



How to Cite:

Mondal, P. K., Noor, R. R., Tiang, J.-J., Biswas, U., & Nahid, A.-A. (2025). Influential socio-demographic traits in caesarean delivery diagnosis through statistical analysis. *International Journal of Health Sciences*, 9(S1), 567–588. <https://doi.org/10.53730/ijhs.v9nS1.15746>

Influential socio-demographic traits in caesarean delivery diagnosis through statistical analysis

Proloy Kumar Mondal

Department of Digital Anti-Aging Healthcare, Inje University, Gimhae 50834, South Korea

Email: proloykumar1996@gmail.com

Raad Ragib Noor

Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

Email: rnoorraad67@gmail.com

Jun-Jiat Tiang

Centre for Wireless Technology (CWT), Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia

Corresponding author email: jjtiang@mmu.edu.my

Uzzal Biswas

Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

Corresponding author email: uzzal.biswas@ece.ku.ac.bd

Abdullah-Al Nahid

Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

Corresponding author email: nahid.ece.ku@gmail.com

Abstract--Cesarean Section (CS) delivery is another mode of parturition other than the vaginal delivery. In some cases (pregnancy complications and so on), it may be necessary to use CS delivery rather than natural birth delivery. Yet, not all CS deliveries are

required. Unwarranted CS delivery may lead both the mother and the baby to some serious health issues. The worldwide rise in CS rates has led to concerns over possible overuse. The chi-square test, crosstab analysis, gamma coefficient analysis, and Cramer's V were employed to identify the most significant socio-demographic effect on the CS delivery rate. This research suggests that other than residence, four factors household income, previous CS delivery, number of antenatal care visits, and watching TV during pregnancy have a stronger associate with CS rate, but the remaining five, smoking during pregnancy, listening to radio during pregnancy, mother's education, and terminated pregnancy had no any associate to CS rate. Additionally, these results will enable the authority to understand the causes of the rising trend of CS delivery and to take measures to limit CS delivery. It will also shed light on some of the common myths concerning why CS is performed.

Keywords---Cesarean section, Statistical analysis, Crosstab analysis, Chi-square test, Gamma Coefficient, Cramér's V.

1 Introduction

Caesarean section (CS) is a surgical procedure by which one or more babies are delivered through an incision in the mother's abdomen (Fiszler et al., 2025). Over the past two decades, the global rate of CS childbirth has seen a significant increase (Niino, 2011). The rates vary significantly among different regions and countries. Fig. 1 illustrates the regions and countries corresponding to CS rates. Figure 1 shows that the C-section rates vary significantly across different regions and countries. There is a wide variation in the use of C-sections around the world, which is reflected in the comparative figures for selected countries. For example, countries such as the Dominican Republic (62.9%), Turkey (58.4%), Brazil (56.4%), South Korea (53.8%), and Mexico (52.6%) have C-section rates that exceed the World Health Organization (WHO) recommended 10-15 percent. The reasons behind these high rates can be attributed to the increasing prevalence of medical interventions, financial incentives for private healthcare, and the role of maternal personal choice. On the other hand, Pakistan has a rate of 24.7 percent, which is in the middle range, higher than the global average but much lower than the countries above. This indicates a dual trend of overuse in urban areas and underuse in rural and resource-limited areas. Among developed countries, the United Kingdom (34.0%) and Poland (42.2%) also have relatively high C-section rates, which may be due to increasing age at delivery, risk of litigation, and modern obstetric practices. On the other hand, countries such as the Netherlands (15.2%) and Iceland (14.3%) appear to be following WHO guidelines. In Chad (1.1%), the rate is so low that it indicates limited availability of surgical delivery facilities. This global variation suggests that C-section use requires policy strategies that are tailored to the relevant needs and ensure that the procedure is used only when it is medically necessary, to reduce maternal and newborn health risks, and to maintain equity in access to care.

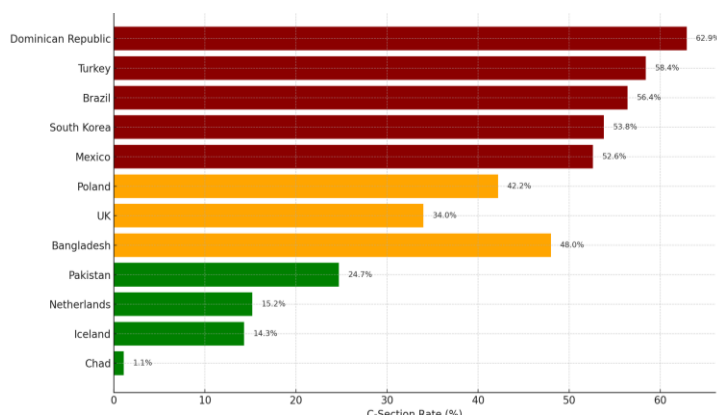


Figure 1. *The global rate of CS*

Although CS was originally intended as a medical procedure that can save mothers' and babies' lives when there are complications in pregnancy but it comes with higher risks when done without medical indication (Villar et al., 2007). A research by Jose Villar et al. showed that both maternal and neonatal mortality and morbidity rates are higher for CS than natural delivery (Villar et al., 2007). Babies born by CS delivery are more prone to developing immune disorders and have a high risk of facing many long-term negative effects, such as allergies, diabetes, obesity, etc. (Munasir & Khoe, 2020). Mothers also have an increased risk of uterine rupture, abnormal placentation, and ectopic pregnancy (Sandall et al., 2018). These problems can cause complications in future pregnancies. If not careful, the mother can suffer from excessive bleeding, which can be life-threatening. Also, unnecessary CS can be considered a waste of resources. A study by Chaudhury et al. showed that during the year 2028 in Bangladesh, almost 70% of CS performed were unnecessary, and almost 48 million USD were spent on these unnecessary CS deliveries. The rise of unnecessary CS delivery is currently a major concern worldwide. A report by the World Health Organization (WHO) suggests that in 2021, 21% of all childbirths were from CS delivery, which is set to increase by 29% by 2030. WHO recommends that the rate of CS should be 10% to 15% (*WHO Recommends the Rate CS Should Be 10 to 15%* - Google Search, n.d.). According to the Pakistan Demographics and Health Survey (PDHS) in 2017-2018, the rate of CS delivery was 22%, which exceeds the WHO recommended range (*According to Pakistan Demographics and Health Survey in 2017-2018 the Rate of CS Delivery Was 22% Which Exceeds the WHO Recommended Range* - Google Search, n.d.). If the reasons for increasing the CS rate can be determined, necessary steps can be taken to reduce the rate of CS. CS related data needs to be analyzed to understand which social or economic factor plays the key role in increasing the CS rate. Statistical analysis is a popular approach to analyze the CS data and to get more insights into the CS delivery rate (Sengupta et al., 2021). Research from MD. Akhtarul Islam et al., based on the data of Indonesia, showed that factors such as living area, mother's age, educational status, and family wealth have a positive association with C-section (Md. A. Islam et al., 2022). Another research from the same author based on the data of Sub-Saharan African Countries showed that mothers' educational status has the most significant impact on CS rate among other influencing factors such

as area of living, family wealth, and father's education (Khawaja & Al-Nsour, 2007a). Khawaja et al.'s research revealed that several crucial factors influenced the likelihood of CS, including multiple births, child-birth weight, advanced maternal age, and antenatal visits. However, the place of residence did not demonstrate a consistent association, and the mother's education showed no significant correlation (Khawaja & Al-Nsour, 2007b). Begum et al. stated that the probability of CS increases with improved socio-economic status, higher education, higher age, and with more antenatal care usage (Begum et al., 2017). Another study by T. Muhammad et al. showed that mass media exposure alongside mothers' educational status and birth order has an association with CS (Ahmmed et al., 2021). In a research based on data of Iranian pregnant women, Soraya N M et al. stated that fear of labor pain, husband's education level, and income were directly associated with increased rate of CS (Motlagh et al., 2020a).

