Yuefu Jiang

Biostatistics Final Project

Cervical Cancer Risk Factor Analysis

Introduction

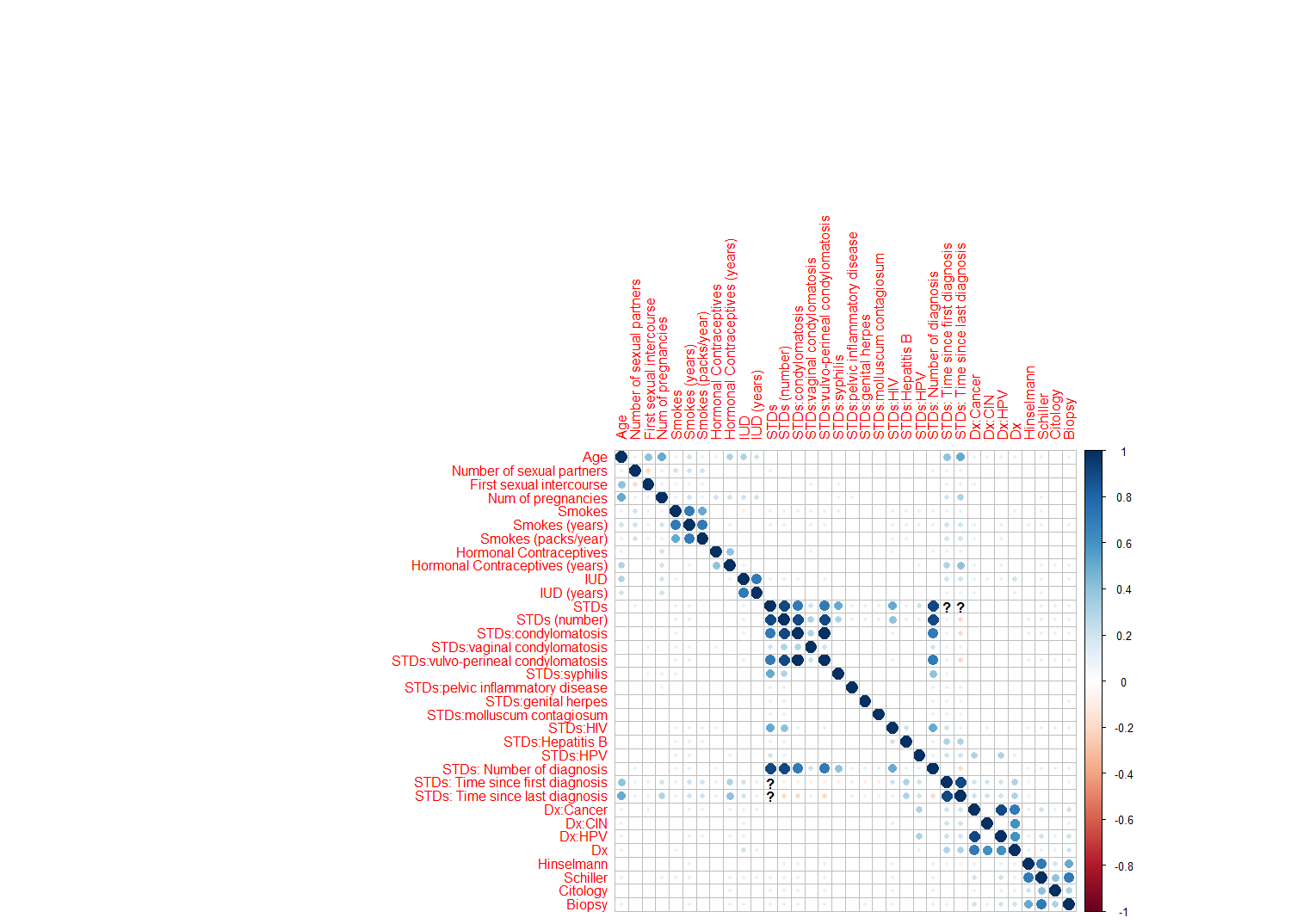
Cervical cancer was a major cause of death in United States in the last century [5]. Although the medical practice to treat this disease has advanced during last few decades, it still has an alarming high diagnosis/death rate, especially in developing countries. Reasons behind this high mortality besides the malicious nature of cancer were believed to be low awareness of gynecological diseases as well as lack of cancer screening. In this analysis, a data set containing 36 attributes from 858 patients in Venezuela was analyzed in order to identify some plausible risk factors of cervical cancer. According to recent researches, HPV infections was one of the major risk factors contributing to the development of the disease; it was also believed that smoking, number of pregnancies as well as some methods of birth control employed could contribute to the acquirement cervical cancer. In this analysis, such factors, along with other factors will be analyzed statistically in seek of significant relations. On the other hand, most of the modern analyses were done in developed countries such as US; as this data set was collected in Venezuela, a developing country, it was plausible that some risk actors may differ from previous results gained in developed countries. The analysis has also included the variance between diagnostic methods of cervical cancer in relation of HPV infection and IUD usage.

