**Results**

Table. 1 shows the number of participants distribution of all variables, including demographic and all questionnaire variables for 12 states (table. 1 is listed in the appendix). Among the 12 states, there were 3797 participants included in the study, 74.27 % of participants did not have STI testing, 20.23 % had STI testing, and 5.5 % were not sure. Michigan (MI) has the highest percentage of participants who had STI testing. 53.28 % of participants from 12 states were female, and 45.72 % were male. The mean BMI of participants from 12 states was 24.20. 63 % of participants from 12 states were White, 14.35% were Black or African American, 14.49 % were Hispanic/Latino, and 8.16 % were all other races. Table. 2 shows the number of participants distribution of all variables, including demographic and all questionnaire variables (table. 2 listed in the appendix). Among the 15 local school districts, there were 2723 participants included in the study; 67.68% of participants did not have STI testing, 26.18 % had STI testing, and 6.12 % were not sure. Philadelphia, PA (PH) has the highest percentage of participants who had STI testing. 54.39 % of participants from 15 local school districts were female, and 45.61 % were male. The mean BMI of participants from 15 local school districts was 23.64. 29.97 % of participants from 15 local school districts were White, 21.26 % were Black or African American, 37.94 % were Hispanic/Latino, and 10.83 % were all other races.

After training the State and local school district datasets separately with multiple ML algorithms, Table. 3 and Table. 4 show the accuracy of all the trained ML algorithms for states and local school districts. Based on the table. 3, the ridge algorithm has the highest mean value of accuracy for the states dataset, which is also shown in the Figure. 1 (a). In addition, random forest, lasso, and ridge algorithms present similar accuracies based on the state datasets. For ML methods trained on local school district data, table. 4 shows that the lasso algorithm has the highest mean accuracy value for the local school district dataset, as shown in Figure. 2 (b). In addition, lasso and ridge algorithms present similar accuracies based on the local school district datasets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Qu | Median | Mean | 3rd Qu | Max |
| multinominal | 0.4357 | 0.4504 | 0.4671 | 0.4690 | 0.4872 | 0.5132 |
| random forest | 0.6974 | 0.7164 | 0.7280 | 0.7219 | 0.7289 | 0.7375 |
| SVM | 0.4042 | 0.4471 | 0.4586 | 0.4595 | 0.4789 | 0.5079 |
| lasso | 0.7297 | 0.7370 | 0.7447 | 0.7437 | 0.7495 | 0.7579 |
| ridge | 0.7297 | 0.7380 | 0.7464 | 0.7447 | 0.7495 | 0.7605 |
| elastic net | 0.4331 | 0.4517 | 0.4658 | 0.4674 | 0.4826 | 0.5132 |
| classification tree | 0.4514 | 0.4721 | 0.4816 | 0.4800 | 0.4852 | 0.5039 |

*Table. 3, resampled accuracy for ML algorithms trained by State data*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Qu | Median | Mean | 3rd Qu | Max |
| multinominal | 0.4485 | 0.4639 | 0.5119 | 0.4985 | 0.5233 | 0.5401 |
| random forest | 0.6496 | 0.6667 | 0.6782 | 0.6768 | 0.6866 | 0.7044 |
| SVM | 0.4359 | 0.4656 | 0.4890 | 0.4864 | 0.5120 | 0.5328 |
| lasso | 0.6788 | 0.7014 | 0.7099 | 0.7079 | 0.7163 | 0.7436 |
| ridge | 0.6765 | 0.6932 | 0.7062 | 0.7054 | 0.7153 | 0.7473 |
| elastic net | 0.5803 | 0.6432 | 0.6545 | 0.6604 | 0.6933 | 0.7143 |
| classification tree | 0.1912 | 0.6273 | 0.6319 | 0.5880 | 0.6447 | 0.6520 |

*Table. 4, resampled accuracy for ML algorithms trained by local school district data.*

图表, 散点图

描述已自动生成 图表, 散点图

描述已自动生成

*Figure. 1 (a and b), resampled accuracy for ML algorithms trained by State data (figure 1 (a) on the left) and local school district data (figure 1 (b) on the right).*

In analyzing state and local school district data, we compared the variable importance using the best-performed ML algorithm selected from Table 3 and Table 4. Table 5 listed the variable for each State with variable importance > 50. Each State shows a different pattern of variable importance. However, variables q17 about physical fights, q21 about dating forcing sex, q41 about alcohol drinking, q52 about heroin use, q53 methamphetamine use, and q89 school grade were observed in most of the States. Datasets of Iowa (IA), Nebraska (NE), and South Carolina (SC) have few than 8 observations for “Not sure” of outcome variable – Q85, which can lead to some bias in our model. Table 6 listed the variable for each State with variable importance > 10. Each local school district demonstrates a distinct pattern of variable importance. However, variables q17 about physical fights, q21 about dating forcing sex, q41 about alcohol drinking, q47 about marijuana, q52 about heroin use, q53 about methamphetamine use, and q89 school grade were observed in most of the local school districts. Datasets of Broward County, FL (FT), Genesee Consortium, MI (GE), and Los Angeles, CA (LO) have few than 8 observations for “Not sure” of outcome variable – Q85, which can lead to some bias in our model. In addition, datasets of Eaton Consortium, MI (EA), Newark, NJ (NW), and Shelby County, TN (ST) did not converge. Datasets of Broward County, FL (FT), Eaton Consortium, MI (EA), Fort Worth, TX (FW), Hillsborough County, FL (HL), Los Angeles, CA (LO), and Philadelphia, PA (PH) did not show any variables that have importance > 10.