Many studies have been conducted on determining the influential factors that play a role in the rising rate of CS. One study used bivariate analysis to determine the relationship between CS and many socio-demographic features such as household wealth, number of ANC visits, etc. (Md. A. Islam et al., 2022). However, since bivariate analysis focuses solely on the relationship between two variables, it may often miss patterns that are present when additional variables are also taken into account (*Bivariate Analysis - GeeksforGeeks*, n.d.). Some studies have focused on a limited number of variables, which may not provide a comprehensive understanding of the issue (Md. A. Islam et al., 2022) (Begum et al., 2017). Also, all of the studies on CS are region-dependent. Based on regions, the factors may significantly vary.

This study is focused on determining the set of important clinical features for diagnosing the occurrence of a CS in the last birth. The focus of this study is to determine the socio-demographic variables so that it can raise awareness and help in reducing unnecessary CS delivery. To achieve this, the inspiration has been drawn from conventional statistical techniques. This analysis is divided into four main sections. In the "Methods" section, we have outlined our proposed approaches and provided a brief overview of the utilized dataset. The "Results and Discussion" section presents the outcomes of our research, including a condensed summary of the analysis, results, and findings. Finally, in the "Conclusion" section, our findings have been summarized and concluded.

2 Materials and Methods

2.1 Dataset Description

The dataset was obtained from a secondary source known as the Demographic Health Survey (DHS). A permission was received [from https://dhsprogram.com/](https://dhsprogram.com/) to access and use the data under strict guidelines. From the DHS database, we used the PDHS secondary datasets that are accessible online ("*The DHS Program - Quality Information to Plan, Monitor and Improve Population, Health, and Nutrition Programs.*" Accessed: Apr. 23, 2023. [Online]. Available: <https://www.Dhsprogram.Com/> - Google Search, n.d.). There are 3038 married women with at least one delivery in the last five years, ranging in age from 15 to

49 years (M. S. Islam et al., 2022). (Figure 2) represents Pakistan's region-based CS rates for the PDHS from 2017 to 2018.

There are 875 features in the PDHS'17-18 datasets; some of the features were related to household data, contraceptive knowledge and practice, postpartum, post-delivery, children's health care, nutritional data, and migratory trends. These features weren't essential for making predictions. Based on a previous study (M. S. Islam et al., 2022). We have used 24 features associated with CS delivery. In the dataset, the target variable known as "Last birth a CS" is represented as a binary value. A non-CS delivery is indicated by the value "0", whereas a CS delivery is represented by the value "1". The dataset utilized for this study is shown in Table 1.

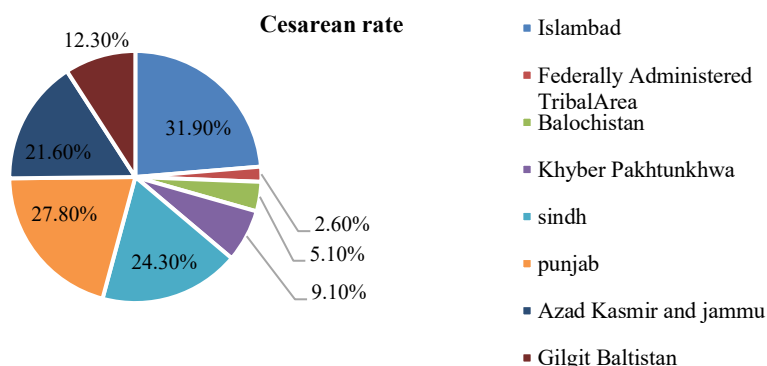


Figure 2. Region based CS rates in Pakistan.

2.2 Feature Importance Method

The feature importance technique is a kind of strategy that may be used while carrying out the process of feature selection (Talin et al., 2022). The basis for feature selection may be provided by the significance of the features. The features that are the least significant may be left out of subsequent analysis, which provides a foundation for finding ways to reduce the overall dimensions. It is a significant part of our investigation. In order to determine the correlation and connection of characteristics with the target variable, we have made use of a few well-known statistical approaches such as the Chi-Square test, Gamma Coefficient analysis, and Cramer's V analysis.

2.3 Statistical Analysis

Statistical analysis is a powerful tool that helps us to understand data by revealing patterns, trends, relationships, and significant differences. It's amazing to see how statistical analysis is utilized in such a diverse range of fields, including business, economics, finance, social sciences, and health sciences (*What You Need to Know About Statistical Analysis,* *Business News Daily*. Accessed: Jul. 22, 2023. [Online]. Available: <https://www.businessnewsdaily.com/6000-statistical-analysis.html> - Google

Search, n.d.). It can help with decision-making, predicting outcomes, and testing hypotheses. Using statistical analysis, we can draw a possible conclusion from the data we have. Different types of statistical tests lead to different inferences and conclusions. By combining the results of various tests and their observed patterns, we can conclude for our dataset, which is a very important part of the research. Using statistical analysis techniques, we can analyze the relationship between two features. For our dataset statistical analysis allows us to determine whether there is a relationship between CS and other features and if so, what kind of relationship it is. There are many kinds of statistical tests such as parametric tests and non-parametric tests (Gerald & Patson, 2021). The primary considerations that determine whether a parametric or non-parametric test should be used are whether the data are in a normal or skewed distribution. In general, parametric tests depend on the dataset has a normal distribution (Jett & Speer, 2016). In this study, we have chosen a random feature name, Mother's BMI to test the distribution of the data. From Figure 3, we can see that this feature represents a skewed distribution. Thus, we have chosen non-parametric tests for our statistical analysis.

In this study, we used several statistical analysis methods, namely the chi-square test, gamma coefficient analysis, Cramer's V, and crosstab analysis. These are described below.