Methods

Correlation matrix was built upon all attributes to select highly correlated attributes for further analysis, excluding HIV diagnosis and cervical condylomatosis as no observation across all patients could be found. Spearman correlation was employed as many diagnostic attributes were binary. For following statistic analysis on specific attributes and cervical cancer diagnosis, general linear model was built as logistic regression was performed at alpha = .05. All analysis was done under R 3.5.1.

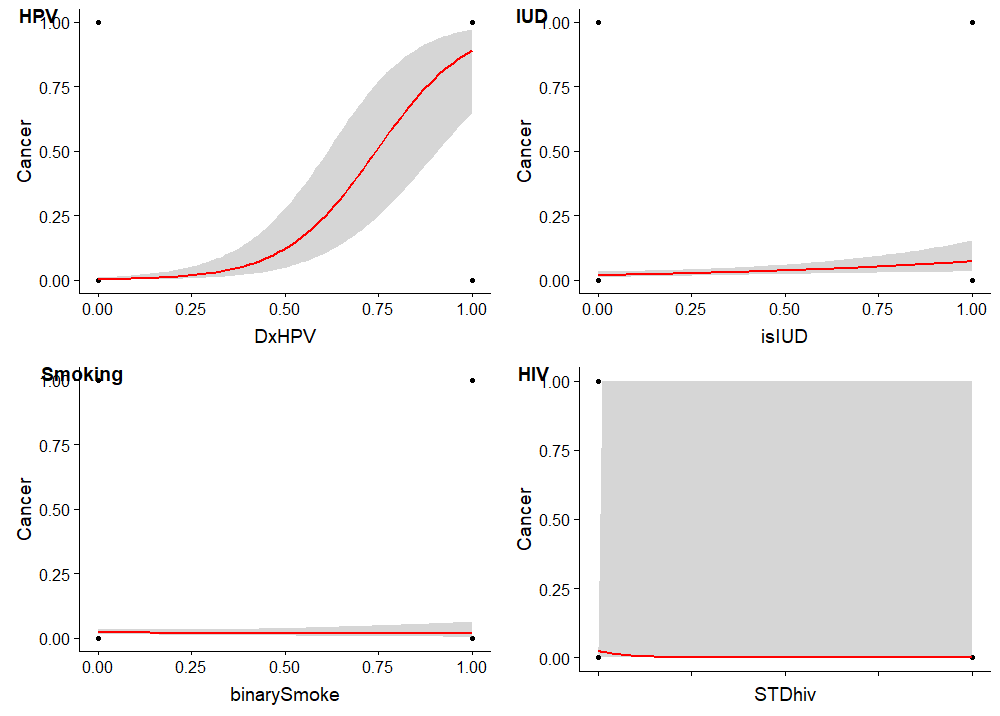
Results

From generated correlation matrix (Fig. 1), it was obvious that HPV was the major positive correlator of cervical cancer, with a r of 0.9. Weaker positive correlation was also observed in both time since first diagnosis of STD as well as time since last diagnosis of STD. (r = 0.2 for both). Other common risk factors were not observed to be significantly correlated to cervical cancer with this analysis.



