Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/4hf0xd34

ANALYSIS TO EXPLORE THE PREVALENCE RATE OF POLIO IN KHYBER PAKHTUNKHWA, PAKISTAN: CAUSES, CONSEQUENCES AND REMEDIES

Sofia[1*], Qamruz Zaman[2], Soofia Iftikhar[3], Najma Salahuddin[3], Mansoor Ahmad[2], Aizaz Shah[2], Mir Ullah[2], Neelam[2]

[1*]College of Home Economics, University of Peshawar, Pakistan
[2]Department of Statistics, University of Peshawar, Pakistan
[3]Department of Statistics, Shaheed Benazir Women University Peshawar, Pakistan

***Corresponding author:** Sofia
*Email: cricsportsresearchgroup@gmail.com

Abstract
This study investigated the association of polio with demographic, social factors, and living conditions in Khyber Pakhtunkhwa, focusing on patients in Peshawar. A case-control approach was employed, collecting data from 222 polio cases and 100 healthy controls, selected from relatives of patients. The analysis found a significantly higher incidence of polio among families with incomes below Rs. 20,000. Chi-square tests and odds ratios indicated that males are at higher risk for polio compared to females (Odds Ratio = 1.586). Individuals without proper sanitation facilities were also more susceptible to polio (Odds Ratio = 0.081). Logistic regression identified several key risk factors, including sources of information about polio, previous exposure to polio patients, dietary practices such as consuming hot food post-vaccination, duration of symptoms, and the effectiveness of polio vaccination in preventing other diseases. Resistance to receiving polio vaccinations and inadequate sanitation also emerged as significant factors. Furthermore, cluster analysis using the gap statistic, silhouette statistic, and elbow method revealed that three clusters best represent the data, providing robust insights and reinforcing the study's findings. This comprehensive analysis highlights the importance of addressing socioeconomic and sanitation-related factors and suggests targeted strategies for effective polio eradication efforts. The results underscore the need for tailored public health interventions to combat polio effectively in Pakistan.

Keywords: A case-control study, Chi-square test, Odds Ratio, Logistic Regression, Cluster Analysis

1. Introduction
Polio remains a major health problem for Pakistani children due to insufficient vaccination. In 2010, Pakistan had 144 cases, more than Nigeria, India, and Afghanistan. Polio mainly affects children under five, causing symptoms like fever and limb pain. It can't be cured but can be prevented with vaccines. Despite government efforts, many children remain unvaccinated and vulnerable. Vaccinations need to be more consistently available across all districts. New cases still occur, especially among unvaccinated children from migrant families, with a recent case in Karachi involving a family from South Waziristan. Polio case numbers have fluctuated over the years, showing a concerning trend (World Health Organization, & Global Polio Eradication Initiative, 2019). In 2009,

89 polio cases were reported across 32 districts in Pakistan, mainly in Khyber Pakhtunkhwa (KPK), Federally Administered Tribal Areas (FATA), Punjab, Sindh, and Baluchistan. By 2010, cases had increased, with KPK and FATA accounting for over 50% of the total cases. High-risk areas include Bajaur Agency, Khyber Agency, Mohmand Agency, Pishin, Killa Abdullah, and Quetta. (National Institute of Health, Pakistan., 2023).

Polio cases in Pakistan are high due to population movements from unsafe areas with unvaccinated children. Along with Afghanistan, Pakistan is one of the most affected countries, with migration worsening the spread. Ineffective strategies have hindered eradication efforts. The Global Polio Eradication Initiative (GPEI), launched in 1988, has significantly reduced polio, but by 2015, it remained endemic in Pakistan and Afghanistan. By 2024, Pakistan has made progress in reducing cases, but challenges remain, especially in high-risk areas. Continued efforts in vaccine delivery and combating misinformation are essential for achieving polio-free status (Global Polio Eradication, 2018).

Polio has been known since 1789, with major outbreaks in the early 20th century. In 1988 the WHO launched the Global Polio Eradication Initiative (GPEI). The U.S. became polio-free in 1994, followed by the Western Pacific in 2000 and Europe in 2002. By 2015, only Pakistan, Afghanistan, and Nigeria remained polio-endemic. Polio spreads through the fecal-oral route, causing muscle weakness and paralysis, in some cases leading to death. While vaccination has eliminated polio in most countries, poor sanitation hinders eradication in Nigeria, Afghanistan, and Pakistan (History of Polio, 2019). In Pakistan, sociocultural factors and shortages of medical resources, along with conspiracy theories about vaccinations being a foreign plot, complicate polio eradication. These theories have a strong influence on public health efforts. Although many reports and studies highlight the difficulties of eradicating polio, the disease has not been eliminated globally. The aim is to compare the economic impact of eradicating polio versus ongoing control measures. Analysis of data from vaccination records, population statistics, and historical costs for immunization, campaigns, surveillance, labs, technical support, outreach, and treatment shows that achieving polio-free status could be more cost-effective than continued control efforts (Zimmermann et.al, 2019). Poliomyelitis, caused by a Picornaviridae virus, is highly infectious and can range from no symptoms to severe paralysis. It includes asymptomatic infection, mild illness, nonparalytic poliomyelitis, and paralytic poliomyelitis (Mehndiratta et.al, 2014). A virus from the Picornaviridae family causes poliomyelitis, a contagious disease. Symptoms range from mild to severe paralysis and are classified as inapparent infection, mild illness, aseptic meningitis, and paralytic poliomyelitis (Mehndiratta et.al, 2014).

Polio continues to threaten Pakistan, Nigeria, and Afghanistan. Sociological, political, and epidemiological factors contribute to its persistence. Key issues include the spread of conspiracy theories, partly due to tactics in the Bin Laden search, and low vaccination education, especially in Pakistan where illiteracy is high. Coordinated efforts from relevant departments can address these challenges (Andrade et.al, 2018). In 1988, a global initiative was started to eradicate polio. By 2010, an economic analysis of the Global Polio Eradication Initiative (GPEI) reviewed post-eradication policies and found economic reasons to continue the program despite increasing costs (Tebbensa et.al, 2010). The GPEI saw major progress, with polio cases dropping by 99% by 2000, from 350,000 in 125 countries to significantly fewer. By 2002, the WHO declared the Americas, Western Pacific, and European regions polio-free. However, by 2005, eradication efforts in Pakistan, India, and Nigeria faced setbacks, leading to 1,500 cases of paralysis. Despite strategic plans addressing these challenges, Pakistan reported no polio cases for five months by 2011 but remained high-risk due to insufficient funding (Aylward et.al, , 2011).

In 1988, there were 350,000 polio cases globally. By 2015, wild poliovirus was found only in Afghanistan and Pakistan. Ensuring children under one receive three doses of oral polio vaccine (OPV) is essential. Data from Nigeria's surveys showed low OPV coverage in Northern and rural areas, and among children of less educated mothers, with coverage at 12% for uneducated mothers versus 81% for educated mothers in 2008. To improve coverage, campaigns should target these areas and educate mothers, boosting vaccination rates and supporting global eradication efforts.

This study examines qualitative challenges faced by polio workers and reasons for parental vaccine refusal. It aims to analyze polio-affected households, identifying risk factors like age, sex, region, and education level. The study will explore high-risk groups' living conditions and specific risk factors for polio. The article is organized as follows: a methodology section covering population, sampling design, sample size, data collection, and statistical methods; a statistical analysis section using descriptive statistics and binary logistic models; and a concluding section with final thoughts.

2. Research Objectives

- To explore the living conditions of the high-risk groups in Khyber Pakhtunkhwa.
- To identify the risk factors associated with the disease in Khyber Pakhtunkhwa.

3. Methodology of the study

The methodology of the study is based on as follows:

3.1 Sampling strategy

The study uses a multistage probability sampling procedure with Khyber Pakhtunkhwa (KP) as the universe. Polio-affected cities in KP are classified through cluster Analysis, followed by a K-mean algorithm to cluster data. Then affected families were randomly selected, followed by a random selection of affected families.

3.2 Sample size

To ensure a sufficient and representative sample of affected families, the study will use a proportion-based formula to estimate sample size. This approach is suitable given the categorical nature of the study variable, Polio presence.

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{e}\right)^2 P(1-P) \qquad (1)$$

with $Z_{\frac{\alpha}{2}}$ set at 1.96 for a 95% confidence level, "e" as the margin of error (0.05), and "P" representing the Polio prevalence (0.70, DHIMS-2018), substituting these values into the formula gives the estimated sample size:

$$n = \left[{}^{1.96}/_{0.05}\right]^2 (0.70)(1-.70) = 322$$

This means 322 Polio-affected families will be selected across KP. Additionally, a sample of 100 families from KP will be included in the control group to assess significant Polio risk factors.