To address the statistical issues that we encountered in applying the lasso and ridge regression algorithms to each State and local school district, we used the random forest method (3rd in accuracy, see Figure. 1 (a) and (b)) to analyze the variable importance for each State and local school district. Table. 7 show the top 3 important variables for each local school district. For the local school district data, BMI is the most important variable for all local school districts except Shelby County (S). Variables q41, q47, and q89 were about alcohol drinking, marijuana use, and school grade, which were the variables that appeared most often in the Table. 7. Based on Table. 8, the top 3 important variables for the State data include BMI, q41, q47, and q89, which were about alcohol drinking, marijuana use, and school grade. BMI is the most important variable associated with STI testing other than HIV for all States.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table. 7, Top 3 important variables for each local school district using the Random Forest method. | | | | | | |
|  | **1st variable** | **variable importance** | **2nd variable** | **variable importance** | **3rd variable** | **variable importance** |
| Broward County, FL (FT) | bmi | 11.96 | q47 | 6.87 | q41 | 6.24 |
| Chicago, IL (CH) | bmi | 12.79 | q47 | 6.56 | q41 | 5.65 |
| Eaton Consortium, MI (EA) | bmi | 8 | q89 | 4.02 | race4 | 3.77 |
| Fort Worth, TX (FW) | bmi | 13.9 | q47 | 7.99 | q89 | 7.79 |
| Genesee Consortium, MI (GE) | bmi | 9,79 | q89 | 5.31 | age | 4.68 |
| Hillsborough County, FL (HL) | bmi | 10.93 | q47 | 9.95 | q17 | 6.99 |
| Los Angeles, CA (LO) | bmi | 7.72 | age | 3.02 | q41 | 2.8 |
| Newark, NJ (NW) | bmi | 10.14 | q47 | 7.64 | q89 | 5.24 |
| Orange County, FL (OL) | bmi | 7.52 | q89 | 4.05 | q47 | 3.61 |
| Palm Beach County, FL (PB) | bmi | 17.2 | q47 | 10.46 | q41 | 9.45 |
| Pasco County, FL (PS) | bmi | 12.13 | q47 | 6.41 | q17 | 5.58 |
| Philadelphia, PA (PH) | bmi | 11.15 | q89 | 8.22 | race4 | 7.61 |
| Portland, OR (PO) | bmi | 15.13 | age | 7.98 | q47 | 7.95 |
| San Francisco, CA (SF) | bmi | 15.11 | q89 | 9.05 | q47 | 7.83 |
| Shelby County, TN (ST) | q47 | 10.46 | bmi | 9.85 | q17 | 6.55 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table. 8, Top 3 important variables for each State using Random Forest method. | | | | | |  |
|  | **1st variable** | **variable importance** | **2nd variable** | **variable importance** | **3rd variable** | **variable importance** |
| Alabama (AL) | bmi | 16.94 | q89 | 9.14 | age | 8.66 |
| Arkansas (AR) | bmi | 16.24 | q47 | 8.06 | q41 | 7.34 |
| Iowa (IA) | bmi | 12.16 | q89 | 6.24 | q41 | 5.98 |
| Illionois (IL) | bmi | 26.05 | q47 | 12.65 | q89 | 12.06 |
| Kentucky (KY) | bmi | 19.83 | q89 | 10.32 | age | 8.68 |
| Michigan (MI) | bmi | 40.92 | q89 | 19.93 | q47 | 18.86 |
| Mississippi (MS) | bmi | 12.27 | q47 | 9.63 | q41 | 7.06 |
| Nebraska (NE) | bmi | 9.53 | q89 | 4.77 | q41 | 4.42 |
| Oklahoma (OK) | bmi | 19.05 | q41 | 10.67 | q89 | 9.43 |
| Pennsylvania (PA) | bmi | 24.66 | q89 | 13.9 | q41 | 13.29 |
| South Carolina (SC) | bmi | 6.8 | q47 | 4.91 | q89 | 4.41 |
| West Virginia (WV) | bmi | 18.95 | q89 | 11.28 | age | 8.48 |