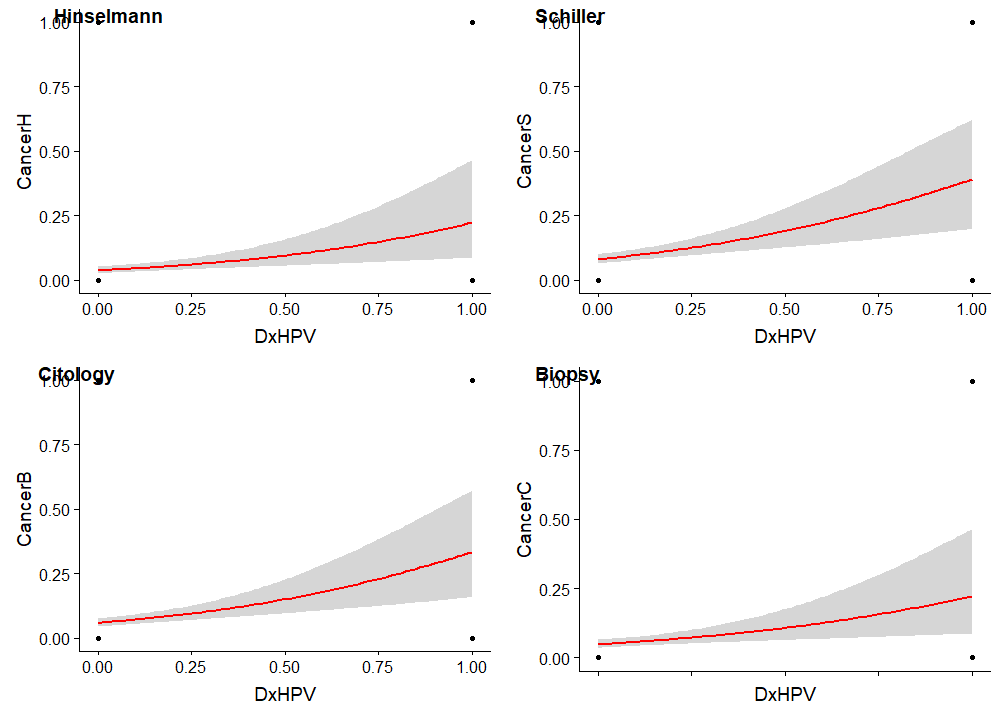
**Figure 1.** *Correlation matrix on all attributes. Size of the dot suggests the potency of correlation, as color of dots suggests direction of correlation. (Red dot for negative correlation and blue dot for positive correlation.)*

Following analysis of logistic regression of four risk factors suggested by previous research as well as the correlational analysis. HPV diagnosis, IUD usage, Smoking and HIV diagnosis were selected for logistic regression against cervical cancer diagnosis (Fig 2). Apparently, HPV diagnosis was the most significant contributor to cervical cancer, as P-Value of 3.51e-15 suggests a potent effect. Analysis regarding to IUD usage also returns to be significant, as a P-value of 0.00531 suggest significant contribution toward cervical cancer. One fact to note is that the estimated coefficient was very small (~1.5 logits), the effect is weak but significant. Smoking and HIV diagnosis both returned to be non-significant regarding to cervical cancer diagnosis.