3.3 Data analysis

The analysis includes descriptive Analysis such as frequency distributions, graphs, and charts, and Cluster Analysis such as Silhouette Analysis and Elbow Method. Inferential statistics involve chi-square tests for association, odds ratio analysis, and Multivariate analysis based on logistic regression modeling for binary dependent variables like Polio presence.

3.3.1 Descriptive analysis

Descriptive statistics summarize and present data, covering qualitative and quantitative variables through tabulation and graphs. Tabulation includes frequency distributions of key demographics, while summary statistics compute measures like mean, median, and mode for representative values, and variance and standard deviation for variability.

3.3.2 Cluster Analysis

Clustering, a key technique in machine learning and data analysis, groups similar data points based on their features to partition a dataset into clusters where items within each cluster are more alike than those in other clusters. This technique aids in exploring and summarizing complex datasets, enhancing data interpretation, and supporting decision-making. Two common methods for evaluating clustering results are Silhouette Analysis and the Elbow Method. Silhouette Analysis measures how similar each data point is to its cluster compared to other clusters, with scores ranging from -1 to 1 to indicate clustering quality. The Elbow Method involves plotting the within-cluster sum of squares against the number of clusters to find the optimal cluster count, identified by a point where the rate of decrease sharply changes. Clustering is widely applied in fields like customer segmentation, image segmentation, outlier detection, and document clustering (Rousseeuw, 1987) (Tibshirani, 2001).

3.3.3 Chi-square test and ODDS ratio

Cross-tabulation, or contingency table analysis, is valuable for analyzing categorical data measured on a nominal scale. It entails creating a two-dimensional table that displays the frequency of respondents with specific characteristics. This method uses Chi-square statistics to test the association between two categorical variables, also known as the chi-square test of association (Rana,et.al, 2015). The null hypothesis states no association (independence), while the alternative hypothesis suggests an association (dependency) (Franke et.al, 2012).
The test statistic is calculated using the formula:

$$\chi^2 = \frac{[ad-bc]^2}{n[(a+b)(c+d)(a+c)(b+d)]} \sim \chi^2_{\alpha(1)} \qquad (2)$$

Here, a, b, c and d represent the frequencies in the respective cells of the contingency table, and n is the total sample size. To assess the strength of association between the two categorical variables, the odds ratio (OR) is commonly used (Bland, et.al, 2000). The odds ratio measures the strength of association between two qualitative variables, assuming each variable has two categories (Kraemer, et.al, 2004).
The odds are defined as:

$$OR = \frac{P_1(1-P_2)}{P_2(1-P_1)} \qquad (3)$$

3.3.4 Logistic Regression

Logistic regression analyzes the impact of multiple independent variables on a categorical response variable, typically binary or nominal. In this context, the response variable indicates the presence (coded as 1) or absence (coded as 0) of a specific characteristic, such as being affected by polio. The logistic regression model determines the mathematical relationship between the binary response variable and a set of predictors using a logit transformation:

$$logit(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \dots \dots \dots \dots \dots \dots + \beta_k X_k \qquad (4)$$

Here, P represents the probability of the presence of polio in a family. The logit transformation $logit(P) = \ln\left[\frac{P}{1-P}\right]$ is the logged odds, where $\frac{P}{1-P}$ is the odds of having the disease. The logistic regression model quantifies how the odds change with each unit increase in an independent variable X_i, while keeping other variables constant (Kleinbaum, et.al, 2002). Specifically, $e^{\beta i}$ represents the odds ratio for X_i, indicating the relative increase or decrease in the odds of the outcome when X_i increases by one unit (LaValley, 2008).

4. Findings and Analysis

The data is analyzed using descriptive statistics and binary logistic models. The results are tabulated and explained.

Table 1: Description of Patients Gender wise

Sex	N (%)	Mean	S. D	S.E(X)	Min	Max	Skew	Kurt	95% C.I	
									Lower	Upper
Male (1)	242	32.79	7.967	0.512	18	54	0.363	-0.186	31.785	33.794
Female (2)	80	33.83	9.347	1.045	19	56	0.650	0.007	31.782	35.878

Table 1 shows the gender distribution of 322 patients in the study. The case group includes 100 males (68.9%) and 109 females (52.7%), while the control group has 46 males (49.5%) and 47 females (50.5%). The overall average age is 33.04 months (± 8.328 SD), ranging from 18 to 56 months.

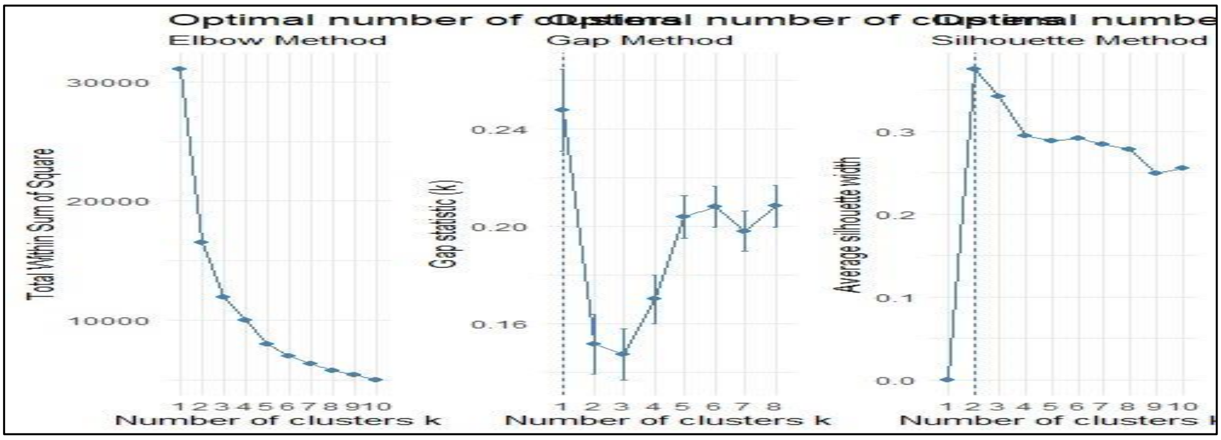

Figure 1: Optimal number of clusters using elbow, gap, and silhouette methods

Figure 1 shows how we can determine the best number of clusters for the dataset using three different methods: the gap statistic, the silhouette statistic, and the elbow method. The elbow method works by plotting the total within-cluster variation against the number of clusters and looking for a "bend" in the curve. In this case, the curve clearly shows that adding more than four clusters doesn't significantly reduce the variation within clusters. This suggests that four clusters might be a good choice. The gap statistic provides another way to decide the optimal number of clusters by comparing the change in within-cluster dispersion to what we would expect under a reference distribution of the data. According to the gap statistic, having three clusters is optimal, but four clusters could also work. The silhouette statistic, which measures how similar an object is to its own cluster compared to other clusters, also indicates that having three clusters is the best option for this dataset. All three methods largely agree that breaking the data into three clusters is the most effective strategy. Thus, based on these analyses, the ideal number of clusters is three, providing a clear and consistent result across the different statistical methods used.

Table 2: Age Distribution of Patients

Age Group	Frequency	% age	Cumulative % age
"16------ 30"	144	44.7	44.7
"31 ------ 45"	152	47.2	91.9
"46 ------ 60"	26	8.1	100.0
Total	**322**	**100.0**	**-------**

Table 2 shows that 242 patients are male and 80 are female. The maximum age for both sexes is 56 months, while the minimum age is 18 months for males and 19 months for females.

Table 3: Gender wise Frequency Table.

Group	Males	Females	Total (%)
Case	173	49	222(68.9)
Control	69	31	100 (31.1)
Total	**242**	**80**	**322**

Table 3 shows that 222 patients (68.9%) are in the case group and 100 in the control group. Among the case group, there are 173 males and 49 females.

The χ^2 value is 2.943 with a P-value of 0.086, indicating no significant association between gender distribution and group.