Table 1
Description of features

Feature Name	Feature Specification
Mother Age	Age of Mother
Residence	Type of Place of Residence
MBMI	Mother's BMI
MEL	Mother's educational level
MO	Mother's Occupations
ANC	Number of ANC Visits
EHTP	Ever had a terminated pregnancy
DMPDP	Decision-Making Power on Delivery Place
WTV	Watching-TV
HE	Husband's Education
HO	Husband's occupation
HWQ	Household Wealth Quantile
SDW	Sources of Drinking Water
HTF	Household Toilet Facility
NOHM	Number of Household Members
TNEBC	Total_children_ever_born
BILFY	Births_in_last_five_years
NOLC	Number_of_living_children
RNOM	Reading_Newspaper_or_Magazine
LR	Listening_to_Radio
Smoker	Smoker
SDV	Suffered_by_Domestic_Violence
HPC	Had Previous C-section
SOCB	Size_of_child_at_birth

2.3.1 Chi-Square Test

The Chi-square test is a statistical test that is used to determine if there is a significant association between two categorical variables (*Chi-Squared Test - Wikipedia.* Accessed: Oct. 30, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Chi-squared_test - Google Search, n.d.). Chi-Square testing is nonparametric. It determines whether predicted and actual values are statistically significant. Researchers repeatedly utilize bivariate tables to test independence between two variables (M. A. Islam et al., 2022). It is a popular way to assess whether two variables are connected or not. Chi square test follows a general formula. We can write the formula as:

$$X^2 = \sum \frac{(u - v)^2}{v} \quad (1)$$

where X^2 denotes the chi-square value, u represents the observed value, and v indicates expected value.

The parameter we use in the Chi-squared test is often referred to as the null hypothesis (Volchok, 2020). According to the null hypothesis for the Chi-squared test, two categorical traits are independent of one another and have no relationship or dependence on one another (Franke et al., 2012). The lowest number to accept the null hypothesis is 0.05 (Rana & Singhal, 2015). The alternative hypothesis is accepted, and the null hypothesis is rejected if its value is less than 0.05. The value being less than 0.05 indicates that there is a relationship or association between the two variables.

In this study, we have performed the chi-square test to see if there is any dependency between each of the independent features and the only dependent feature, named CS delivery. This result will help us to check the strength of the association later.

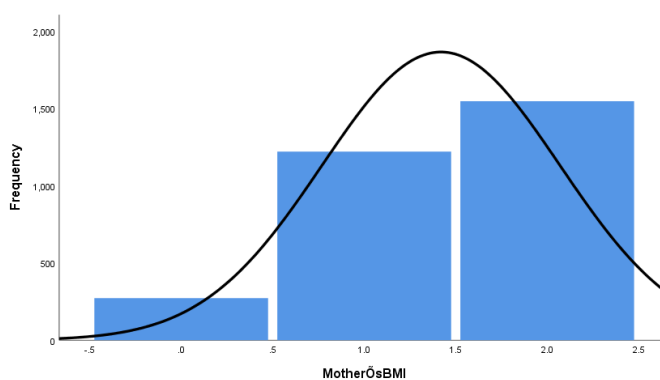


Figure 3. Skewed distribution of feature named Mother's BMI

2.3.2 Gamma Coefficient

The gamma coefficient, which has values between -1 and 1, symmetrically measures the degree of association between two ordinal variables (Kearney et al., 2019). Its purpose is to evaluate the two aspects' monotonic connection (Raveh, 1986). The gamma value, which may be positive or negative, determines the strength of the correlation. When two ordinal qualities are positively correlated, there is a proportionate relationship between them; when they are negatively correlated, there is an inverse relationship. To be more precise, a gamma value between 0.01 and 0.29 denotes a weak correlation, 0.3 and 0.49 a moderate relationship, 0.50 and 0.69 a strong association, and a gamma value over 0.70 the greatest conceivable association (Cohen et al., 2013). A value of 1.00 indicates a perfect association (Cohen et al., 2013).

In this study, to determine the association strength of those features on CS delivery, we have performed the gamma coefficient test.

2.3.3 Cramér's V

Named after the Swedish statistician Harald Cramér, Cramer's V is a measure of association or correlation between two nominal variables (Bergsma, 2013). It is denoted as V and the value of it ranges from 0 to 1, where 0 corresponds to no association between the variables and 1 corresponds to perfect association between the variables. The value of Cramer's V can be determined by equation (3.2)

$$V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}} \quad (2)$$

Where χ^2 is the chi-square statistic for the contingency table, n is the total number of observations, and k and r are the number of columns and rows.

To determine Cramer's V between our desired features, in our contingency table, we used "CS delivery" as the row and all other nominal categorical features as columns. Every nominal feature was used to create an individual contingency table. The strength of association was considered strong only if Cramer's V was higher than 0.70. And if Cramer's V was less than 0.05, the feature and target feature were considered to have no association between them. Cramer's V, ranging from 0.05 to 0.30, is considered a weak association between the feature and the target feature.

2.3.4 Crosstab Analysis

One useful technique for examining the relationship between two categorical features is crosstab analysis. A contingency table that represents the frequency of outcomes based on the feature is created for the crosstab analysis (Wong et al., 2008). Through the use of crosstab analysis, researchers are able to get deeper insights into their datasets. Two categorical variables are compared in a crosstab; one is shown in the table's rows, and the other in its columns (Yates, 1948). The table's cells show how often certain events occur.

In our study, we used just the dependent variable CS delivery and conducted a crosstab analysis on each independent categorical feature. The relationship between CS delivery and other independent categorical factors is shown by this analysis. We have derived a frequency distribution table from this analysis, which shows the distribution of the CS delivery occurrence as a result of the percentage and numerical value for each feature.

3 Results and Discussions

We have used distinct methods in order to determine the features that are likely to be the most influential in causing CS delivery. Our dataset contains a significant amount of data that has nominal and ordinal data types. Among the twenty-four features in our dataset, we find that twelve features are of nominal types, and another twelve features are ordinal types. In regard to the fact that our dataset does not follow a normal distribution, we have used the Chi-square test and the crosstab analysis for each feature. On the other hand, for nominal variables, we have used Cramer's V Coefficient, and for ordinal variables type we have analyzed the gamma coefficient. The outcome of these analyses is outlined in the following subsections.

3.1 Crosstab Analysis Outcome

From the earlier analysis, we've assessed the degree of association between CS delivery and each categorical variable. This section focuses on the categorical relationship with CS delivery, employing crosstab analysis. While a crosstab analysis has been conducted for each categorical feature, this section highlights the relationships between CS delivery and four specific features, chosen due to their moderate or strong associations with CS delivery. The graphical representation of their relationships is depicted in Figure 5. For the remaining features, the crosstab analysis results are provided in Table 2.

