Supplementary Figure 1: Item characteristics curve after one-parameter logistic model

Supplementary Figure 2: Algorithm showing the linkage of factors associated with self-reported genital tract infection

Supplementary Table 1: Proportion of Sexually transmitted infection by socio-demographic characteristics among men and women aged 15-49 year, GDHS 2003–2014

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **STIs proportion** | **95%CI** | **Rao-Scott Wald Chi** |
| **Wealth quintile** |  |  |  |
| poorest | 10.9 | 9.6 to 12.4 | **13.27\*\*\*** |
| poorer | 11.1 | 10.0 to 12.3 |  |
| middle | 12.5 | 11.4 to 13.7 |  |
| richer | 13.3 | 12.2 to 14.6 |  |
| richest | 11 | 10.0 to 12.1 |  |
| **Household members** |  |  | **71.74\*\*\*** |
| ≤3 | 14 | 13.1 to 15.0 |  |
| 4-5 | 12.3 | 11.5 to 13.3 |  |
| 6-8 | 9.5 | 8.7 to 10.3 |  |
| 9+ | 9.2 | 8.0 to 10.5 |  |
| **Region** |  |  | **73.96\*\*\*** |
| western | 9.3 | 7.8 to 11.0 |  |
| central | 11 | 9.6 to 12.5 |  |
| greater accra | 12.1 | 10.7 to 13.7 |  |
| volta | 13.1 | 11.3 to 15.1 |  |
| eastern | 10.9 | 9.6 to 12.4 |  |
| ashanti | 12.9 | 11.5 to 14.4 |  |
| brong ahafo | 14.2 | 12.1 to 16.5 |  |
| northern | 13.8 | 12.0 to 15.8 |  |
| upper west | 7 | 5.6 to 8.5 |  |
| upper east | 7.6 | 5.8 to 10.0 |  |
| **Place of residence** |  |  | **17.30\*\*\*** |
| urban | 13 | 12.2 to 13.9 |  |
| rural | 10.6 | 9.9 to 11.4 |  |
| **Sanitation** |  |  | **178.53\*\*\*** |
| Unimproved | 8.4 | 7.8 to 9.0 |  |
| Improved | 16.2 | 15.2 to 17.3 |  |
| **Sex** |  |  | **508.68\*\*\*** |
| Female | 16.1 | 15.3 to 17.0 |  |
| Male | 5.2 | 4.7 to 5.7 |  |
| **Age group** |  |  | **324.82\*\*\*** |
| 15-19 | 7.1 | 6.4 to 7.9 |  |
| 20-24 | 17.2 | 16.0 to 18.5 |  |
| 25-29 | 15.5 | 14.2 to 16.8 |  |
| 30-34 | 12.7 | 11.5 to 14.0 |  |
| 35-39 | 11.9 | 10.6 to 13.2 |  |
| 40-44 | 10 | 8.7 to 11.3 |  |
| 45-49 | 6.5 | 5.5 to 7.7 |  |
| **Educational level** |  |  | **7.17\*** |
| None | 11 | 9.8 to 12.3 |  |
| Primary | 11.6 | 10.6 to 12.7 |  |
| Secondary | 12.3 | 11.6 to 13.0 |  |
| Higher | 10.4 | 8.8 to 12.2 |  |
| **Sexual initiation** |  |  | **166.69\*\*\*** |
| Late | 13.7 | 13.0 to 14.4 |  |
| Early | 8.5 | 7.8 to 9.2 |  |
| **Currently working** |  |  | **17.51\*\*\*** |
| No | 10.2 | 9.4 to 11.1 |  |
| Yes | 12.3 | 11.7 to 13.0 |  |
| **Staying with partner** |  |  | **7.93\*\*** |
| No | 11.6 | 11.0 to 12.2 |  |
| Yes | 13.7 | 12.3 to 15.3 |  |

$$sGTIs=-206.3026+ 0.2012\*p1a \_{i}+ 0.1287\*p1b\_{i} + 0.1506\*p1c\_{i} + 0.0893\*p1d\_{i} + 1.4802\*p2b\_{i} + 0.2496\*p3a\_{i} + 0.4514\*p3b\_{i} + 0.1934\*p3c\_{i} + 0.5585\*p3d\_{i} + 0.1327\*p4a\_{i}-0.1863\*p4c\_{i} - 0.2292\*p4d \_{i}- 0.2307\*p5a \_{i}- 0.0336\*p5b\_{i} + 0.3024\*p5d\_{i} + 0.0094\*p5e\_{i} + 0.1797\*p5f\_{i} + 0.3554\*p5g\_{i} + 0.5028\*p5h\_{i}- 0.3033\*p5i\_{i} - 0.0300\*p5j\_{i}+ 0.1701\*p6a\_{i} + 0.0015\*p7a\_{i} - 0.1371\*p8a\_{i} + 0.4918\*p8b\_{i} + 0.2904\*p8c\_{i} + 0.0507\*p8d\_{i} - 0.3001\*p8f\_{i} - 0.6623\*p8g\_{i} - 0.2108\*p9a\_{i} + 0.0416\*p9c\_{i} - 0.1887\*p9d\_{i} + 0.4167\*p10a\_{i} - 0.1874\*p11a\_{i} + 0.0845\*p12a\_{i} + 0.1008\*p13a\_{i}+ e\_{i} $$

Where; $i$ is the individual person aged 15-49 years, and e is the error term $e\_{i}\~N(0,σ^{2})$. The variable definition and their respective codes are presented in Table 2.

 Supplementary Figure 1. LASSO penalized 10-fold cross-validation regression model discrimination and calibration belt from bootstrapping analysis technique