Table 4 Proper Flesh System Facility

Group	Yes	No	Total (%)
Case	20	202	222(68.9)
Control	55	45	100 (31.1)
Total	**75**	**247**	**322**

The Odds Ratio of 1.586 indicates that males have higher odds of contracting polio compared to females. This means that being male is associated with a greater likelihood of getting polio. Table 4 shows data from a study involving 222 patients. Out of these, 20 patients come from families with proper sanitation facilities, while 202 patients are from families without such facilities. The study found a significant association between the presence of proper sanitation facilities and the incidence of polio, as shown by the chi-square test result ($\chi^2 = 81.620$, $p < 0.001$). This strong statistical significance suggests that there is a clear relationship between having adequate sanitation systems and the occurrence of polio cases. The Odds Ratio of 0.081 further supports this finding by indicating that families without proper sanitation facilities have much higher odds of polio cases compared to families with proper facilities. This suggests that inadequate sanitation significantly increases the likelihood of polio, emphasizing the importance of proper hygiene and sanitation in reducing the risk of the disease.

Table 5 Socio-Economic Status of Patients

Category	Frequency	Percentage
<=10000	19	8.6%
10000<Income<=20000	188	84.7%
20000<Income<=30000	9	4.1%
>30000	6	2.7%
Total	**222**	**100%**

Table 5 shows that 19 (8.6%), 188 (84.7%), 9 (4.1%), and 6 (2.7%) patients are from families earning less than Rs. 10,000, less than Rs. 20,000, less than Rs. 30,000, and more than Rs. 30,000, respectively. Polio incidence was significantly higher among families earning less than Rs. 20,000.

Table 6 Washing hands with soap after attending washroom

	Frequency	Percentage
Yes	14	6.3
No	208	93.7
Total	**222**	**100.0**

From Table 6, it is observed that out of 222 patients, only 14 (6.3%) reported washing hands with soap after using the washroom, while the majority, 93.7%, did not.

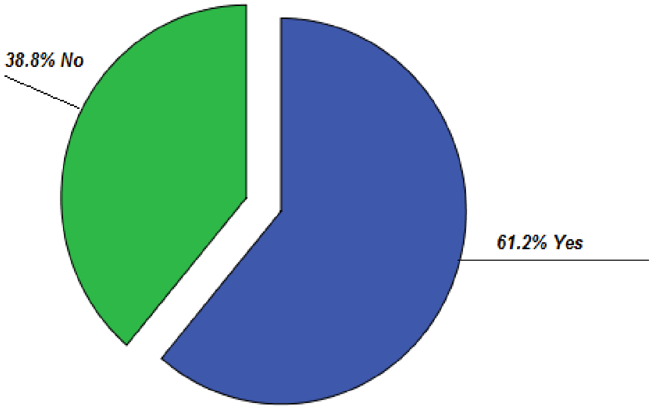

Figure 2: Hot food stuff can be given just after the Polio drop

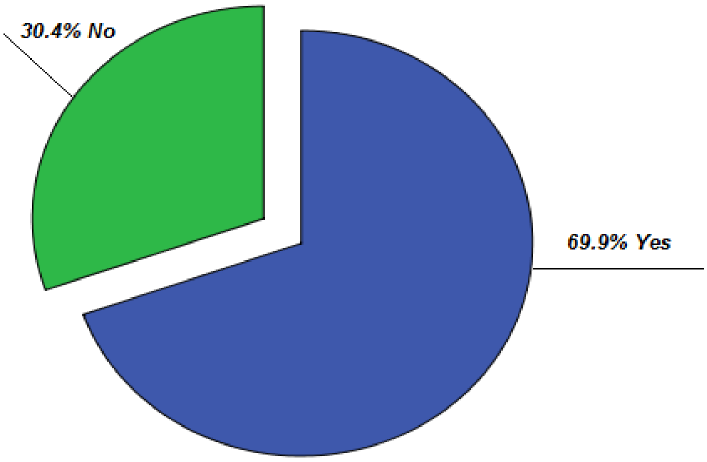

Figure 3: Transmission of Polio form one person to other

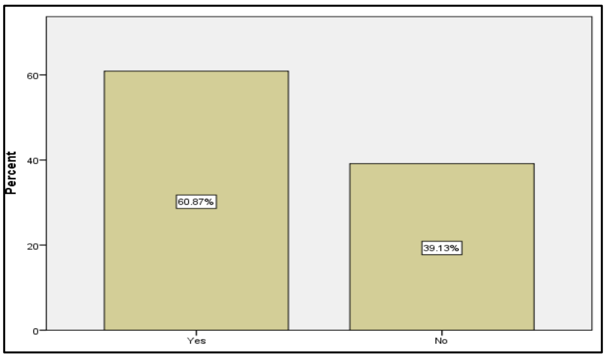

Figure 4: Are you forced of not giving Polio dose

Data from 322 individuals were analyzed. In Figure 2, 61.2% agree that a child can eat fast food within 30 minutes of a polio drop, while 38.8% disagree. Figure 3 shows that 69.9% believe polio is transmissible, and 30.1% disagree.

In Figure 4, 60.87% of participants said they were pressured by others to withhold the dose from their children, while 39.13% made independent decisions without external influence.

4.1 Model diagnostics for overall Polio Respondent

Model diagnostics tests include three plots to detect outliers: Cook's distance, standardized residuals, and leverage values. The analysis of the diagnostic model is divided into three steps: overall, males, and females.

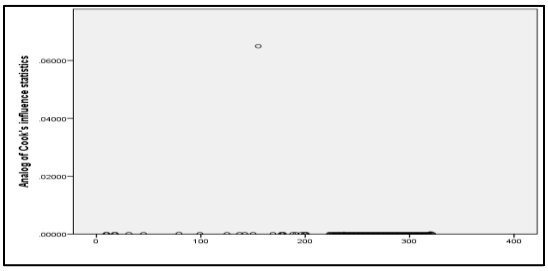

Figure 5: Index plot of Cook's distance

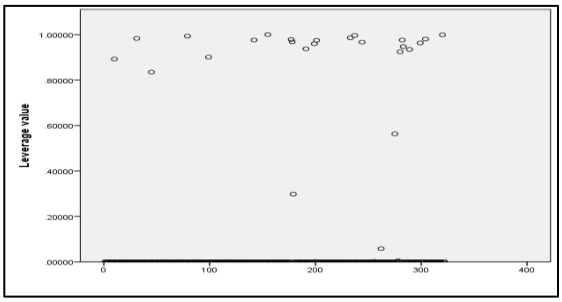

Figure 6: Index plot of leverage values

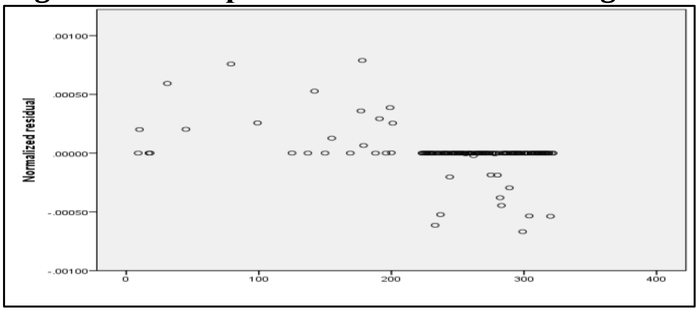

Figure 7: Index plot of standardized residuals

Figure 5 shows observations closely grouped, with one slightly distant point, though minimally so. Figures 6 and 7 depict leverage points and standardized residuals. Figure 6 confirms no outliers or influential observations. Figure 4.7 shows standardized residuals within the ±3 range, indicating no outliers.

4.2 Model Fitting of Overall Polio Respondents

The forward selection procedure was employed to select the best model based on Likelihood, Cox & Snell R Square, Nagelkerke R Square criteria, followed by a parsimony approach to choose representative factors. The procedure is summarized in thirteen steps, as detailed in the table below.

Table 7: Model Summary

Step	Cox & Snell R Square	Nagelkerke R Square
1	.399	.561
2	.505	.711
3	.550	.774
4	.584	.822
5	.617	.868
6	.639	.900
7	.654	.921
8	.665	.936
9	.675	.950
10	.684	.963
11	.704	.990
12	.710	1.000
13	.710	1.000

Table 7 summarizes a 13-step forward selection procedure to find the best model based on Cox & Snell R-Square and Nagelkerke R-Square values. The values gradually increased until the seventh step, after which they showed slower increments. Using a parsimony approach, the model selected at the seventh step was considered the best. Table 8 provides a detailed summary of all steps in the procedure.