Table 2
Outcomes of crosstab analysis

Feature Name	Categories	Labeled As	CS Delivery Status	
			No	Yes
Mother Age	15-24	0	583 85.4%	100 14.6%
	25-35	1	1393 79.7%	354 20.3%
	36-49	2	527 86.7%	81 13.3%
	Total		2503	535
Residence	Urban	1	1074 75.9%	341 24.1%
	Rural	2	1429 88.0%	194 12%
	Total		2503 82.4%	535 17.6%
MBMI	Underweight (<18.50)	0	244 89.7%	28 10.3%
	Normal (18.50-24.9)	1	1058 86.7%	162 13.3%

Feature Name	Categories	Labeled As	CS Delivery Status	
			No	Yes
	Overweight (>=25)	2	1201 77.7%	345 22.3%
	Total		2503 82.4%	535 17.6%
MEL	No education	0	1269 82.9%	262 17.1%
	Primary	1	338 83.0%	69 17.0%
	Secondary	2	515 80.8%	122 19.2%
	Higher	3	381 82.3%	82 17.7%
	Total		2503 82.4%	535 17.6%
MO	Not Working	0	2188 81.8%	487 18.2%
	Working	1	313 86.7%	48 13.3%
	Total		2501 82.4%	535 17.6%
ANC	No Visits	0	338 98.8%	4 1.2%
	No of visits (1-3)	1	571 88.5%	74 11.5%
	No of visits (>=4)	2	646 70.8%	266 29.2%
	Total		1555 81.9%	344 18.1%
EHTP	No	0	1727 82.2%	374 17.8%
	Yes	1	776 82.8%	161 17.2%
	Total		2503 82.4%	535 17.6%
DMPDP	Self	0	162 77.5%	47 22.5%
	Both	1	757 76.0%	239 24.0%
	Husband alone	2	1207 87.3%	176 12.7%
	Someone else	3	350 86.6%	54 13.4%
	Total		2476 82.8%	516 17.2%
WTV	No	0	1288 92.0%	112 8.0%
	Yes	1	1215 74.2%	423 25.8%
	Total		2503 82.4%	535 17.6%
HE	No Education	0	830 90.2%	90 9.8%
	Primary	1	423 91.0%	42 9%
	Secondary	2	753 78.6%	205 21.4%
	Higher	3	469 72.4%	179 27.6%
	Total		2475 82.4%	516 17.6%

Feature Name	Categories	Labeled As	CS Delivery Status	
			No	Yes
			82.7%	17.3%
HE	No Work	0	86	9
			90.5%	9.5%
	Agriculture	1	437	56
			88.6%	11.4%
	Professional	2	461	168
			73.3%	26.7%
	Business	3	396	112
		78.0%	22.0%	
HWQ	Others	4	1092	170
			86.5%	13.5%
	Total		2472	515
			82.2%	17.2%
	Poorest	0	652	33
			95.2%	4.8%
	Poor	1	640	56
		92.0%	8%	
SDW	Middle	2	455	89
			83.6%	16.4%
	Rich	3	421	149
			73.9%	26.1%
	Richest	4	335	208
			61.7%	38.3%
	Total		2503	535
		82.4%	17.6%	
HTF	No Improved Water	0	330	38
			89.7%	10.3%
	Improved Water	1	2173	497
		81.4%	18.6%	
NOHM	Total		2503	535
			82.4%	17.6%
	Unhygienic Toilet	0	705	66
		91.4%	8.6%	
TNEBC	Hygienic Toilet	1	1798	469
			79.3%	20.7%
	Total		2503	535
		82.4%	17.6%	
BILFY	1-5	1	416	139
			75%	25%
	6-10	2	1209	231
			84%	16%
TNEBC	>11	3	878	165
			84.2%	15.8%
	Total		2503	535
			82.4%	17.6%
BILFY	1-2	0	831	259
			76.2%	23.8%
	3-4	1	831	204
			80.3%	19.7%
BILFY	>=5	2	841	72
			92.1%	7.9%
	Total		2503	535
			82.4%	17.6%
BILFY	1-2	0	2175	463
			82.4%	17.6%
	3-4	1	326	72
			81.9%	18.1%
BILFY	>=5	2	2	0
			100%	0.0%

Feature Name	Categories	Labeled As	CS Delivery Status	
			No	Yes
	Total		2503 82.4%	535 17.6%
NOLC	1-2	0	900 75.9%	286 24.1%
	3-4	1	880 82.4%	188 17.6%
	>=5	2	723 92.2%	61 7.8%
	Total		2503 82.4%	535 17.6%
RNOM	No	0	2227 85.0%	393 15%
	Yes	1	273 65.8%	142 34.2%
	Total		2500 82.4%	535 17.6%
LR	No	0	2269 82.5%	482 17.5%
	Yes	1	234 81.5%	53 18.5%
	Total		2503 82.4%	535 17.6%
Smoker	No	0	2410 82.2%	521 17.8%
	Yes	1	91 86.7%	14 13.3%
	Total		2501 82.4%	535 17.6%
SDV	No	0	1251 74.6%	426 25.4%
	Yes	1	1182 92.1%	102 7.9%
	Total		2433 82.2%	528 17.8%
HPC	No	0	2487 98.1%	48 1.9%
	Yes	1	10 2%	487 98%
	Total		2497 82.4%	535 17.6%
SOCB	1-2	1	172 79.3%	45 20.7%
	3-4	2	2251 82.5%	478 17.5%
	>=5	3	73 85.9%	12 14.1%
	Total		2496 82.3%	535 17.7%

The variable HPC exhibits a robust association with the occurrence of the last birth from CS, as depicted in Fig. 4. Fig. 4 illustrates that mothers who have previously undergone a CS are likely to have subsequent deliveries via CS, whereas the incidence of CS is considerably lower among mothers who have never had a CS.

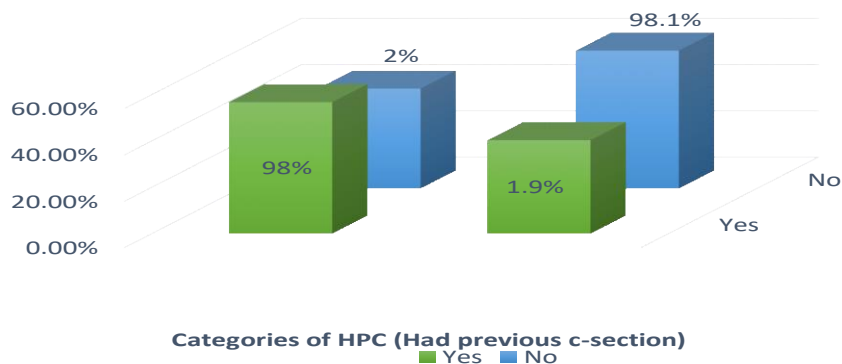


Figure 4. Occurrence of CS delivery based on HPC

HTF maintains a moderate relationship with CS delivery. Fig. 5 represents the categorical relationship between HTF and CS delivery. From Fig. 5, we can see that there are two categories for this feature. It is clear from this figure that women who had hygienic toilet facilities are more likely to have CS for childbirth. So, we can say that HTF is more related to CS delivery

Fig. 6 illustrates the correlation between the mother's BMI and the occurrence of CS during their most recent childbirth. There exists a moderate relationship between these two traits. The mother's BMI is classified into three distinct groups. The initial group pertains to instances where the mother's BMI number is below 18.50. The subsequent category encompasses values ranging from 18.50 to 24.9. Lastly, the last category encompasses values over 25. In the initial two groups, there is a lower prevalence of CS delivery among women. However, in the third category, approximately 22.30% of women have undergone a CS delivery. Based on the findings of this analysis, it can be inferred that mothers with higher BMI are more likely to have a CS delivery.

