should aim to be prospective cohort studies so researchers can observe how seropositivity and breast cancer interact over time.

Additionally, age is just one factor that can affect the associaiton between hepatitus C and breast cancer. Future studies should dive deeper into more factors such as weight, reproductive history, and alcohol consumption. Addiitonally, future studies can explore how seropositivity affects tumor size, disease agression, and more. As breast cancer is so prevalent in Egypt, further understanding the risk factors is becoming even more important. The collaboartion between Egypt and Qatar is very promising for emerging research in the health sector. With resources and knowledge from Qatar, it is hopeful that Egypt can research further into Hepatitis C and breast cancer for the good of their population.