Table 8 : Summary of the Overall Models through Forward Selection Procedure

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Sgn	1.985	.196	102.089	1	.000	7.276
	Constant	-1.058	.211	25.169	1	.000	.347
Step 2	Drp	2.961	.437	45.829	1	.000	19.308
	Sgn	2.075	.251	68.384	1	.000	7.966
	Constant	-5.639	.788	51.230	1	.000	.004
Step 3	F	2.553	.514	24.695	1	.000	12.847
	Drp	2.886	.468	38.101	1	.000	17.922
	Sgn	2.045	.269	57.602	1	.000	7.731
	Constant	-9.855	1.312	56.396	1	.000	.000
Step 4	F	2.886	.607	22.613	1	.000	17.917
	T	-2.398	.516	21.584	1	.000	.091
	Drp	2.686	.521	26.592	1	.000	14.678
	Sgn	1.971	.291	45.949	1	.000	7.181
	Constant	-6.481	1.511	18.405	1	.000	.002
Step 5	F	3.829	.800	22.913	1	.000	46.029
	T	-2.965	.672	19.480	1	.000	.052
	Drp	2.614	.614	18.125	1	.000	13.652
	Sgn	2.686	.428	39.316	1	.000	14.678
	Hotf	3.445	.829	17.275	1	.000	31.345
	Constant	-12.226	2.421	25.510	1	.000	.000
Step 6	F	4.083	.920	19.713	1	.000	59.350
	Pn	-2.742	.700	15.339	1	.000	.064
	T	-2.901	.746	15.129	1	.000	.055
	Drp	2.608	.711	13.437	1	.000	13.570
	Sgn	2.692	.494	29.676	1	.000	14.757
	Hotf	4.247	1.054	16.224	1	.000	69.872
	Constant	-9.592	2.748	12.180	1	.000	.000
Step 7	F	4.773	1.171	16.610	1	.000	118.243
	Pn	-2.809	.811	12.009	1	.001	.060
	T	-3.213	.926	12.041	1	.001	.040
	Drp	2.813	.792	12.632	1	.000	16.667
	Sgn	2.973	.622	22.864	1	.000	19.548
	Hotf	4.645	1.228	14.314	1	.000	104.093
	Pv	-1.829	.571	10.276	1	.001	.161
	Constant	-10.005	3.136	10.177	1	.001	.000
Step 8	F	5.247	1.410	13.848	1	.000	190.022
	Pn	-3.702	1.161	10.176	1	.001	.025
	T	-3.796	1.155	10.800	1	.001	.022
	Drp	2.744	.875	9.842	1	.002	15.546

	Sgn	3.022	.729	17.180	1	.000	20.534
	Hotf	5.347	1.620	10.890	1	.001	209.954
	Pexp	2.677	1.001	7.147	1	.008	14.544
	Pv	-1.770	.626	7.995	1	.005	.170
	Constant	-13.942	4.145	11.311	1	.001	.000
Step 9	F	5.889	1.733	11.546	1	.001	361.024
	Pn	-4.177	1.505	7.706	1	.006	.015
	T	-3.972	1.340	8.792	1	.003	.019
	Drp	3.889	1.227	10.040	1	.002	48.861
	Sgn	2.811	.823	11.677	1	.001	16.621
	Hotf	5.888	2.127	7.664	1	.006	360.654
	Pexp	3.082	1.199	6.607	1	.010	21.801
	Pv	-2.193	.793	7.643	1	.006	.112
	S	1.016	.400	6.452	1	.011	2.761
	Constant	-19.183	5.878	10.650	1	.001	.000
Step 10	F	6.621	2.509	6.965	1	.008	750.508
	H	6.259	2.987	4.391	1	.036	522.621
	Pn	-8.018	3.308	5.876	1	.015	.000
	T	-5.885	2.307	6.509	1	.011	.003
	Drp	4.607	1.715	7.215	1	.007	100.231
	Sgn	5.171	2.029	6.494	1	.011	176.020
	Hotf	9.479	3.736	6.438	1	.011	13085.572
	Pexp	5.797	2.535	5.230	1	.022	329.365
	Pv	-3.203	1.371	5.457	1	.019	.041
	S	1.408	.606	5.402	1	.020	4.090
	Constant	-35.381	13.383	6.990	1	.008	.000
Step 11	F	195.284	1661.021	.014	1	.906	6.468E+84
	H	306.565	2687.655	.013	1	.909	1.378E+133
	Pn	-250.460	2129.422	.014	1	.906	.000
	T	-132.012	1110.657	.014	1	.905	.000
	Drp	206.930	1766.861	.014	1	.907	7.390E+89
	Sgn	186.123	1605.393	.013	1	.908	6.793E+80
	Hotf	283.709	2444.197	.013	1	.908	1.634E+123
	WtisPolio	174.570	1537.760	.013	1	.910	6.529E+75
	Pexp	119.937	1050.981	.013	1	.909	1.225E+52
	Pv	-77.495	671.573	.013	1	.908	.000
	S	76.818	671.573	.013	1	.909	22992792753
	Constant	-1723.622	14927.495	.013	1	.908	.000
Step 12	F	128.051	1825.402	.005	1	.944	4.090E+55
	H	193.324	2969.196	.004	1	.948	9.113E+83
	Pn	-139.142	2138.156	.004	1	.948	.000
	T	-101.703	1490.291	.005	1	.946	.000
	Drp	145.727	1970.713	.005	1	.941	1.944E+63
	Sgn	109.377	1485.844	.005	1	.941	3.175E+47
	Hotf	206.085	2835.911	.005	1	.942	3.173E+89
	WtisPolio	117.386	2055.489	.003	1	.954	9.548E+50

	Age	3.420	97.270	.001	1	.972	30.559
	Pexp	79.291	1271.087	.004	1	.950	27256000597
	Pv	-66.586	951.717	.005	1	.944	.000
	S	37.570	545.733	.005	1	.945	20723149147
	Constant	-1232.361	17643.277	.005	1	.944	.000
Step 13	F	128.117	1833.477	.005	1	.944	4.368E+55
	H	193.395	2979.383	.004	1	.948	9.782E+83
	Ipp	-51.016	8108.256	.000	1	.995	.000
	Pn	-88.191	8159.826	.000	1	.991	.000
	T	-101.760	1497.325	.005	1	.946	.000
	Drp	145.814	1980.387	.005	1	.941	2.119E+63
	Sgn	109.411	1488.519	.005	1	.941	3.284E+47
	Hotf	206.158	2842.023	.005	1	.942	3.414E+89
	WtisPolio	117.426	2065.113	.003	1	.955	9.938E+50
	Age	3.424	98.305	.001	1	.972	30.695
	Pexp	79.303	1275.362	.004	1	.950	27602208497
	Pv	-66.626	956.555	.005	1	.944	.000
	S	37.583	547.226	.005	1	.945	20995428356
	Constant	-1232.880	17706.183	.005	1	.944	.000

Steps 11 to 13, which consisted entirely of insignificant factors, were excluded, resulting in the reduction of the table to ten steps.

Table 9 : Summary of the Ten Selected Overall Models through Forward Selection Procedure

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Sgn	1.985	.196	102.089	1	.000	7.276
	Constant	-1.058	.211	25.169	1	.000	.347
Step 2	Drp	2.961	.437	45.829	1	.000	19.308
	Sgn	2.075	.251	68.384	1	.000	7.966
	Constant	-5.639	.788	51.230	1	.000	.004
Step 3	F	2.553	.514	24.695	1	.000	12.847
	Drp	2.886	.468	38.101	1	.000	17.922
	Sgn	2.045	.269	57.602	1	.000	7.731
	Constant	-9.855	1.312	56.396	1	.000	.000
Step 4	F	2.886	.607	22.613	1	.000	17.917
	T	-2.398	.516	21.584	1	.000	.091
	Drp	2.686	.521	26.592	1	.000	14.678
	Sgn	1.971	.291	45.949	1	.000	7.181
	Constant	-6.481	1.511	18.405	1	.000	.002
Step 5	F	3.829	.800	22.913	1	.000	46.029
	T	-2.965	.672	19.480	1	.000	.052
	Drp	2.614	.614	18.125	1	.000	13.652
	Sgn	2.686	.428	39.316	1	.000	14.678
	Hotf	3.445	.829	17.275	1	.000	31.345
	Constant	-12.226	2.421	25.510	1	.000	.000
Step 6	F	4.083	.920	19.713	1	.000	59.350
	Pn	-2.742	.700	15.339	1	.000	.064

