

## Statistical Appendix

This survey adopts three-stage sampling design in District, community and household level respectively. Therefore, the population weighting is required when estimating the overall prevalence.

### 1. Individual sample weight

#### (1) Basic sample weight $W_{base}$

$$W_{base} = W_{s1} \times W_{s2} \times W_{s3}$$

The sample weight is the reciprocal of the sampling probability. The construction method of basic sample weight in this survey is shown in Supplementary Table 1.

Supplementary Table 1 The construction method of basic sample weight

Sampling Stage	Sampling Unit	Stratum	Sampling Unit weighting
1	District	High/medium/ low/zero incidence district	$W_{s1} = \frac{\text{Population in each stratum}}{\text{Population of sampled districts in the stratum}}$
2	Community	District	$W_{s2} = \frac{\text{Population in certain district}}{\text{Population of communities sampled in the district}}$ Sampled households
3	Household	Age group (1~9, 10~19,20~39, 40~59 & ≥60ys)	$W_{s3} = \frac{\text{Total households in certain community}}{\text{Households sampled in the community}}$ Added households $W_{s3} = \frac{\text{Population in certain age group of non – sampling households of community}}{\text{Population of added households in the age group}}$

#### (2) Post-stratification adjustment of demography

The post-stratification adjustment is to adjust the weight with the real total population. The sum of all individual sample weights (the basic sample weight after adjusting no response) is the estimate of the total target population. Due to selective biases such as omissions, this estimate is not completely consistent with that of the true total population. Therefore, the sample weights need to be post-stratified modified to the real total population.

The demography adjusted weight of individual in certain age and gender group is expressed by  $W_{adj}$ .

$$W_{adj} = \frac{N_{ij}}{\sum W_{base}}$$

$N_{ij}$  is census population in the group of age  $i$  and gender  $j$ ;  $\sum W_{base}$  is the total basic sample weight in in the group of age  $i$  and gender  $j$ .

#### (3) Individual final weight ( $W_i$ )

$$W_i = W_{base} \times W_{adj}$$

## 2. The seroprevalence estimation

If  $y_i$  is final testing result with 1 for seropositivity and 0 for seronegativity, then the point value of the seroprevalence is estimated as:

$$p = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

## 3. Variance and 95%CI estimation

Taylor series linearization is used to estimate the variance of seroprevalence  $\hat{V}[p]$ .

When the prevalence ranges 25%-75%, 95%CI can be estimated by Wald (linearization):

$$p \pm t_{\alpha/2,df} \times \sqrt{\hat{V}[p]}$$

in which  $t_{\alpha/2,df}$  is  $100(1-\alpha/2)$ th quantile of the  $t$  distribution with the degree of freedom  $df$ .

When the prevalence <25% or >75%, 95%CI can be estimated by Wilson (score):

$$\left( p + z_{\alpha/2}^2 / 2n_e^* \right) \pm \left( z_{\alpha/2} \sqrt{\left( p(1-p) + z_{\alpha/2}^2 \right) / 4n_e^*} \right) / \left( 1 + z_{\alpha/2}^2 / n_e^* \right)$$

The effective sample size adjusted for degrees of freedom

$$n_e^* = n_e \left( t_{\alpha/2, n-1}^2 / t_{\alpha/2, df}^2 \right)$$

The effective sample size

$$n_e = n / deff$$