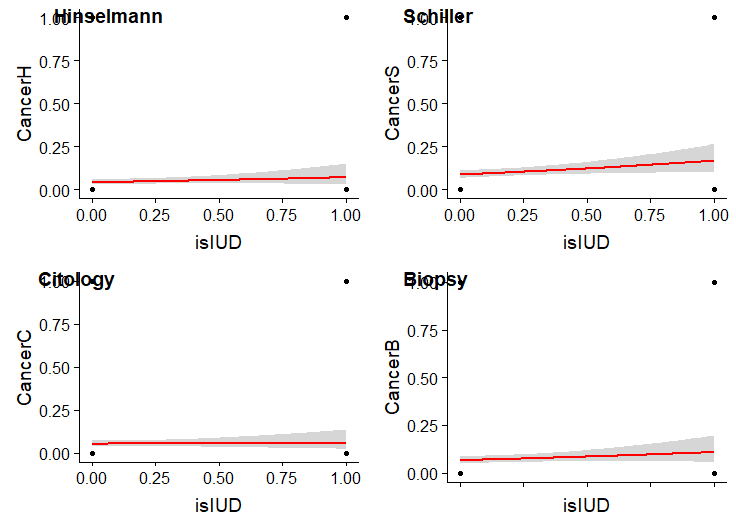


**Figure 2.** *Logistic models of HPV diagnosis, IUD usage, Smoking and HIV diagnosis against cancer. HPV diagnosis and IUD usage against cancer were both statistically significant (P = 3.51e-15 (\*\*\*\*) and P = 0.00531(\*\*)) while both other results were non-significant (P > 0.05).*

Final analysis was done on diagnostic method of cervical cancer (Hinselmann, Schiller, Citology and Biopsy) in relation to HPV diagnosis and IUD usage. Result of HPV against diagnostic methods shows that there were no significant differences in diagnostic method with patients with HPV, as all 4 analysis show similar P-values, with Cytology method show a lower P-value. Result of IUD usage against diagnostic methods shows that there much discrepancies among diagnostic methods of cervical cancers against IUD usage, as only Schiller’s test displayed significant relationship in this case (P-value = 0.022(\*)).



**Figure 3.** *Different diagnostic methods of cervical cancer against HPV diagnosis. All P-values are very significant (\*\*\*\*) except Citology with P = 0.00312 (\*\*).*



**Figure 4**. *Different diagnostic methods of cervical cancer against IUD usage. Only Schiller’s test was significantly (P = 0.022(\*)) related to patients with IUD. All other diagnostic methods were not significant (P>0.05).*

Discussion

Among the course of the analysis, the correlational matrix has suggested some plausible correlations judging by the potency of the effect, which only leads to HPV infection with a higher correlation coefficient. This expected as most of the observations are binary, even Spearman’s correlation would not be very suggestive. The following analysis of four risk factors reveals that the HPV infection was still the most potent contributor of cervical cancer; as Burd [1] concluded in his review, HPV disturb the cell cycle of infected cell and therefore induce cancer. The usage of IUD also was significantly linked to cervical cancer; unlike Castellsagué X et al [2] suggested, here the IUD seems to increase the risk of cervical cancer rather than decreasing it. In Venezuela, most woman uses copper IUD under support of the government [4]; as copper is a potential carcinogen, there could potentially be relationship between high rate of copper IUD usage and cervical cancer. Upon different diagnostic methods of cervical cancer, all were still tightly connected to HPV infection; the IUD relations, however, is different. IUD usage was significantly related to cancer diagnosed by Schiller’s test but no other test. This is interesting as Schiller’s test was the least invasive test as it was usually the earlier step of cervical cancer screening; such test could be given during the routine IUD sessions [3].

Conclusion

This analysis on patients in Venezuela has confirmed the strong relationship between HPV infection and cervical cancer again. In addition to this obvious result, what was interesting was that the IUD usage in Venezuela has positive relationship with cervical cancer, which contradict previous studies on the topic. The analysis on different methods of diagnosis of cervical cancer may have suggested that clinical visits in relation of IUD could associated to diagnosis of cancer.

Reference:

1. Burd, Eileen M. *Human Papillomavirus and Cervical Cancer*. Clin Microbiol Rev. 2003 Jan; 16(1): 1–17.
2. Castellsagué X et al. *Intrauterine device use, cervical infection with human papillomavirus, and risk of cervical cancer: a pooled analysis of 26 epidemiological studies*. Lancet Oncol. 2011 Oct;12(11):1023-31.
3. Heatley, M. K. *A critical evaluation of the use of the Schiller test in selecting blocks from the uterine cervix in suspected intraepithelial neoplasia.* Ulster Med J. 1995 Oct; 64(2): 147–150.
4. Rodriguez-Arnas O et al. *Use of IUDs in Venezuela, South America*. Contraceptive Delivery Systems. 1980 Jul; 1(3):272.
5. *NIH fact sheet of Cervical Cancer*, NIH: <https://report.nih.gov/nihfactsheets/viewfactsheet.aspx?csid=76>