	T	-2.901	.746	15.129	1	.000	.055
	Drp	2.608	.711	13.437	1	.000	13.570
	Sgn	2.692	.494	29.676	1	.000	14.757
	Hotf	4.247	1.054	16.224	1	.000	69.872
	Constant	-9.592	2.748	12.180	1	.000	.000
Step 7	F	4.773	1.171	16.610	1	.000	118.243
	Pn	-2.809	.811	12.009	1	.001	.060
	T	-3.213	.926	12.041	1	.001	.040
	Drp	2.813	.792	12.632	1	.000	16.667
	Sgn	2.973	.622	22.864	1	.000	19.548
	Hotf	4.645	1.228	14.314	1	.000	104.093
	Pv	-1.829	.571	10.276	1	.001	.161
	Constant	-10.005	3.136	10.177	1	.001	.000
Step 8	F	5.247	1.410	13.848	1	.000	190.022
	Pn	-3.702	1.161	10.176	1	.001	.025
	T	-3.796	1.155	10.800	1	.001	.022
	Drp	2.744	.875	9.842	1	.002	15.546
	Sgn	3.022	.729	17.180	1	.000	20.534
	Hotf	5.347	1.620	10.890	1	.001	209.954
	Pexp	2.677	1.001	7.147	1	.008	14.544
	Pv	-1.770	.626	7.995	1	.005	.170
	Constant	-13.942	4.145	11.311	1	.001	.000
Step 9	F	5.889	1.733	11.546	1	.001	361.024
	Pn	-4.177	1.505	7.706	1	.006	.015
	T	-3.972	1.340	8.792	1	.003	.019
	Drp	3.889	1.227	10.040	1	.002	48.861
	Sgn	2.811	.823	11.677	1	.001	16.621
	Hotf	5.888	2.127	7.664	1	.006	360.654
	Pexp	3.082	1.199	6.607	1	.010	21.801
	Pv	-2.193	.793	7.643	1	.006	.112
	S	1.016	.400	6.452	1	.011	2.761
	Constant	-19.183	5.878	10.650	1	.001	.000
Step 10	F	6.621	2.509	6.965	1	.008	750.508
	H	6.259	2.987	4.391	1	.036	522.621
	Pn	-8.018	3.308	5.876	1	.015	.000
	T	-5.885	2.307	6.509	1	.011	.003
	Drp	4.607	1.715	7.215	1	.007	100.231
	Sgn	5.171	2.029	6.494	1	.011	176.020
	Hotf	9.479	3.736	6.438	1	.011	13085.572
	Pexp	5.797	2.535	5.230	1	.022	329.365
	Pv	-3.203	1.371	5.457	1	.019	.041
	S	1.408	.606	5.402	1	.020	4.090
	Constant	-35.381	13.383	6.990	1	.008	.000

Upon comparing the models at the 7th and 10th steps, it was found that the values of the selection criteria increased insignificantly. Therefore, the model selected at the 7th step was deemed the best, as summarized in Table 10.

Table 10: Summary of Best Selected Model

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 7	F	4.773	1.171	16.610	1	.000	118.243
	Pn	-2.809	.811	12.009	1	.001	.060
	T	-3.213	.926	12.041	1	.001	.040
	Drp	2.813	.792	12.632	1	.000	16.667
	Sgn	2.973	.622	22.864	1	.000	19.548
	Hotf	4.645	1.228	14.314	1	.000	104.093
	Pv	-1.829	.571	10.276	1	.001	.161
	Constant	-10.005	3.136	10.177	1	.001	.000

So, the above table was summarized in the equation form as:

$$\boldsymbol{Logit(\widehat{Y}) = -10.005 - 1.829\ P_v + 4.645\ Hotf + 2.973\ Drp - 3.213\ T - 2.809\ Pn + 4.773\ F}$$

$$\boldsymbol{Logit(\widehat{Y}) = -10.005 - 1.829 \textit{Polio is caused by virus} + 4.645\ Can\ hot\ food\ stuff + 2.973\ \ Polio\ signs\ Duration + 2.813\ Polio\ prevent\ other\ diseases\ - 3.213 Do\ \ Polio\ tested\ - 2.809\ not\ giving\ Polio\ dose + 4.773\ proper\ flesh\ system}$$

The model identified seven significant factors: awareness about the virus, hot food, duration of polio symptoms, drops curing other diseases, polio testing procedure, being forced not to give doses, and availability of a proper flesh system. Table 4.10, which includes regression coefficients, standard error, Wald statistics, P-Value, and odds ratio, shows that gender was not significant. There is a strong association between polio and virus awareness, with unawareness reducing the odds to 0.161 times. The odds ratio for hot food is 104.093. Knowledge of symptom duration increases the odds by 19.548 times. Polio symptoms are strongly associated with disease prevention. The polio testing procedure and being forced not to give doses are significant but have opposite coefficients. Knowledge of polio symptoms and a proper flesh system are also strongly associated. Since all factors are categorical, Table 11 further elaborates on the risk factors.

Table 11: Detailed Analysis of Best Selected Model

Factors		B	S.E.	Wald	df	Sig.	Exp(B)
	Pv(1)	2.972	1.180	6.347	1	.012	19.538
	Pv(2)	-.077	1.154	.004	1	.947	.926
	Hotf(1)	-5.152	1.402	13.502	1	.000	.006
	Sgn(1)	-6.124	1.379	19.720	1	.000	.002
	Sgn(2)	-2.175	1.525	2.035	1	.154	.114
	Drp(1)	-3.109	.848	13.431	1	.000	.045
	T(1)	3.416	1.026	11.076	1	.001	30.444
	Pn(1)	2.757	.854	10.410	1	.001	15.745
	F(1)	-5.081	1.241	16.775	1	.000	.006
	Constant	6.314	1.878	11.306	1	.001	552.276

Table 11 shows that Pv(1) is significant, indicating most respondents believe the disease is caused by a virus. The negative coefficient for Hotf(1) suggests a misconception about giving hot food within half an hour after the polio drop. Symptom duration has negative coefficients for both 1-2 days and 3-35 days. The belief that polio drops cannot prevent other diseases is significant. Most people know polio can be tested through stool or throat tests. Being forced by family or society members is a major risk factor. Lastly, the use of a proper flesh system has a negative coefficient.

3.3 Model Diagnostics for Male Model

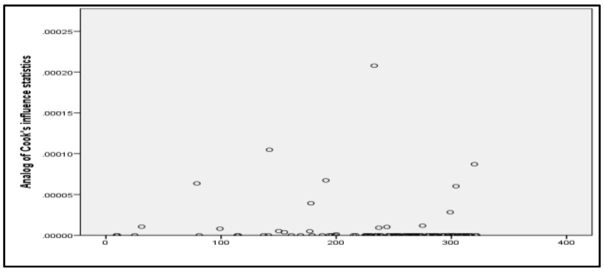

Figure 8: Index plot of Cook's Distance male

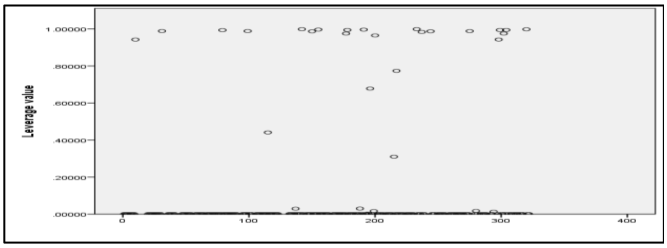

Figure 9: Index plot of leverage values for male

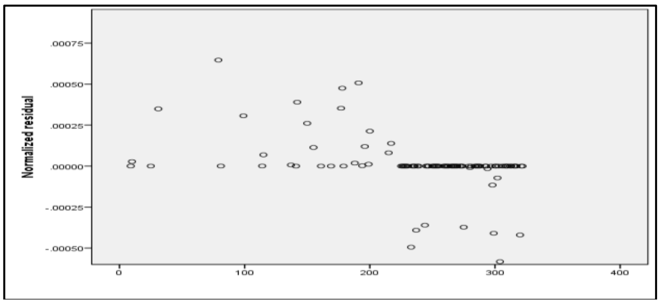

Figure 10: Index plot of standardized residuals for male

Figure 8 to 10 show the diagnostics tests for the male respondents. These figures give the satisfactory results and indicated that the data is free from any type of outliers

4.3 Model Fitting of Male Polio Respondents

Table 12 summarizes a 13-step forward selection procedure, detailing Cox & Snell R-Square and Nagel Kreke R-Square values for each step. Values showed a gradual increase from the 9th to the 13th steps, leading to the selection of the 8th step model as the best using a parsimony approach.

Table 12: Model Summary for Male

	Step	Cox & Snell R Square	Nagelkerke R Square
Male	1	.358	.513
	2	.491	.703
	3	.539	.773
	4	.584	.838
	5	.613	.879
	6	.630	.904
	7	.641	.919
	8	.657	.942
	9	.666	.955
	10	.678	.973
	11	.697	1.000
	12	.697	1.000
	13	.697	1.000

The details of each model are summarized in Table 13, which includes the selected model comprising eight significant factors.