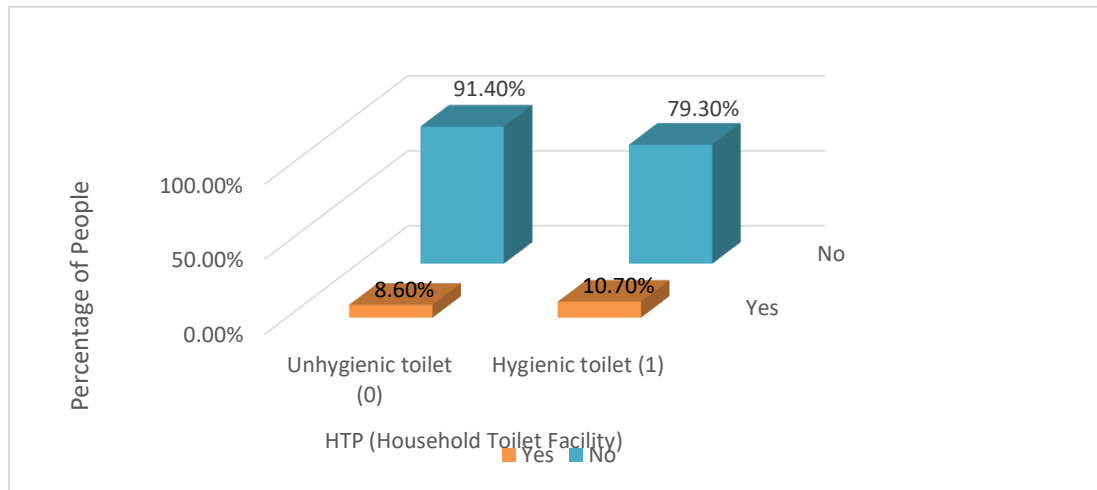


Figure 5. Occurrence of CS delivery based on HTF

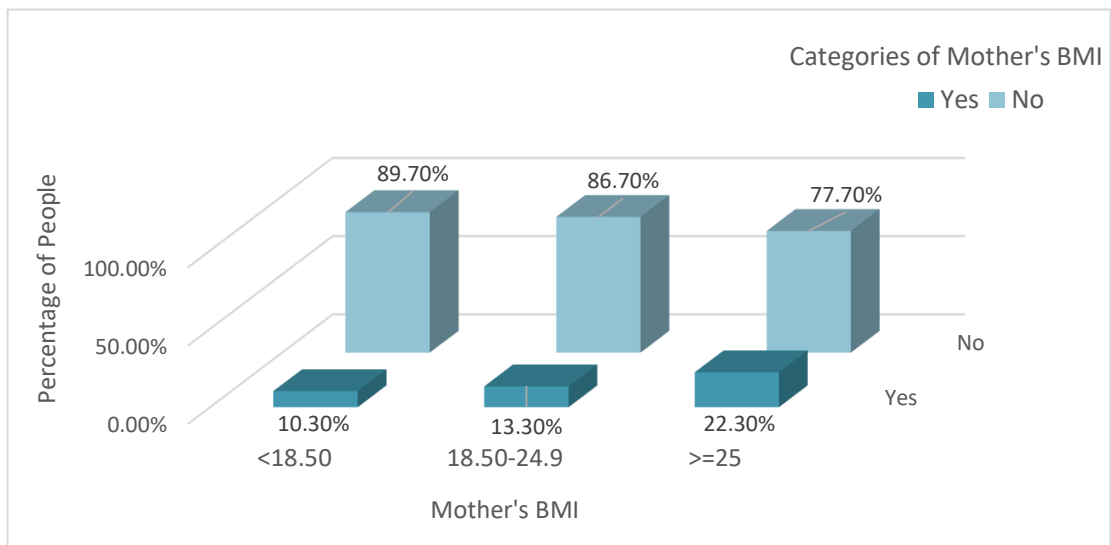


Figure 6. Occurrence of CS delivery based on Mother's BMI.

HWQ (Household Wealth Quintile) also demonstrates a strong association with the occurrence of the CS delivery, as illustrated in Fig. 7, showcasing the categorical relationship between HWQ and CS delivery. Notably, there are five categories for this feature, and it's evident from the figure that not all categories share an equal relationship with the prediction of CS delivery. The significance of the "rich" and the combination of the "richest" categories is particularly pronounced in CS delivery prediction. In the "rich" category, 26.10% of individuals have a relationship with CS delivery, while in the "richest" category, the percentage rises to 38.30%. Hence, it can be concluded that the categories "rich" and "richest" exhibit a stronger association with CS delivery compared to other categories within the HWQ feature.

Fig. 8 illustrates how some specific outcomes of a feature affect the CS rate. From Fig. 9 we can see that 98% mother who went through CS previously goes through it again when expecting a child. 10.70% of the pregnant women who have hygienic household toilet facilities and 22.30% of the mothers who has a BMI of over 25 resort to CS delivery. Also, from Fig. 9 we can conclude that 38.30% of mothers who are considered to be the richest based on their wealth under-goes CS. This analysis reveals that HPC and HWQ have a heavy influence on CS delivery, while the influence of HTP and MBMI is comparatively weaker.