**Appendix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table. 1 Population characteristics of study population for States (n = 3,797)** | | | | | |
| **Participants’ characteristics** | **STI testing - Yes** | **STI testing - No** | **STI testing - Not sure** | **No. of participants** | **%** |
| **n (%)** | **n (%)** | **n (%)** |
| **Participants** | 768 (20.23) | 2820 (74.27) | 209 (5.5) | **3797** |  |
| **(Total = 3797)** |
| **sitename** (n=3797 participants) | |  |  |  |  |
| Alabama (AL) | 57 | 234 | 17 | 308 | 8.11% |
| Arkansas (AR) | 52 | 234 | 17 | 303 | 7.98% |
| Illinois (IL) | 94 | 286 | 18 | 398 | 10.48% |
| Iowa (IA) | 41 | 187 | 6 | 234 | 6.16% |
| Kentucky (KY) | 60 | 231 | 22 | 313 | 8.24% |
| Michigan (MI) | 152 | 395 | 34 | 581 | 15.30% |
| Mississippi (MS) | 42 | 207 | 14 | 263 | 6.93% |
| Nebraska (NE) | 30 | 163 | 8 | 201 | 5.29% |
| Oklahoma (OK) | 65 | 263 | 19 | 347 | 9.14% |
| Pennsylvania (PA) | 83 | 326 | 31 | 440 | 11.59% |
| South Carolina (SC) | 25 | 95 | 5 | 125 | 3.29% |
| West Virginia (WV) | 67 | 199 | 18 | 284 | 7.48% |
| **Sex** (n=3,797 participants) |  |  |  |  |  |
| Female | 502 (24.81) | 1431 (70.74) | 90 (4.45) | 2023 | 53.28% |
| Male | 266 (14.99) | 1389 (78.30) | 119 (6.71) | 1774 | 46.72% |
| **Age** (n=3,797 participants) | | | | | |
| 13 years old | NA | 5 | NA | 5 | 0.13% |
| 14 years old | 22 | 102 | 6 | 130 | 3.42% |
| 15 years old | 102 | 466 | 46 | 614 | 16.17% |
| 16 years old | 174 | 777 | 60 | 1011 | 26.63% |
| 17 years old | 289 | 898 | 64 | 1251 | 32.95% |
| ≥ 18 years | 181 | 572 | 33 | 786 | 20.70% |
| **Grade** (n=3797 participants) | | | | | |
| 9th grade | 87 (17.40) | 374 (74.80) | 39 (7.80) | 500 | 13.17% |
| 10th grade | 144 (17.04) | 656 (77.63) | 45 (5.33) | 845 | 22.25% |
| 11th grade | 253 (20.74) | 887 (72.70) | 80 (6.56) | 1220 | 32.13% |
| 12th grade | 284 (23.05) | 903 (73.30) | 45 (3.65) | 1232 | 32.45% |
| **Race** (n=3797 participants) |  |  |  |  |  |
| White | 433 (18.1) | 1827 (76.38) | 132 (5.52) | 2392 | 63.00% |
| Black or African American | 147 (26.97) | 366 (67.16) | 32 (5.87) | 545 | 14.35% |
| Hispanic/Latino | 119 (21.64) | 399 (72.55) | 32 (5.82) | 550 | 14.49% |
| All Other Races | 69 (22.26) | 228 (73.55) | 13 (4.19) | 310 | 8.16% |
| **Body mass index, BMI** (n=3797 participants)**,** | | | | | |
| Min 13.25 | 1st Quartile 20.62 | Median 22.89 | Mean 24.20 | 3rd Quartile 26.51 | Max 54.40 |
| **qnothhpl** (n=3797 participants) | |  |  |  |  |
| Yes | 384 (28.28) | 912 (67.16) | 62 (4.56) | 1358 | 35.77% |
| No | 384 (15.74) | 1908 (78.23) | 147 (6.03) | 2439 | 64.23% |
| **Q17** (n=3797patients) | |  |  |  |  |
| 0 times | 505 | 2033 | 143 | 2681 | 70.61% |
| 1 times | 106 | 371 | 25 | 502 | 13.22% |
| 2 or 3 times | 96 | 272 | 23 | 391 | 10.30% |
| 4 or 5 times | 23 | 65 | 8 | 96 | 2.53% |
| 6 or 7 times | 10 | 25 | 3 | 38 | 1.00% |
| 8 or 9 times | 6 | 14 | 1 | 21 | 0.55% |
| 10 or 11 times | 4 | 4 | NA | 8 | 0.21% |
| 12 or more times | 18 | 36 | 6 | 60 | 1.58% |
| **Q21** (n=3797 participants) | |  |  |  |  |
| No date or go out with anyonw during the past 12 months | 63 | 222 | 18 | 303 | 7.98% |
| 0 times | 641 | 2412 | 178 | 3231 | 85.09% |
| 1 times | 27 | 81 | 4 | 112 | 2.95% |
| 2 or 3 times | 22 | 61 | 7 | 90 | 2.37% |
| 4 or 5 times | 4 | 25 | NA | 29 | 0.76% |
| 6 or more times | 11 | 19 | 2 | 32 | 0.84% |
| **Participant’s characteristics** | **Pneumonia** | **Pneumonia** | **Non- pneumonia** | **No. of patients** | **%** |