Table 13: Summary of the Male Models through Forward Selection Procedure

			B	S.E.	Wald	df	Sig.	Exp(B)
Male	Step 1[a]	Sgn	1.837	.219	70.269	1	.000	6.275
		Constant	-.797	.233	11.661	1	.001	.451
	Step 2[b,n]	Drp	3.281	.539	37.104	1	.000	26.601
		Sgn	2.035	.299	46.463	1	.000	7.654
		Constant	-6.003	.985	37.129	1	.000	.002
	Step 3[c,o]	T	-2.421	.534	20.526	1	.000	.089
		Drp	3.232	.590	29.954	1	.000	25.324
		Sgn	1.996	.327	37.226	1	.000	7.358
		Constant	-2.271	1.170	3.769	1	.052	.103
	Step 4[d,p]	F	3.343	.805	17.252	1	.000	28.317
		T	-3.032	.720	17.711	1	.000	.048
		Drp	3.770	.755	24.937	1	.000	43.379
		Sgn	2.121	.390	29.606	1	.000	8.341
		Constant	-7.991	1.924	17.242	1	.000	.000
	Step 5[e]	F	3.475	.904	14.793	1	.000	32.301
		Pn	-2.815	.777	13.112	1	.000	.060
		T	-3.349	.817	16.807	1	.000	.035
		Drp	4.321	.981	19.384	1	.000	75.230
		Sgn	2.007	.432	21.604	1	.000	7.439
		Constant	-3.933	2.335	2.836	1	.092	.020
	Step 6[f]	F	4.402	1.151	14.615	1	.000	81.591
		Pn	-3.016	.864	12.174	1	.000	.049
		T	-3.588	.940	14.569	1	.000	.028
		Drp	3.860	1.032	14.000	1	.000	47.471
		Sgn	2.600	.573	20.603	1	.000	13.465
		Hotf	3.203	1.133	7.991	1	.005	24.613
		Constant	-8.950	3.199	7.827	1	.005	.000
	Step 7[g]	F	5.011	1.344	13.902	1	.000	150.011
		Pn	-3.037	.956	10.095	1	.001	.048
		T	-3.638	1.035	12.353	1	.000	.026
		Drp	4.420	1.193	13.733	1	.000	83.092
		Sgn	2.621	.651	16.193	1	.000	13.747
		Hotf	3.840	1.255	9.359	1	.002	46.526
		S	.754	.319	5.583	1	.018	2.126
		Constant	-13.234	4.220	9.833	1	.002	.000
	Step 8[h]	F	5.572	1.743	10.224	1	.001	262.876
		Pn	-3.607	1.338	7.266	1	.007	.027
		T	-3.891	1.177	10.922	1	.001	.020
		Drp	5.199	1.692	9.448	1	.002	181.140
		Sgn	2.277	.714	10.158	1	.001	9.744
		Hotf	3.991	1.642	5.912	1	.015	54.125
		Pexp	3.906	1.566	6.217	1	.013	49.687
		S	1.356	.552	6.028	1	.014	3.879
		Constant	-21.701	7.606	8.141	1	.004	.000

	Step 9[i]	F	5.235	2.014	6.759	1	.009	187.742
		H	4.805	2.312	4.318	1	.038	122.151
		Pn	-5.933	2.353	6.358	1	.012	.003
		T	-4.195	1.364	9.463	1	.002	.015
		Drp	6.001	2.245	7.149	1	.007	404.015
		Sgn	2.996	1.048	8.173	1	.004	20.004
		Hotf	4.428	1.841	5.783	1	.016	83.765
		Pexp	5.136	2.149	5.710	1	.017	170.102
		S	1.994	.890	5.015	1	.025	7.346
		Constant	-31.441	11.606	7.338	1	.007	.000
	Step 10[j]	F	11.666	6.082	3.679	1	.055	116528.097
		H	13.706	7.774	3.108	1	.078	895883.771
		Pm	-6.928	3.724	3.461	1	.063	.001
		Pn	-17.231	9.685	3.165	1	.075	.000
		T	-6.797	3.407	3.981	1	.046	.001
		Drp	13.537	7.494	3.263	1	.071	757045.523
		Sgn	6.787	3.263	4.327	1	.038	885.874
		Hotf	10.339	5.720	3.267	1	.071	30899.694
		Pexp	12.966	6.728	3.714	1	.054	427513.967
		S	4.987	2.716	3.371	1	.066	146.487
		Constant	-68.864	35.918	3.676	1	.055	.000
	Step 11[k]	F	128.783	2128.408	.004	1	.952	8.510E+55
		H	231.284	4065.613	.003	1	.955	2.788E+100
		Pm	-77.826	1426.050	.003	1	.956	.000
		Pn	-206.582	3623.554	.003	1	.955	.000
		T	-126.361	2189.941	.003	1	.954	.000
		Drp	179.396	2783.184	.004	1	.949	8.145E+77
		Sgn	54.213	1729.175	.001	1	.975	350147102
		Hotf	32.396	3404.358	.000	1	.992	117364804
		WtisPolio	100.875	1921.835	.003	1	.958	6.451E+43
		Pexp	130.469	2978.504	.002	1	.965	4.592E+56
		S	52.314	1054.006	.002	1	.960	524658358
		Constant	-824.832	14617.46	.003	1	.955	.000
	Step 12[k]	F	169.373	4781.602	.001	1	.972	3.614E+73
		H	345.884	5682.155	.004	1	.951	1.643E+150
		Pm	-99.485	2989.325	.001	1	.973	.000
		Pn	-293.530	4614.715	.004	1	.949	.000
		T	-177.398	5930.825	.001	1	.976	.000
		Drp	246.019	4045.029	.004	1	.952	6.992E+106
		Sgn	59.960	760.726	.006	1	.937	109748892
		WtisPolio	151.175	5822.371	.001	1	.979	4.512E+65

		Pexp	191.675	3731.811	.003	1	.959	1.752E+83
		S	72.324	910.401	.006	1	.937	256872321
		Constant	-1125.72	13949.65	.007	1	.936	.000
	Step 13[l]	F	146.796	10478.98	.000	1	.989	5.656E+63
		H	354.631	15736.39	.001	1	.982	1.034E+154
		Pm	-109.246	15425.37	.000	1	.994	.000
		Pinf	-21.860	9711.768	.000	1	.998	.000
		Pn	-292.339	4118.295	.005	1	.943	.000
		T	-196.736	30742.34	.000	1	.995	.000
		Drp	255.311	15675.76	.000	1	.987	7.589E+110
		Sgn	59.843	752.715	.006	1	.937	975828302
		WtisPolio	170.349	30807.56	.000	1	.996	9.582E+73
		Pexp	181.499	15466.21	.000	1	.991	6.669E+78
		S	72.158	896.781	.006	1	.936	217664797
		Constant	-1057.0	31730.77	.001	1	.973	.000

Table 14: Cox & Snell R-Square and Nagel Kreke R-Square of best selected model.

Gender	Cox & Snell R Square	Nagelkerke R Square
Male	.668	.958

Table 14 displays the Cox & Snell R-Square and Nagel Kreke R-Square values for the best-selected model. These values represent the variation in the dependent variable explained by the model, with Nagel Kreke R-Square explaining 96% of the variation in the dependent variable.