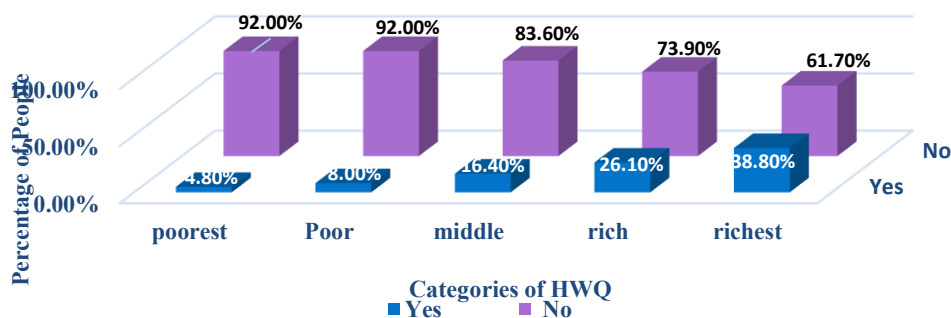


Figure 7. Occurrence of CS delivery based on HWQ

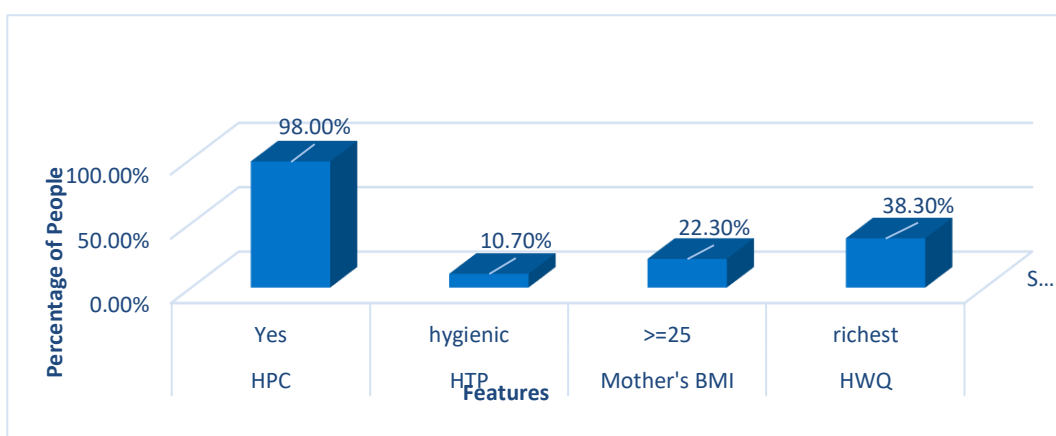


Figure 8. Relationship of important features with CS delivery

3.2 Chi-Square Test Outcome

As stated earlier, we employed the Chi-square test to identify the features associated with CS delivery. The outcomes of the Chi-square test are presented in Table 3.

Table 3
Outcomes of Chi-square test

Features Name	Chi-square Test (p value)	Association
Mother Age	.000	Yes
Residence	.000	Yes
MBMI	.000	Yes
MEL	0.699	No
MO	0.022	Yes
ANC	0.000	Yes
EHTP	0.679	No
DMPDP	0.000	Yes
WTV	0.000	Yes
HE	0.000	Yes
HO	0.000	Yes
HWQ	0.000	Yes
SDW	0.000	Yes
HTF	0.000	Yes
NOHM	0.000	Yes
TNEBC	0.000	Yes
BILFY	0.789	No
NOLC	0.000	Yes
RNOM	0.000	Yes
LR	0.689	No
Smoker	0.240	No
SDV	0.000	Yes
HPC	0.000	Yes
SOCB	0.674	No

In the study of the Chi-square test for independence, the null hypothesis represents no association or relationship between two variables. On the other hand, the alternative hypothesis proposes an actual dependency or association between the studied variables. The acceptance of the alternative hypothesis and the rejection of the null hypothesis occur when the calculated p-value from the chi-square test is below the significance level of 0.05. According to Table 3, the alternative hypothesis is endorsed for a total of 18 variables, as the associated p-values for these variables are found to be less than 0.05. The study's findings imply a significant correlation between CS delivery and 18 out of the 24 examined factors. Table 3 represents the names of these 18 associated features. According to Fig. 9 and Table 3, the p-value is estimated to 0 for the following 17 features: Mother Age, Residence, MBMI, MO, ANC, DMPDM, WTV, HE, HO, HWQ, SDW, HTF, NOHM, TNEBC, NOLC, RNOM, SDV, HPC. From this analysis, we can assume that these features are closely related to CS delivery. Further, their strength of association with CS delivery will help us to determine the most important feature for the prediction of CS delivery.

Fig. 9 is a visual representation of features that are associated with CS delivery and their p-value. All the features shown in Fig. 9 have a p-value less than 0.05, which means the null hypothesis is rejected and there is a relationship between the features and CS delivery.

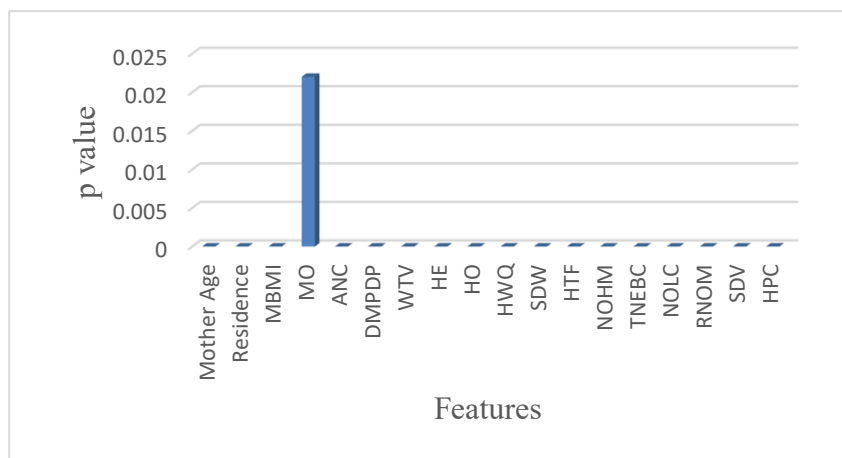


Fig. 9: Features related to CS delivery

3.3 Gamma Coefficient Analysis Outcome

In this study, we have used the gamma coefficient to observe the strength of association between each of the features that consist of an ordinal variable and CS delivery. Table 4 shows the outcomes of the gamma coefficient analysis.

From Table 4, we can see that 12 features have an association with CS delivery prediction, but not all features have the same strength of association. Moderate, strong, and strongest associations are significant for finding the important features. Analyzing the value of the gamma coefficient, we found that MBIM, HE, and HTF features have a moderate association with the target feature, and ANC and HWQ features have a strong association with the target feature.

Table 4
Strength of association based on value of gamma coefficient.

Features Name	Value of gamma coefficient	Strength of Association
Mother Age	-0.016	Negligible
MBMI	0.309	Moderate
MEL	0.029	Negligible
ANC	0.670	Strong
HE	0.375	Moderate
HO	-0.070	Negligible
HWQ	0.556	Strong
HTF	0.472	Moderate
NOHM	-0.150	Negligible
TNEBC	-0.353	Negligible
BILFY	0.015	Negligible
NOLC	-0.363	Negligible

3.4 Cramér's V Analysis Outcome

In this study, we used the Cramer's V test to assess the strength of association between features with nominal variables and CS delivery. Table 5 shows the outcomes of the Cramer's V analysis. From Table 5 it can be noticed that "HPC"

feature has almost perfect association to “CS delivery” target variable based on Cramer’s V, which means that the CS delivery status can be determined very accurately by knowing whether the mother had previous CS or not. Also, other features such as type of place of residence, watching TV, sources of drinking water, reading newspaper or magazine, suffered by domestic violence and decision-making power of mother on delivery place have a weak association on the target feature. Other nominal features have no association to the target feature.