| **n (%)** | **n (%)** | **n (%)** |
| **Q23** (n=3797patients) | |  |  |  |  |
| Yes | 205 | 580 | 56 | 841 | 22.15% |
| No | 563 | 2240 | 153 | 2956 | 77.85% |
| **Q26** (n=3797 participants) | |  |  |  |  |
| Yes | 242 | 659 | 45 | 946 | 24.91% |
| No | 526 | 2161 | 164 | 2851 | 75.09% |
| **Q30** (n=3797patients) | |  |  |  |  |
| Yes | 397 | 1127 | 106 | 1630 | 42.93% |
| No | 371 | 1693 | 103 | 2167 | 57.07% |
| **Q41** (n=3797 participants) | |  |  |  |  |
| 0 days | 387 | 1532 | 110 | 2029 | 53.44% |
| 1 or 2 days | 169 | 654 | 50 | 873 | 22.99% |
| 3 to 5 days | 106 | 311 | 26 | 443 | 11.67% |
| 6 to 9 days | 54 | 185 | 8 | 247 | 6.51% |
| 10 to 19 days | 34 | 103 | 11 | 148 | 3.90% |
| 20 to 29 days | 7 | 21 | 2 | 30 | 0.79% |
| All 30 days | 11 | 14 | 2 | 27 | 0.71% |
| **Q47** (n=3797patients) | |  |  |  |  |
| 0 times | 419 | 1912 | 138 | 2469 | 65.03% |
| 1 or 2 times | 96 | 310 | 20 | 426 | 11.22% |
| 3 or 9 times | 76 | 221 | 15 | 312 | 8.22% |
| 10 to 19 times | 49 | 131 | 13 | 193 | 5.08% |
| 20 to 39 times | 44 | 93 | 8 | 145 | 3.82% |
| 40 or more times | 84 | 153 | 15 | 252 | 6.64% |
| **Q49** (n=3797 participants) | |  |  |  |  |
| 0 times | 559 | 2316 | 157 | 3032 | 79.85% |
| 1 or 2 times | 75 | 235 | 13 | 323 | 8.51% |
| 3 to 9 times | 49 | 126 | 20 | 195 | 5.14% |
| 10 to 19 times | 36 | 59 | 2 | 97 | 2.55% |
| 20 to 39 times | 18 | 43 | 8 | 69 | 1.82% |
| 40 or more times | 31 | 41 | 9 | 81 | 2.13% |
| **Q50** (n=3797patients) | |  |  |  |  |
| 0 times | 693 | 2713 | 195 | 3601 | 94.84% |
| 1 or 2 times | 32 | 66 | 3 | 101 | 2.66% |
| 3 or 9 times | 20 | 18 | 6 | 44 | 1.16% |
| 10 to 19 times | 11 | 10 | 3 | 24 | 0.63% |
| 20 to 39 times | 2 | 3 | 1 | 6 | 0.16% |
| 40 or more times | 10 | 10 | 1 | 21 | 0.55% |
| **Q52** (n=3797 participants) | |  |  |  |  |
| 0 times | 750 | 2797 | 206 | 3753 | 98.84% |
| 1 or 2 times | 9 | 11 | 2 | 22 | 0.58% |
| 3 to 9 times | 1 | 2 | NA | 3 | 0.08% |
| 10 to 19 times | 1 | 1 | NA | 2 | 0.05% |
| 20 to 39 times | NA | 2 | NA | 2 | 0.05% |
| 40 or more times | 7 | 7 | 1 | 15 | 0.40% |
| **Q53** (n=3797patients) | |  |  |  |  |
| 0 times | 738 | 2788 | 204 | 3730 | 98.24% |
| 1 or 2 times | 11 | 15 | 1 | 27 | 0.71% |
| 3 or 9 times | 3 | 4 | 3 | 10 | 0.26% |
| 10 to 19 times | 4 | 3 | NA | 7 | 0.18% |
| 20 to 39 times | 2 | 1 | NA | 3 | 0.08% |
| 40 or more times | 10 | 9 | 1 | 20 | 0.53% |
| **Q63** (n=3797 participants) | |  |  |  |  |
| Yes | 321 | 1558 | 120 | 1999 | 52.65% |
| No | 447 | 1262 | 89 | 1798 | 47.35% |
| **Q65** (n=3797patients) | |  |  |  |  |
| Female | 255 | 1369 | 117 | 1741 | 45.85% |
| Males | 418 | 1269 | 74 | 1761 | 46.38% |
| Female and males | 95 | 182 | 18 | 295 | 7.77% |
| **Q84** (n=3797 participants) | |  |  |  |  |
| Yes | 532 | 182 | 25 | 739 | 19.46% |
| No | 172 | 2477 | 27 | 2676 | 70.48% |
| Not sure | 64 | 161 | 157 | 382 | 10.06% |
| **Q87** (n=3797patients) | |  |  |  |  |
| Yes | 233 | 659 | 53 | 945 | 24.89% |
| No | 497 | 2085 | 136 | 2718 | 71.66% |
| Not sure | 38 | 76 | 20 | 134 | 3.53% |
| **Q89** (n=3797 participants) | |  |  |  |  |
| Mostly A's | 245 | 1118 | 63 | 1426 | 37.56% |
| Mostly B's | 274 | 1036 | 85 | 1395 | 36.74% |
| Mostly C's | 162 | 465 | 40 | 667 | 17.57% |
| Mostly D's | 42 | 94 | 11 | 147 | 3.87% |
| Mostly F's | 13 | 35 | 2 | 50 | 1.32% |