Table 15: Categorical Variables Coding of Best Selected Male Model

Gender			Frequency	Parameter coding (1)	(2)	(3)	(4)
Male	From which source you have heard the word Polio?	TV	86	1.000	.000	.000	.000
		Radio	20	.000	1.000	.000	.000
		Newspaper	34	.000	.000	1.000	.000
		Healthworker	20	.000	.000	.000	1.000
		Other	82	.000	.000	.000	.000
	How long does it take to show the Polio signs?	Dont know	82	1.000	.000		
		1 to 2 days	19	.000	1.000		
		3 to 35 days	141	.000	.000		
	Are you forced of not giving Polio dose?	Yes	150	1.000			
		No	92	.000			
	Do you know how Polio tested (through stool or throat)?	Yes	158	1.000			
		No	84	.000			

	Can Polio drops prevent other diseases also?	No	83	1.000			
		Yes	159	.000			
	Can hot food stuff be given just after(within half an hour) administration of Polio drop	Yes	150	1.000			
		No	92	.000			
	Past experience with Polio patients	Yes	89	1.000			
		No	153	.000			
	Do you have proper flesh system?	Yes	52	1.000			
		No	190	.000			

For the best-selected model focusing on males, categorical variables are coded in Table 15. It includes six binary variables, one ordinal variable, and one nominal variable, specifically the source of getting information.
So the above table was summarized in the equation form as:

$$Logit\left(\hat{Y}\right) = -21.701 + 1.356\ S + 3.906\ P\exp + 3.991\ Hotf + 2.277\ Sgn + 5.199\ Drp$$
$$-3.891\ T - 3.607 Pn + 5.572\ F$$

$$\begin{aligned} \boldsymbol{Logit(\hat{Y}) = -21.701 + 1.356\ From\ which\ souce\ you\ heard\ the\ word\ Polio}? \\ \boldsymbol{+3.906\ Past\ exprience\ with\ Polio\ patients} \\ \boldsymbol{+3.991\ Can\ hot\ food\ stuffbe\ given\ just\ after}\ (\boldsymbol{within\ half\ an\ hour}) \\ \boldsymbol{ad\min i\ stration\ of\ Polio\ drop} \\ \boldsymbol{+2.277 How\ long\ does\ it\ take\ to\ show\ the\ Polio\ signs}? \\ \boldsymbol{+5.199\ Can\ Polio\ drops\ prevent\ other\ diseases\ also}? \\ \boldsymbol{-3.891\ Do\ you\ how\ Polio\ tested}(\boldsymbol{through\ stool\ or\ throat})? \\ \boldsymbol{-3.607\ Are\ you\ forced\ of\ not\ giving\ Polio\ dose}? \\ \boldsymbol{+5.572\ Do\ you\ have\ proper\ flesh\ system}? \end{aligned}$$

The thirteen-step procedure selected the best model at stage eight, incorporating eight factors. Among these, two factors (T and Pn) had negative coefficients, while F, Drp, Sgn, Hotf, Pexp, and S had positive coefficients. The table highlights a strong association between these factors and Polio. Larger regression coefficients correspond to higher odds ratios, while negative coefficients approach zero. The highest odds ratio (262.876) was for knowledge about the proper flesh system, indicating those informed about Polio are significantly more likely to support proper flesh system use. Similarly, the odds ratio for duration of showing Polio symptoms is 9.744, suggesting Polio-affected individuals are more knowledgeable about symptom onset periods. Conversely, odds ratios for Pn (forced not to give Polio dose) and T (knowledge about Polio testing) were 0.027 and 0.020 respectively, indicating lower awareness among Polio-affected families regarding these aspects. An odds ratio of 181.140 for whether Polio drops can prevent other diseases shows that Polio-affected families are significantly more aware of this prevention aspect compared to unaffected families. Factors like offering hot food, Polio patient experiences, and information sources also showed higher odds ratios, reflecting greater awareness among affected families about Polio compared to unaffected ones.

3.4 Model Diagnostics for Female Model

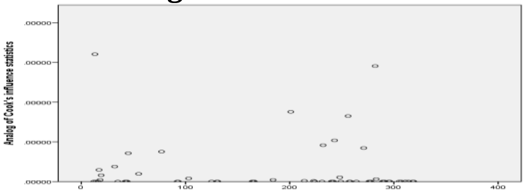

Figure 11: Index plot of Cook's Distance female

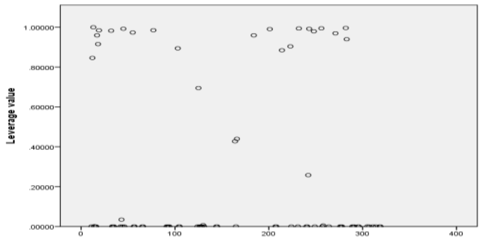

Figure 12: Index plot of leverage values for female

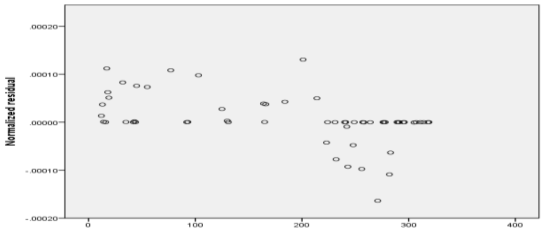

Figure 13: Index plot of standardized residuals for female

Figure 11 to 13 show the diagnostics tests for the female respondents. These figures give the satisfactory results and indicated that the data is free from any type of outliers

4.4 Model Fitting of Female Polio Respondents

Table 16: Model Summary for Female

	Step	Cox & Snell R Square	Nagelkerke R Square
Female	1	.522	.708
	2	.593	.805
	3	.653	.886
	4	.696	.945
	5	.737	1.000

Table 16 summarizes the 5-step forward selection procedure, showing Cox & Snell R-Square and Nagel Kreke R-Square values for each step. These values indicate the variation explained by the model. Since the values increased slowly at steps 4 and 5, the model at stage 3 was chosen as the best using the parsimony approach. The detailed of each model is summarized in the following tables:

Table 17: Summary of the Female Models through Forward Selection Procedure

Female	Step 1[a]	Sgn	2.600	.483	29.005	1	.000	13.461
		Constant	-2.045	.545	14.081	1	.000	.129
	Step 2[b,n]	Sgn	2.661	.617	18.629	1	.000	14.315
		Constant	-7.696	2.238	11.824	1	.001	.000
		F	3.343	1.131	8.742	1	.003	28.297
	Step 3[c,o]	Sgn	3.576	1.037	11.892	1	.001	35.713
		Constant	-17.205	5.342	10.371	1	.001	.000
		F	4.747	1.605	8.750	1	.003	115.251
		Hotf	4.201	1.624	6.696	1	.010	66.782
	Step 4[d,p]	F	5.094	2.180	5.460	1	.019	163.050

		Sgn	4.409	1.783	6.118	1	.013	82.215
		Constant	-19.385	8.022	5.839	1	.016	.000
		Hotf	6.540	3.137	4.346	1	.037	692.490
		Pv	-2.773	1.185	5.478	1	.019	.062
	Step 5	F	62.887	5207.370	.000	1	.990	2049667018828
		Pn	-61.321	4705.824	.000	1	.990	.000
		Sgn	47.158	3587.752	.000	1	.990	3021832780637
		Constant	-146.990	21388.074	.000	1	.995	.000
		Hotf	79.860	11293.096	.000	1	.994	4815990753040
		Pv	-30.699	2578.628	.000	1	.991	.000

Table 18 shows that the best-selected model explains 89% of the variation in the dependent variable according to Nagel Kreke R-Square.

Table 18: Cox & Snell R-Square and Nagel Kreke R-Square of best selected female

Gender	Cox & Snell R Square	Nagelkerke R Square
Female	.657	.891

Table 19 describes the coding of the categorical variables. The best model includes one ordinal variable and two binary variables.

Table 19: Categorical Variables Coding of Best Selected Female Model

Gender			Frequency	Parameter coding (1)	Parameter coding (2)
Female	How long does it take to show the Polio signs?	Dont know	29	1.000	.000
		1 to 2 days	10	.000	1.000
		3 to 35 days	41	.000	.000
	Can hot food stuff be given just after(within half an hour) administration of Polio drop	Yes	47	1.000	
		No	33	.000	
	Do you have proper flesh system?	Yes	23	1.000	
		No	57	.000	

The forward selection procedure identified the best model at stage 3, which includes three significant factors: duration of polio symptoms, knowledge about the proper flesh system, and provision of hot food within an hour after the polio drop.

Table 20 : Summary of Best Selected Model for female

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 8	Sgn	3.576	1.037	11.892	1	.001	35.713
	Constant	-17.205	5.342	10.371	1	.001	.000
	F	4.747	1.605	8.750	1	.003	115.251
	Hotf	4.201	1.624	6.696	1	.010	66.782

So the above table was summarized in the equation form as:

$$Logit(\widehat{Y}) = -17.205 + 4.201\ Hotf + 3.576\ Sgn + 4.747\ F$$

$$Logit(\widehat{Y}) = -17.205$$
$$+3.991\ Can\ hot\ food\ stuff be\ given\ just\ after\ (within\ half\ an\ hour)\ ad\min i\ stration\ of\ Polio\ drop$$
$$+2.277 How\ long\ does\ it\ take\ to\ show\ the\ Polio\ signs?$$
$$+5.199\ Can\ Polio\ drops\ prevent\ other\ diseases\ also?$$
$$+5.572\ Do\ you\ have\ proper\ flesh\ system?$$

The best model for female respondents (Table 4.2202) shows three significant factors with positive regression coefficients. The odds ratio for knowing the duration of polio symptoms is 35.713, indicating higher awareness among females from polio-affected families. The odds ratio for using a proper flesh system is 115.251, and for giving hot food within half an hour after polio drops, it is 66.782, both showing significantly greater knowledge among these females.