Table 5
Strength of association based on value of cramer’s v analysis

Feature name	Cramer’s V	Strength of Association
Residence	0.159	Weak
MO	0.042	No
EHTP	0.007	No Association
DMPDP	0.141	Weak
WTV	0.233	Weak
SDW	0.071	Weak
RNOM	0.173	Weak
LR	0.007	No Association
Smoker	0.021	No Association
SDV	0.226	Weak
HPC	0.933	Strong
SOCB	0.013	No Association

3.5 Discussion

Our study aimed to determine the socio-demographic variables that influences the CS rate. We used statistical analysis methods such as crosstab analysis, Chi-square test, gamma coefficient analysis and Cremer’s V analysis. The outcomes of Chi-square test showed that variables such as mother’s age, residence, mother’s BMI, occupation, number of antenatal care visits, decision making power on delivery place, watching TV, husband’s education, occupation, wealth, drinking water source, household toilet facility, number of members in family, total number of children ever born, number of living children, suffering from domestic violence, reading newspaper or magazine and having previous CS has association to CS delivery. However, from Chi-square test, we cannot determine the strength of association. To determine the strength of association we used gamma coefficient analysis on ordinal variables. From gamma coefficient analysis we discovered that number of antenatal care visits and household wealth has the strongest strength of association to CS delivery and mother’s BMI, husband’s education and household toilet facility has moderate strength of association to CS delivery. Cremer’s V analysis was used to determine the strength of association of nominal variables. The analysis indicates that having previous CS has strong association to CS delivery while residence, watching TV, source of drinking water, suffering from domestic violence and decision making power of birth place has weak association to CS delivery.

Our findings show similarities to many other studies conducted in similar settings. A study conducted by S. N. Motlegh et al. based on Iranian pregnant women back in 2020 also used Chi-square analysis (Motlagh et al., 2020b). Their

findings indicated that mother's age, education, occupation, wealth, BMI has association to CS delivery. Our study outcome matches their finding. M. N. Mia et al. had the similar findings in their research (Mia et al., 2019). They researched on a Health and Demographic Surveillance Systems (HDSS) area operated by International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) and found that higher income, higher number of ANC visits, delivery of firstborn had a positive impact to the CS delivery. Our findings also supports the results of studies conducted in 2022 by Md. Akhtarul Islam et al. on Indonesia Demographic and Health Survey (IDHS) datasets of 2007, 2012 and 2017 In which they concluded that place of residence, wealth index, education, and ANC visits influence the rate of CS delivery (Md. A. Islam et al., 2022). A. K. Manyeh et al. in their research conducted based on the secondary data from the Dodowa Health and Demographic Surveillance System (DHDSS) stated that CS delivery is associated with maternal age, level of education, occupation, parity and ANC visit (Manyeh et al., 2016).

In 2017, S. Mumtaz et al. studied on the PDHS dataset and stated that mother's education, wealth and residence had strong relation to having CS delivery (Mumtaz et al., 2017). Saman Nazir et al. also studied the PDHS dataset. They determined that mother's education, wealth, residence and receiving institutional antenatal care had an influential role to CS delivery (Nazir, 2015). All of these studies further strengthen our findings. Our study was based on PDHS' 17-18 dataset. The dataset had 24 variables which strengthens the outcome of the study because the result is more accurate for more variables. Using multiple statistical analysis methods enhances the findings by allowing cross-verification, which increases the accuracy of the results. Also, crosstab analysis helps visualizing the data to make the findings clearer. However, this study has some limitations that need to be considered. First, although we have 24 variables available, it still might not be enough. Second, since our findings were based on PDHS dataset, our findings may have dependency on region. Additionally, pregnancy complications, which play a significant role in CS delivery, are not available in our dataset. Taking these strengths and limitations into account, we believe our study can help understand the reasons of increasing CS rate and help in decision making process.

4 Conclusion

CS delivery can be lifesaving when it is needed. But unnecessary CS delivery causes health risk to both mother and baby. Besides all the health risk, it is also a waste of very important medical resources. The accurate detection of CS plays a pivotal role in ensuring maternal and neonatal health during a childbirth. Some socio-demographic factors play a predominant role in dictating the rate of CS delivery. This study aimed to discover the key socio-demographic features that influences the rate of CS using various statistical methods. Our analysis has identified that having had a previous CS, household wealth quantile, number of antenatal care visit has the strongest association to having CS delivery. This study can help the authority to create awareness in order to reduce unnecessary CS deliveries and to ensure that every childbirth is met with the highest standard of safety and well-being.

Acknowledgments

Not Applicable.

References

- According to Pakistan demographics and Health Survey in 2017-2018 the rate of CS delivery was 22% which exceeds the WHO recommended range—Google Search. (n.d.).
- Ahmed, F., Manik, M. M. R., & Hossain, M. J. (2021). Caesarian section (CS) delivery in Bangladesh: A nationally representative cross-sectional study. *PloS One*, 16(7), e0254777.
- Begum, T., Rahman, A., Nababan, H., Hoque, D. M. E., Khan, A. F., Ali, T., & Anwar, I. (2017). Indications and determinants of caesarean section delivery: Evidence from a population-based study in Matlab, Bangladesh. *PloS One*, 12(11), e0188074.
- Bergsma, W. (2013). A bias-correction for Cramér's V and Tschuprow's T. *Journal of the Korean Statistical Society*, 42(3), 323–328.
- Bivariate Analysis—GeeksforGeeks. (n.d.). Retrieved August 9, 2024, from <https://www.geeksforgeeks.org/bivariate-analysis/>
- Caesarean section rates continue to rise, amid growing inequalities in access." Accessed: Jul. 02, 2023. [Online].
- Chi-squared test—Wikipedia." Accessed: Oct. 30, 2023. [Online].
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge. <https://www.taylorfrancis.com/books/mono/10.4324/9780203774441/applied-multiple-regression-correlation-analysis-behavioral-sciences-jacob-cohen-patricia-cohen-stephen-west-leona-aiken>
- Ever-increasing Caesarean section and its economic burden in Bangladesh | PLOS ONE. (n.d.). Retrieved August 9, 2024, from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208623>
- Fiszer, E., Rabkin, V., Aptekman, B., Reider, E., Chavez, R., Lavie, A., Matot, I., & Weiniger, C. F. (2025). Impact of an interdisciplinary process to increase utilization of neuraxial anesthesia for cesarean delivery: A retrospective database analysis. *International Journal of Obstetric Anesthesia*, 61, 104267. <https://doi.org/10.1016/j.ijoa.2024.104267>
- Franke, T. M., Ho, T., & Christie, C. A. (2012). The Chi-Square Test: Often Used and More Often Misinterpreted. *American Journal of Evaluation*, 33(3), 448–458. <https://doi.org/10.1177/1098214011426594>
- Gerald, B., & Patson, T. F. (2021). Parametric and Nonparametric Tests: A Brief Review. *International Journal of Statistical Distributions and Applications*, 7(3), Article 3. <https://doi.org/10.11648/j.ijds.20210703.12>
- Islam, M. A., Sathi, N. J., Hossain, M. T., Jabbar, A., Renzaho, A. M., & Islam, S. M. S. (2022). Caesarean delivery and its association with educational attainment, wealth index, and place of residence in Sub-Saharan Africa: A meta-analysis. *Scientific Reports*, 12(1), 5554.
- Islam, M. S., Awal, M. A., Laboni, J. N., Pinki, F. T., Karmokar, S., Mummenin, K. M., Al-Ahmadi, S., Rahman, M. A., Hossain, M. S., & Mirjalili, S. (2022). HGSORF: Henry Gas Solubility Optimization-based Random Forest for C-