| None of these grades | 2 | 9 | NA | 11 | 0.29% |
| Not sure | 30 | 63 | 8 | 101 | 2.66% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table. 2 Population characteristics of study population for local school district(n = 2723)** | | | | | |
| **Participants’ characteristics** | **STI testing - Yes** | **STI testing - No** | **STI testing - Not sure** | **No. of participants** | **%** |
| **n (%)** | **n (%)** | **n (%)** |
| **Participants** | 713(26.18) | 1843(67.68) | 167(6.13) | **2723** |  |
| **(Total = 3797)** |
| **sitename** (n=3797 participants) | |  |  |  |  |
| Broward County, FL (FT) | 48 | 113 | 6 | 167 | 6.13% |
| Chicago, IL (CH) | 48 | 98 | 11 | 157 | 5.77% |
| Eaton Consortium, MI (EA) | 21 | 99 | 9 | 129 | 4.74% |
| Fort Worth, TX (FW) | 45 | 174 | 18 | 237 | 8.70% |
| Genesee Consortium, MI (GE) | 38 | 120 | 8 | 166 | 6.10% |
| Hillsborough County, FL (HL) | 37 | 196 | 14 | 247 | 9.07% |
| Los Angeles, CA (LO) | 26 | 79 | 4 | 109 | 4.00% |
| Newark, NJ (NW) | 40 | 87 | 13 | 140 | 5.14% |
| Orange County, FL (OL) | 22 | 92 | 9 | 123 | 4.52% |
| Palm Beach County, FL (PB) | 58 | 224 | 23 | 305 | 11.20% |
| Pasco County, FL (PS) | 37 | 151 | 9 | 197 | 7.23% |
| Philadelphia, PA (PH) | 102 | 59 | 10 | 171 | 6.28% |
| Portland, OR (PO) | 76 | 128 | 9 | 213 | 7.82% |
| San Francisco, CA (SF) | 73 | 116 | 14 | 203 | 7.46% |
| Shelby County, TN (ST) | 42 | 107 | 10 | 159 | 5.84% |
| **Sex** (n=3797 participants) | | | | | |
| Female | 463 | 940 | 78 | 1481 | 54.39% |
| Male | 250 | 903 | 89 | 1242 | 45.61% |
| **Age** (n=3,797 participants) | | | | | |
| 14 years old | 19 | 81 | 11 | 111 | 4.08% |
| 15 years old | 82 | 313 | 25 | 420 | 15.42% |
| 16 years old | 180 | 496 | 52 | 728 | 26.74% |
| 17 years old | 281 | 612 | 50 | 943 | 34.63% |
| ≥ 18 years | 151 | 341 | 29 | 521 | 19.13% |
| **Grade** (n=3797 participants) | | | | | |
| 9th grade | 51 | 197 | 27 | 275 | 10.10% |
| 10th grade | 129 | 441 | 39 | 609 | 22.37% |
| 11th grade | 238 | 534 | 50 | 822 | 30.19% |
| 12th grade | 295 | 671 | 51 | 1017 | 37.35% |
| **Race** (n=3797 participants) |  |  |  |  |  |
| White | 166 | 606 | 44 | 816 | 29.97% |
| Black or African American | 194 | 353 | 32 | 579 | 21.26% |
| Hispanic/Latino | 271 | 686 | 76 | 1033 | 37.94% |
| All Other Races | 82 | 198 | 15 | 295 | 10.83% |
| **Body mass index, BMI** (n=3797 participants)**,** | | | | | |
| Min 13.58 | 1st Quartile 20.38 | Median 22.49 | Mean 23.64 | 3rd Quartile 25.72 | Max 58.89 |
| **qnothhpl** (n=3797 participants) | |  |  |  |  |
| Yes | 285 | 405 | 42 | 732 | 26.88% |
| No | 428 | 1438 | 125 | 1991 | 73.12% |
| **Q17** (n=3797patients) | |  |  |  |  |
| 0 times | 467 | 1334 | 143 | 1905 | 69.96% |
| 1 times | 86 | 231 | 25 | 344 | 12.63% |
| 2 or 3 times | 92 | 179 | 23 | 296 | 10.87% |
| 4 or 5 times | 29 | 55 | 8 | 89 | 3.27% |
| 6 or 7 times | 10 | 17 | 3 | 28 | 1.03% |
| 8 or 9 times | 4 | 8 | 1 | 13 | 0.48% |
| 10 or 11 times | 7 | 3 | NA | 10 | 0.37% |
| 12 or more times | 18 | 16 | 4 | 38 | 1.40% |
| **Q21** (n=3797 participants) | |  |  |  |  |
| No date or go out with anyonw during the past 12 months | 55 | 171 | 13 | 239 | 8.78% |
| 0 times | 588 | 1559 | 145 | 2292 | 84.17% |
| 1 times | 33 | 54 | 4 | 91 | 3.34% |
| 2 or 3 times | 24 | 2713 | 4 | 55 | 2.02% |
| 4 or 5 times | 4 | 13 | NA | 17 | 0.62% |