5. Conclusion

Polio remains a persistent issue in Pakistan, largely due to the substantial challenges involved in ensuring that all children receive adequate vaccine doses. In 2010, Pakistan reported 144 polio cases, a number that exceeded the combined cases of Nigeria, India, and Afghanistan. This alarming statistic underscores the critical nature of polio as a public health concern within the country. Predominantly affecting children under the age of five, polio initially manifests through a series of symptoms such as fever, fatigue, headache, vomiting, neck stiffness, and limb pain. With no available cure, prevention through immunization remains the most effective strategy to combat the spread of this debilitating disease.

Pakistan, being a developing nation where a significant portion of the population lives below the poverty line, faces myriad barriers that impede progress in polio eradication. Among these barriers are limited education and inadequate access to healthcare services, both of which significantly hinder efforts to reach all segments of the population with necessary vaccines. The present study aimed to model polio incidence among patients attending hospitals in Peshawar, one of the regions most affected by the disease. Utilizing a detailed questionnaire, data was collected from 332 respondents, comprising 222 polio patients and 100 individuals from polio-free families. This dataset was strategically analyzed across three stages: overall, male-specific, and female-specific, to gain a nuanced understanding of polio's impact across different demographics.

The analysis identified several significant factors influencing polio incidence, including the duration of polio symptoms, the availability of adequate sanitation facilities, and the practice of immediate consumption of hot food following polio drops. These factors emerged consistently across various models, highlighting their critical role in understanding polio dynamics. Notably, a comparison between overall and male-specific models revealed shared factors such as hot food consumption post-polio drops, the duration of symptoms, the role of polio drops in preventing other diseases, testing methods, resistance to receiving the polio vaccine, and access to sanitation facilities. However, the presence of the virus causing polio was significant in the overall model but not in the male-specific model. Conversely, sources of information about polio and prior experience with polio patients were significant factors in the male-specific model, indicating varying influences across different population segments.

To explore these relationships, the study employed descriptive statistics for initial data exploration, Chi-Square tests to assess associations between categorical variables, and odds ratios to quantify the strength of these relationships. The findings revealed a significant gender association with polio incidence (Odds Ratio = 1.586), indicating that males are at a higher risk compared to females. Additionally, households lacking proper sanitation facilities exhibited higher odds of polio incidence (Odds Ratio = 0.081), as did families with an income below Rs. 20,000. Further insights into the data revealed that only 6.3% of patients consistently practiced handwashing with soap, while 69.9% of respondents perceived polio as a transmissible disease. The majority (60.87%) reported experiencing external pressure against vaccinating children, whereas 39.13% asserted autonomy in their decision-making process regarding vaccinations.

In addition to the above analysis, cluster analysis was utilized to gain deeper insights into the data. This analytical approach aimed to identify patterns and groupings within the dataset that might not be immediately apparent through traditional statistical methods. By employing three distinct clustering techniques—the gap statistic, silhouette statistic, and elbow method—the analysis revealed that the dataset is optimally divided into three clusters. Each of these methods corroborated the conclusion that three clusters provide the most robust insights into the data, reinforcing the identified factors and highlighting potential intervention points. The elbow method, for instance, showed a distinct "bend" in the curve at three clusters, indicating an optimal partitioning of data with minimal within-cluster variation. Similarly, the gap statistic and silhouette statistic both pointed to three clusters as the most effective solution, underscoring the consistency and reliability of these findings.

This systematic clustering approach not only provides a deeper understanding of the underlying data structure but also reinforces the validity of the identified factors, such as sanitation access, socioeconomic status, and resistance to vaccination. These insights offer valuable guidance for public health interventions aimed at targeting specific clusters within the population that are most at risk. By doing so, health authorities can design more effective strategies that address the unique challenges faced by each cluster, thereby optimizing resource allocation and enhancing the overall impact of polio eradication efforts.

The findings of this study underscore the persistent threat of polio in Pakistan and highlight the critical need for targeted strategies to combat the disease. By understanding the nuanced interplay of factors contributing to polio incidence and leveraging advanced analytical techniques like cluster analysis, stakeholders can develop more comprehensive and effective interventions. These efforts are essential to overcoming barriers to immunization, improving public health outcomes, and ultimately achieving the goal of a polio-free Pakistan. The study's insights not only contribute to the existing body of knowledge on polio eradication but also provide a foundation for future research and policy-making aimed at addressing one of the most pressing public health challenges facing the nation today.

Acknowledgements

Not applicable.

Author contributions

Study conception and design: Qamruz Zaman and Sofia Gul

Data collection: Mansoor Ahmad and Soofia Iftikhar.

Interpretation of results: Syed Habib Shah, Mir Ullah and Aizaz Shah. Draft manuscript

Preparation: Qamruz Zaman and Neelam.

Supervision: Qamruz Zaman and Soofia Iftikhar. All authors reviewed the results and approved the final version of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical approval

Our study did not require an ethical board approval because there was no human interaction in this study.

Informed consent
Not applicable.

Trial registration
Not applicable.
ORCID iD
Qamruz Zaman: https://orcid.org/0009-0007-5971-2685

Supplemental material
Primary Data was collected for the article.

References
1. Andrade et.al. (2018). Polio in Pakistan: Political, Sociological, and Epidemiological Factors. *Open Access Review. , 10.7759/cureus.3502,*, 10(10).
2. Aylward et.al, . (2011). The global polio eradication initiative lessons learned and prospects for success. *vaccine.*, 29, 80-85.
3. Bland, et.al. (2000). The odds ratio. *Bmj* , 320(7247), 1468.
4. Franke et.al. (2012). The chi-square test: Often used and more often misinterpreted. . *American journal of evaluation,* , 33(3), 448-458.
5. Global Polio Eradication. (2018). *initiative, WHO polio weekly global surveillance updata (2013-2018)[Internet]*. Geneva: GPEI.
6. History of Polio. (2019). *[https://polioeradication.org/polio-today/history-of-polio/]*. polioeradication.org.
7. Kleinbaum, et.al. (2002). *Logistic regression (p. 536).* . New York: : Springer-Verlag.
8. Kraemer, et.al. (2004). Reconsidering the odds ratio as a measure of 2× 2 association in a population. *Statistics in Medicine*, 23(2), 257-270.
9. LaValley. (2008). Logistic regression. . *Circulation*, 117(18), 2395-2399.
10. Mehndiratta et.al. (2014). Historical Facts, Epidemiology, and Current Challenges in Eradication. *The Neurohospitalist.*, 4(4), 223-229.
11. Mehndiratta et.al. (2014). Historical Facts, Epidemiology, and Current Challenges in Eradication. *The Neurohospitalist.* , 4(4), 223-229.
12. National Institute of Health, Pakistan. (2023). *Polio eradication program. Retrieved from https://www.nih.org.pk/polio-eradication-program.*
13. Rana,et.al. (2015). Chi-square test and its application in hypothesis testing. *Journal of Primary Care Specialties,*, 1(1), 69-71.
14. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. . *Journal of computational and applied mathematics*, 20, 53-65.
15. Tebbensa et.al. (2010). Economic analysis of the global polio eradication initiative, . *vaccine.*, 29, 334-343.
16. Tibshirani, R. W. (2001). Estimating the number of clusters in a data set via the gap statistic. . *Journal of the Royal Statistical Society*, 63(2), 411-423.
17. World Health Organization, & Global Polio Eradication Initiative. (2019). *Polio endgame strategy 2019-2023: eradication, integration, certification and containment (No. WHO/Polio/19.04).* World Health Organization.
18. Zimmermann et.al. (2019). Projection of Costs of Polio Eradication Compared to Permanent Control,. *THe Journal of Infectious Diseases*, 1-5.

APPENDIX – I

APPENDIX: Distribution of data among the affected districts

Districts	Case	Control	Total
Peshawar	15	8	23
Charsadda	16	9	25
Nowshera	10	8	18
Mardan	13	8	21
Swabi	10	8	18
Buner	12	6	18
Torghar	14	6	20
Malakand	12	6	18
Swat	21	7	28
Kohat	20	6	26
Bannu	20	10	30
LakiMarwat	20	2	22
Mohmand	17	4	21
Khyber	6	3	9
Bajaur	7	4	11
Others	9	5	14
Total	222	100	322

Appendix table I provides a breakdown of data concerning the most affected districts in Khyber Pakhtunkhwa. It presents the number of patients categorized into both the case and control groups from these districts, offering insights into regional impacts and distributions within the study.