- Section prediction and XAI-based cause analysis. *Computers in Biology and Medicine*, 147, 105671.
- Islam, Md. A., Shanto, H. H., Jabbar, A., & Howlader, Md. H. (2022). Caesarean Section in Indonesia: Analysis of Trends and Socio-Demographic Correlates in Three Demographic and Health Surveys (2007–2017). *Dr. Sulaiman Al Habib Medical Journal*, 4(3), 136–144. <https://doi.org/10.1007/s44229-022-00011-0>
- Jett, D., & Speer, J. (2016). Comparison of parametric and nonparametric tests for differences in distribution. *Proceedings of The National Conference On Undergraduate Research (NCUR)*, 7–9. <https://libjournals.unca.edu/ncur/wp-content/uploads/2021/06/2065-Jett-Speers-FINAL.pdf>
- Kearney, S., Raffini, L. J., Pham, T. P., Lee, X. Y., Von Mackensen, S., Landorph, A., Takedani, H., & Oldenburg, J. (2019). Health-related quality-of-life and treatment satisfaction of individuals with hemophilia A treated with turoctocog alfa pegol (N8-GP): A new recombinant extended half-life FVIII. *Patient Preference and Adherence*, Volume 13, 497–513. <https://doi.org/10.2147/PPA.S196103>
- Khawaja, M., & Al-Nsour, M. (2007a). Trends in the prevalence and determinants of caesarean section delivery in Jordan: Evidence from three demographic and health surveys, 1990–2002. *World Health & Population*, 9(4), 17–28.
- Khawaja, M., & Al-Nsour, M. (2007b). Trends in the prevalence and determinants of caesarean section delivery in Jordan: Evidence from three demographic and health surveys, 1990–2002. *World Health & Population*, 9(4), 17–28.
- Manyeh, A., Kukula, V., Odonkor, G., Rosemond, E., Adjei, A., Narh-Bana, S., Akpakli, D., & Gyapong, M. (2016). Socioeconomic and demographic determinants of birth weight in southern rural Ghana: Evidence from Dodowa Health and Demographic Surveillance System. *BMC Pregnancy and Childbirth*, 16, 160. <https://doi.org/10.1186/s12884-016-0956-2>
- Mia, M. N., Islam, M. Z., Chowdhury, M. R., Razzaque, A., Chin, B., & Rahman, M. S. (2019). Socio-demographic, health and institutional determinants of caesarean section among the poorest segment of the urban population: Evidence from selected slums in Dhaka, Bangladesh. *SSM-Population Health*, 8, 100415.
- Motlagh, S. N., Asadi-piri, Z., Bajoulvand, R., Mohseni, F. S., Bakhtiar, K., Birjandi, M., & Mansouri, M. (2020a). Factors contributing to Iranian pregnant Women's tendency to choice cesarean section. *Medical-Surgical Nursing Journal*, 9(2). <https://brieflands.com/articles/msnj-109483.pdf>
- Motlagh, S. N., Asadi-piri, Z., Bajoulvand, R., Mohseni, F. S., Bakhtiar, K., Birjandi, M., & Mansouri, M. (2020b). Factors Contributing to Iranian Pregnant Women's Tendency to Choice Cesarean Section. *Medical - Surgical Nursing Journal*, 9(2), Article 2. <https://doi.org/10.5812/msnj.109483>
- Mumtaz, S., Bahk, J., & Khang, Y.-H. (2017). Rising trends and inequalities in cesarean section rates in Pakistan: Evidence from Pakistan Demographic and Health Surveys, 1990-2013. *PloS One*, 12(10), e0186563.
- Munasir, Z., & Khoe, L. C. (2020). Immune development of children born from Caesarean section. *World Nutrition Journal*, 4(S2), 17–25.
- Nazir, S. (2015). Determinants of cesarean deliveries in Pakistan. PIDE Islamabad, Pakistan. <https://www.academia.edu/download/101153444/WorkingPaper-122.pdf>

- Niino, Y. (2011). The increasing cesarean rate globally and what we can do about it. *Bioscience Trends*, 5(4), 139–150.
- Rana, R., & Singhal, R. (2015). Chi-square test and its application in hypothesis testing. *Journal of Primary Care Specialties*, 1(1), 69–71.
- Raveh, A. (1986). On Measures of Monotone Association. *The American Statistician*, 40(2), 117–123.
<https://doi.org/10.1080/00031305.1986.10475372>
- Sandall, J., Tribe, R. M., Avery, L., Mola, G., Visser, G. H., Homer, C. S., Gibbons, D., Kelly, N. M., Kennedy, H. P., & Kidanto, H. (2018). Short-term and long-term effects of caesarean section on the health of women and children. *The Lancet*, 392(10155), 1349–1357.
- Sengupta, A., Sabastin Sagayam, M., & Reja, T. (2021). Increasing trend of C-section deliveries in India: A comparative analysis between southern states and rest of India. *Sexual & Reproductive Healthcare*, 28, 100608.
<https://doi.org/10.1016/j.srhc.2021.100608>
- “STOP Unnecessary Cesarean Section (UnCS) CAMPAIGN Developing and influencing national policy guidelines to reduce unnecessary cesarean sections in Bangladesh—Google Search.” Accessed: Jul. 02, 2023. [Online]. Available:
- Talin, I. A., Abid, M. H., Khan, M. A.-M., Kee, S.-H., & Nahid, A.-A. (2022). Finding the influential clinical traits that impact on the diagnosis of heart disease using statistical and machine-learning techniques. *Scientific Reports*, 12(1), 20199.
- “The DHS Program—Quality information to plan, monitor and improve population, health, and nutrition programs.” Accessed: Apr. 23, 2023. [Online].
- Villar, J., Carroli, G., Zavaleta, N., Donner, A., Wojdyla, D., Faundes, A., Velazco, A., Bataglia, V., Langer, A., & Narváez, A. (2007). Maternal and neonatal individual risks and benefits associated with caesarean delivery: Multicentre prospective study. *Bmj*, 335(7628), 1025.
- Volchok, E. (2020). Clear-Sighted Statistics: Module 17: Chi-Square Tests.
https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1133&context=q_b_oers
- What You Need to Know About Statistical Analysis,” *Business News Daily*. Accessed: Jul. 22, 2023. [Online].
- WHO recommends the rate CS should be 10 to 15%—Google Search. (n.d.). Retrieved April 9, 2024, from <https://www.google.com/search?>
- “WHO Statement on Caesarean Section Rates—PMC.” Accessed: Jul. 02, 2023. [Online].
- Wong, E., Wei, T., Qi, Y., & Zhao, L. (2008). A crosstab-based statistical method for effective fault localization. 2008 1st International Conference on Software Testing, Verification, and Validation, 42–51.
<https://ieeexplore.ieee.org/abstract/document/4539531/>
- Yates, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika*, 35(1/2), 176–181.