| 6 or more times | 9 | 19 | 1 | 29 | 1.07% |
|  |  |  |  |  |  |
| **Table. 2 Population characteristics of study population from states (n = 3,797)** | | | | | |
| **Participant’s characteristics** | **Pneumonia** | **Pneumonia** | **Non- pneumonia** | **No. of patients** | **%** |
| **n (%)** | **n (%)** | **n (%)** |
| **Q23** (n=3797patients) | |  |  |  |  |
| Yes | 122 | 266 | 22 | 410 | 15.06% |
| No | 591 | 1577 | 145 | 2313 | 84.94% |
| **Q26** (n=3797 participants) | |  |  |  |  |
| Yes | 192 | 412 | 43 | 647 | 23.76% |
| No | 521 | 1431 | 124 | 2076 | 76.24% |
| **Q30** (n=3797patients) | |  |  |  |  |
| Yes | 223 | 492 | 59 | 774 | 28.42% |
| No | 490 | 1351 | 108 | 1949 | 71.58% |
| **Q41** (n=3797 participants) | |  |  |  |  |
| 0 days | 375 | 1055 | 101 | 1531 | 56.22% |
| 1 or 2 days | 152 | 405 | 32 | 589 | 21.63% |
| 3 to 5 days | 101 | 222 | 17 | 340 | 12.49% |
| 6 to 9 days | 50 | 94 | 10 | 154 | 5.66% |
| 10 to 19 days | 26 | 51 | 6 | 83 | 3.05% |
| 20 to 29 days | 5 | 10 | NA | 15 | 0.55% |
| All 30 days | 4 | 6 | 1 | 11 | 0.40% |
| **Q47** (n=3797patients) | |  |  |  |  |
| 0 times | 371 | 1109 | 93 | 1573 | 57.77% |
| 1 or 2 times | 108 | 239 | 25 | 372 | 13.66% |
| 3 or 9 times | 75 | 201 | 16 | 292 | 10.72% |
| 10 to 19 times | 51 | 100 | 11 | 162 | 5.95% |
| 20 to 39 times | 49 | 83 | 12 | 144 | 5.29% |
| 40 or more times | 59 | 111 | 10 | 180 | 6.61% |
| **Q49** (n=3797 participants) | |  |  |  |  |
| 0 times | 570 | 1490 | 130 | 2190 | 80.43% |
| 1 or 2 times | 48 | 165 | 17 | 230 | 8.45% |
| 3 to 9 times | 39 | 91 | 6 | 136 | 4.99% |
| 10 to 19 times | 27 | 44 | 6 | 77 | 2.83% |
| 20 to 39 times | 10 | 19 | 1 | 30 | 1.10% |
| 40 or more times | 19 | 34 | 7 | 60 | 2.20% |
| **Q50** (n=3797patients) | |  |  |  |  |
| 0 times | 654 | 1756 | 161 | 2571 | 94.42% |
| 1 or 2 times | 23 | 55 | 3 | 81 | 2.97% |
| 3 or 9 times | 22 | 24 | 2 | 48 | 1.76% |
| 10 to 19 times | 4 | 3 | 1 | 8 | 0.29% |
| 20 to 39 times | 5 | 1 | NA | 6 | 0.22% |
| 40 or more times | 5 | 4 | NA | 9 | 0.33% |
| **Q52** (n=3797 participants) | |  |  |  |  |
| 0 times | 702 | 1826 | 167 | 2695 | 98.97% |
| 1 or 2 times | 2 | 10 | NA | 12 | 0.44% |
| 3 to 9 times | 2 | 1 | NA | 3 | 0.11% |
| 10 to 19 times | 3 | 3 | NA | 6 | 0.22% |
| 20 to 39 times | 1 | NA | NA | 1 | 0.04% |
| 40 or more times | 3 | 3 | NA | 6 | 0.22% |
| **Q53** (n=3797patients) | |  |  |  |  |
| 0 times | 694 | 1824 | 164 | 2682 | 98.49% |
| 1 or 2 times | 6 | 11 | 3 | 20 | 0.73% |
| 3 or 9 times | 4 | 1 | NA | 5 | 0.18% |
| 10 to 19 times | NA | 1 | NA | 1 | 0.04% |
| 20 to 39 times | 6 | 1 | NA | 7 | 0.26% |
| 40 or more times | 3 | 5 | NA | 8 | 0.29% |
| **Q63** (n=3797 participants) | |  |  |  |  |
| Yes | 323 | 1093 | 92 | 1508 | 55.38% |
| No | 390 | 750 | 75 | 1215 | 44.62% |
| **Q65** (n=3797patients) | |  |  |  |  |
| Female | 234 | 909 | 90 | 1233 | 45.28% |
| Males | 387 | 767 | 69 | 1223 | 44.91% |
| Female and males | 92 | 167 | 8 | 267 | 9.81% |
| **Q84** (n=3797 participants) | |  |  |  |  |
| Yes | 512 | 162 | 31 | 705 | 25.89% |
| No | 151 | 1559 | 28 | 1738 | 63.83% |
| Not sure | 50 | 122 | 108 | 280 | 10.28% |
| **Q87** (n=3797patients) | |  |  |  |  |
| Yes | 210 | 437 | 36 | 683 | 25.08% |
| No | 471 | 1343 | 120 | 1934 | 71.02% |
| Not sure | 32 | 63 | 11 | 106 | 3.89% |
