**DATA ANALYSIS FOR PATIENTS SEEKING TREATMENT AFTER TESTING POSITIVE TO COVID-19**

**Objective**

To find any relationship between various independent variables with the outcome of question Q1a. The question stated for Q1a is as follows:  
 ‘If you tested positive for COVID-19, would you consider seeking treatment from   
 your doctor?’

The study only concerns patients from Australia.

**Data**

The supplied data file contains 1514 records in total for both Australian and New Zealand patients and 88 columns.

**Data Cleaning**

As this study is only for Australian patients, all rows pertaining to New Zealand patients were removed. This reduced the number of records down to 1009 in total.

A large proportion of columns were removed for the following reasons:

1. The columns were specific to New Zealand patients
2. All values were the same for each record
3. Most patients did not answer the question meaning that estimating missing values would be pointless
4. The values in the column were a combination of values in two separate columns
5. The data had no value for the study
6. Grouping data with a high number of varied responses was already catered for by another column

Some columns were merged into the one column where a separate column only catered for responses not covered in the list of the first column. Options from the list that were either “Other, please specify” or “Something else, please specify:” were replaced by what was typed in the separate column.

**List Of Columns Kept for the Study**

|  |  |
| --- | --- |
| **COLUMN NAME FOR STUDY** | **ORIGINAL NAME** |
| AE2 | ‘AE2: Do you agree…’ |
| Age | ‘S3: Please enter your age below’ |
| Age\_Bracket | ‘hAge: Age’ |
| Asthma\_moderate | ‘Q1r1: Asthma (moderate)…’ |
| Asthma\_severe | ‘Q1r2: Asthma (severe…’ |
| Autoimmune | ‘Q1r15: Autoimmune disease…’ |
| Cancer | ‘Q1r16: Cancer…’ |
| Cardiovascular | ‘Q1r3: Cardiovascular disease…’ |
| 'Chronic\_Infections' | ‘Q1r17: Chronic infections…’ |
| 'Congenital\_Acquired\_Heart\_Disease' | ‘Q1r18: Congenital/acquired heart…’ |
| COPD\_moderate | ‘Q1r4: COPD (moderate)…’ |
| COPD\_severe | ‘Q1r5: COPD (severe…’ |
| Dementia\_Alzheimer | ‘Q1r6: Dementia/Alzheimer's…’ |
| Diabetes | ‘Q1r7: Diabetes…’ |
| Disabilities | ‘Q1r8: Disabilities…’ |
| Ethnicity\_Background | Merging of ‘S4:…’ & ‘S4r12oe:’ |
| GCCSA | ‘hGCCSA: HIDDEN QUESTION’ |
| Gender | ‘S1: To start with, please select…’ |
| HIGH\_RISKr1 | ‘HIGH\_RISKr1: Aged 70+…’ |
| HIGH\_RISKr4 | ‘HIGH\_RISKr4: Aged 18+…’ |
| HIGH\_RISKr5 | ‘HIGH\_RISKr5: Aged 50-69…’ |
| HIGH\_RISKr6 | ‘HIGH\_RISKr6: Aged 18+ and CODE 23…’ |
| Hypertension | ‘Q1r9: Hypertension…’ |
| Interview\_Time | ‘qtime: Total Interview Time’ |
| Liver | ‘Q1r10: Liver dysfunction…’ |
| Neurodevelopmental | ‘Q1r11: Neurodevelopmental disorder…’ |
| Obesity | ‘Q1r12: Obesity…’ |
| Other\_Chronic\_Respiratory | ‘Q1r13: Other chronic respiratory disease…’ |
| Other | Merging of ‘Q1r22:…’ & ‘Q1r22oe:…’ |
| Q1a | ‘Q1a: If you tested positive for COVID-19…’ |
| Q2\_New | ‘Q2\_New: When was your most recent…’ |
| Q3 | ‘Q3: Did you contact a Healthcare…’ |
| Q6 | ‘Q6: Have you seen or heard any…’ |
| Renal\_Kidney | ‘Q1r14: Renal/kidney dysfunction…’ |
| Sickle | ‘Q1r20: Sickle cell disease…’ |
| Solid\_Organ | ‘Q1r21: Solid organ…’ |
| State | ‘hState: HIDDEN QUESTION hState’ |

NB: Not all columns were used but were still kept as they had potential or could   
 possibly aggregated into a new column.

**Analysis**

With Q1a as the outcome (y), both logistic regression and the chi square test were used to find any relationship between Q1a and the other variables in the dataset.

Q1a has 3 separate as follows:

Qty

Yes 607

No 285

Not sure 117

For the purposes of interpretability, records containing ‘Not sure’ for Q1a were removed to make the outcome binary. This reduced the total number of records to 892.

Various logistic regression models were tested with either categorical or continuous variables or both. While some variables appeared to be significant, they ultimately explained very little of the relationship between the independent variables and the outcome variable.

More success was found when conducting the chi square test. For this test only binary variables were used. A total of 27 variables were tested against Q1a and for 6 of the variables a conclusion could be made regarding the type of patient and their likelihood of seeking treatment from a doctor if they tested positive for COVID-19. Odds ratio was used to determine the scale in which a patient would seek treatment compared to those that answered ‘No’.

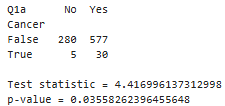
**Results**

People with hypertension are 1.76 times more likely to seek treatment from a doctor if testing positive to COVID-19. The Confidence interval is (1.35, 2.17).

A number and a mathematical equation

AI-generated content may be incorrect.

People with cancer are 2.91 times more likely to seek treatment from a doctor if testing positive to COVID-19. The Confidence interval is (1.95, 3.87).



People with HIGH\_RISKr1 are 2.76 times more likely to seek treatment from a doctor if testing positive to COVID-19. The Confidence interval is (2.25, 3.27).

A number and numbers on a white background

AI-generated content may be incorrect.

People with HIGH\_RISKr5 are 2.18 times more likely to seek treatment from a doctor if testing positive to COVID-19. The Confidence interval is (1.53, 2.82).

A number and a mathematical equation

AI-generated content may be incorrect.

People with HIGH\_RISKr6 are 3.95 times more likely to seek treatment from a doctor if testing positive to COVID-19. The Confidence interval is (3.01, 4.89).

A number and a mathematical equation

AI-generated content may be incorrect.

NB: Although it appeared there was some significance for people with   
 renal/kidney dysfunction, one of the assumptions of the chi square test   
 was not met as one of the cells (equal to < 80% of the cells) had a count   
 less than 5.

A number and numbers on a white background

AI-generated content may be incorrect.

Cancer, people previously admitted with Covid-19 and HIGH\_RISKr6   
 contain one cell with just a count of 5 which means they are on the border   
 of meeting the requirements of the chi square test.