| **Q89** (n=3797 participants) | |  |  |  |  |
| Mostly A's | 235 | 593 | 43 | 871 | 31.99% |
| Mostly B's | 259 | 743 | 60 | 1062 | 39.00% |
| Mostly C's | 132 | 344 | 42 | 518 | 19.02% |
| Mostly D's | 35 | 61 | 8 | 104 | 3.82% |
| Mostly F's | 22 | 23 | 5 | 50 | 1.84% |
| None of these grades | 4 | 9 | 1 | 14 | 0.51% |
| Not sure | 26 | 70 | 8 | 104 | 3.82% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table. 5 variable importance table for each State, variable importance > 50** | | | |  |  |  |
|  | **Class 1** | **Importance valuel for class 1** | **Class 2** | **Importance valuel for class 2** | **Class 3** | **Importance valuel for class 3** |
| Alabama (AL) | Q53-2 | 100 | Q53-2 | 88.54 | Q17-6 | 92.06 |
|  | Q89-6 | 65.18 | Q49-6 | 76.45 |  |  |
|  | Q41-7 | 65.18 | Q50-2 | 56.33 |  |  |
|  | Q52-2 | 65.18 | Q87-3 | 54.83 |  |  |
|  | Q89-7 | 54.94 | Q89-6 | 53.76 |  |  |
|  | Q49-4 | 54.94 | Q41-7 | 53.76 |  |  |
|  |  |  | Q52-2 | 53.76 |  |  |
|  |  |  | Q41-6 | 50.05 |  |  |
|  |  |  | Q21-4 | 50.05 |  |  |
| Arkansas (AR) | Q49-4 | 77.58 | Q21-5 | 66.16 | Q17-4 | 100 |
|  | Age-3 | 75.46 | Q47-6 | 64.62 | Q87-3 | 65.01 |
|  | Q89-4 | 73.72 | Q89-6 | 63.5 | Q17-5 | 54.31 |
|  | Q52-2 | 56.12 | Q41-7 | 63.41 | Q21-4 | 53.98 |
|  | Q17-4 | 54.61 | Q89-5 | 62.92 |  |  |
|  | Q89-6 | 53.23 | Q17-5 | 62.73 |  |  |
|  | Q41-7 | 53.12 |  |  |  |  |
|  | Q21-5 | 51.72 |  |  |  |  |
|  | Q50-2 | 50.71 |  |  |  |  |
| Illinois (IL) | Q17-7 | 100 | Q17-7 | 90.57 |  |  |
|  | Q41-6 | 64.66 | Q41-6 | 57.7 |  |  |
|  | Q49-4 | 54.71 | Q41-7 | 56.11 |  |  |
| Iowa (IA) | Q52-2 | 100 | Q52-2 | 96.89 |  |  |
|  | Q17-8 | 100 | Q17-8 | 96.89 |  |  |
|  | Q17-6 | 100 | Q17-6 | 96.89 |  |  |
|  | Q53-2 | 100 | Q53-2 | 96.89 |  |  |
|  | Q41-7 | 70.46 | Q41-7 | 67.31 |  |  |
|  | Q65-4 | 51.7 | Q65-4 |  |  |  |
| Kentucky (KY) | Q53-2 | 64.57 | Q41-6 | 100 | Q41-6 | 58.23 |
|  |  |  | Q53-2 | 55.01 |  |  |
| Michigan (MI) | Q41-7 | 94.99 | Q17-7 | 87.18 | Q89-6 | 100 |
|  | Q89-6 | 75.8 | Q41-7 | 81.42 | Q17-6 | 89.65 |
|  | Q17-7 | 70.84 | Q41-6 | 79.99 | Q49-5 | 53.1 |
|  | Q53-2 | 67.19 | Q49-6 | 69.25 |  |  |
|  | Q17-6 | 67.03 | Q17-5 | 69.04 |  |  |
|  | Q41-6 | 63.84 | Q21-5 | 60.12 |  |  |
|  | Q17-5 | 51.84 | Q21-6 | 57.52 |  |  |
|  | Q52-2 | 51.78 | Q49-5 | 54.49 |  |  |
| Mississippi (MS) | Q21-6 | 86.85 | Q49-5 | 100 | Q41-5 | 77.47 |
|  | Q89-4 | 72.19 | Q21-5 | 94.54 | Q17-7 | 66.41 |
|  | Q47-6 | 69.18 | Q89-4 | 88.2 |  |  |
|  | Q52-2 | 67.35 | Q17-8 | 76.14 |  |  |
|  | Q21-5 | 63.89 | Q47-5 | 70.18 |  |  |
|  | Q41-6 | 63.06 | Q89-6 | 69.33 |  |  |
|  | Q17-4 | 62.48 | Q47-6 | 63.93 |  |  |
|  | Q49-6 | 60.37 | Q87-3 | 60.37 |  |  |
| Nebraska (NE) | Q41-6 | 94.32 | Q17-8 | 100.00 | Q17-8 | 81.22 |
|  | Q47-5 | 78.55 | Q41-6 | 85.21 | Q89-7 | 76.66 |
|  | Q17-5 | 59.37 | Q87-3 | 79.09 | Q7-3 | 61.62 |
|  | Q21-4 | 56.70 | Q47-5 | 66.80 | Q41-7 | 55.38 |
|  |  |  | Q89-7 | 55.47 |  |  |
|  |  |  | Q17-5 | 52.82 |  |  |
| Oklahoma (OK) | Q17-6 | 100.00 | Q17-6 | 81.70 | Q89-4 | 50.82 |
|  | Q17-7 | 53.65 | Q17-8 | 64.25 |  |  |
|  |  |  | Q89-4 | 58.23 |  |  |
|  |  |  | Q17-5 | 53.24 |  |  |
| Pennsylvania (PA) | q17-7 | 100.00 | q17-7 | 81.69 | q21-6 | 60.65 |
|  | q21-6 | 58.17 | q17-6 | 72.77 | q50-2 | 50.22 |
|  | q17-6 | 55.77 | q89-5 | 70.93 |  |  |
|  | q89-5 | 53.65 | q49-5 | 61.95 |  |  |
|  |  |  | q50-2 | 55.79 |  |  |
|  |  |  | q17-5 | 55.61 |  |  |
| South Carolina (SC) | q17-6 | 100.00 | q17-6 | 94.06 |  |  |
|  | q52-2 | 59.37 | q52-2 | 53.33 |  |  |
|  | q17-8 | 51.76 |  |  |  |  |
|  | q50-2 | 51.76 |  |  |  |  |
| West Virginia (WV) | q89-7 | 100.00 | q89-7 | 84.34 | q49-5 | 65.77 |
|  | q89-6 | 93.99 | q89-6 | 80.94 |  |  |
|  | q17-6 | 57.84 | q17-6 | 69.58 |  |  |
|  | q49-5 | 56.82 |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table. 6 variable importance table for each local school district, variable importance > 10** | | | | | |  |
|  | **Class 1** | **Importance valuel for class 1** | **Class 2** | **Importance valuel for class 2** | **Class 3** | **Importance valuel for class 3** |
| Broward County, FL (FT) | None | 0 | None | 0 | None | 0 |
| Chicago, IL (CH) | q49-4 | 100 | q41-4 | 15.28 | None | 0 |
| Eaton Consortium, MI (EA) | None | 0 | None | 0 | None | 0 |
| Fort Worth, TX (FW) | None | 0 | None | 0 | None | 0 |
| Genesee Consortium, MI (GE) | q17-7 | 100.00 | q49-6 | 72.61 | q49-4 | 43.78 |
|  | q53-2 | 89.72 | q41-5 | 72.05 | q17-4 | 22.79 |
|  | q21-3 | 72.04 | q17-5 | 42.66 |  |  |
|  | q21-5 | 67.37 | q89-4 | 39.12 |  |  |
|  | q87-2 | 35.14 | q41-4 | 33.77 |  |  |
|  | q65-4 | 34.62 | qnothhpl-2 | 32.85 |  |  |
|  | q23-2 | 31.05 | q41-2 | 32.56 |  |  |
|  | q49-4 | 18.93 | q49-3 | 27.81 |  |  |
|  | q17-2 | 18.34 | q89-7 | 27.60 |  |  |
|  | grade-2 | 16.36 | q47-4 | 25.29 |  |  |
|  |  |  | q49-2 | 25.03 |  |  |
|  |  |  | q47-2 | 21.74 |  |  |
|  |  |  | q52-2 | 18.98 |  |  |
| Hillsborough County, FL (HL) | None | 0 | None | 0 | None | 0 |
| Los Angeles, CA (LO) | None | 0 | None | 0 | None | 0 |
| Newark, NJ (NW) | qnothhpl-2 | 23.34 | q47-6 | 100.00 | q87-3 | 82.23 |
|  | q49-4 | 10.09 |  |  |  |  |
| Orange County, FL (OL) | qnothhpl2 | 36.47 | q213 | 76.12 | q896 | 100.00 |
|  | q895 | 34.02 | q474 | 35.63 | q474 | 21.83 |
|  | q654 | 25.25 | q475 | 24.85 | q872 | 16.35 |
|  | q633 | 22.42 | q496 | 16.33 |  |  |
|  | age6 | 13.31 | race43 | 11.22 |  |  |
| Palm Beach County, FL (PB) | qnothhpl-2 | 10.67 | q21-4 | 100.00 | q17-6 | 30.71 |
|  |  |  | q17-6 | 57.22 | q87-2 | 18.93 |
|  |  |  | q49-6 | 36.09 | q47-2 | 11.51 |
|  |  |  | q63-3 | 22.22 |  |  |
|  |  |  | q89-4 | 20.08 |  |  |
|  |  |  | q17-4 | 19.50 |  |  |
|  |  |  | q89-7 | 17.55 |  |  |
|  |  |  | q53-2 | 12.42 |  |  |
| Pasco County, FL (PS) | q17-8 | 100.00 | qnothhpl-2 | 14.87 | q47-5 | 74.58 |
|  | q49-3 | 17.75 | q50-2 | 14.04 |  |  |
|  | sex-2 | 17.38 |  |  |  |  |
|  | grade-4 | 12.55 |  |  |  |  |
|  | q30-2 | 10.61 |  |  |  |  |
| Philadelphia, PA (PH) | None | 0.00 | None | 0.00 | None | 0.00 |
| Portland, OR (PO) | qnothhpl-2 | 90.60 | age-7 | 100.00 | None | 0.00 |
|  | sex-2 | 75.05 | grade-2 | 95.15 |  |  |
| San Francisco, CA (SF) | q17-8 | 92.36 | qnothhpl-2 | 70.66 | q49-2 | 100.00 |
|  | q41-5 | 85.03 | q30-2 | 44.69 |  |  |
|  | q47-4 | 42.31 | age-4 | 42.26 |  |  |
|  | q17-5 | 40.17 | q49-5 | 33.95 |  |  |
|  | age-6 | 15.93 | q65-3 | 21.01 |  |  |
|  | age-4 | 15.75 | q21-2 | 19.26 |  |  |
|  | q87-3 | 12.80 |  |  |  |  |
| Shelby County, TN (ST) | q17-6 | 86.44 | q47-2 | 41.69 | q21-6 | 100.00 |
|  | q41-4 | 78.60 | q89-7 | 25.18 | q17-8 | 41.63 |
|  | sex-2 | 18.16 | race4-2 | 23.41 |  |  |
|  | q47-3 | 15.50 | q47-3 | 13.75 |  |  |