**Survey design and Sample size**

The aim of this study is to provide estimates on seroprevalence of COVID-19 at various administrative levels (regions and districts). Developing the sample design for such national surveys generally requires relatively large sample sizes if highly precise estimates are desired at lower administrative levels. We thus examined sample sizes under various scenarios. Our sampling strategy will use a stratified two-stage design with census enumeration areas (EAs) in each district as the primary sampling units (PSUs) and households (HHs) as the second-stage units. There are 48 districts within 8 Local Government Areas (LGAs), which constituted a total of 4098 EAs. The number of EAs per district range from 7 in the Janjanbureh district to 704 in the Kombo North district. For comparison, we consider three approaches to distribute a given total sample sizes of interest (N) across each district. These are:

* 1. proportional allocation of the samples across the districts (PPS)
  2. equal sample size allocation for each district
  3. a compromise sample allocation that falls between the proportional and equal allocations (Kish 1976, 1988).

For each scenario, we calculated the precision for estimating 95% CI for the seroprevalence. Hence, the selection of a specific sample size can be based on achieving a desired precision for estimating seroprevalence (in addition to budget). The PPS approach favours precision at national level while the equal allocation approach favours precision at district level. Thus, we used the compromise approach (3 above) which provides good precision for national level estimates but still better precision than the PPS approach for district level estimates. For example, for a total sample size of 1500 and estimated seroprevalence of 10%, our method provides a margin of error about 3% at the national level. Assuming 10 HHs from each EA (one member from each HH ), the number of EAs to be selected per district ranges between 2 and 17 for the smallest and largest districts with 7 and 704 EAs respectively. The selection of HH members will take into account the need to balance for age groups. One simple approach here can be to select 2 HH members from each of the five age groups 5-9, 10-19, 20-39, 40-59 and ≥60 in each EA (i.e. 2x5=10HHs per EA). The 10 HH per EA will be randomly selected based on sampling frames to be prepared in collaboration with the Gambia Bureau of Statistics (GBoS). That is, HH listing will be undertaken for all selected EAs prior to data collection.

**Appendix**

1. There are 48 districts within 8 LGAs, which constituted a total of 4098 EAs. The number of EAs per district ranged from 7 in the Janjanbureh district to 704 in the Kombo North district (see Table 1).
2. For the purpose of sample size calculations, we explored a range of proportions/ seroprevalence (P) between 5%-10% assuming these would be similar for all the districts in the country.
3. The range of overall samples sizes (n) considered was 1000-15000 in steps of 1000.
4. The PPS weights for each district ranged from 0.0017 to 0.1717 (see Table 1).
5. We allowed for clustering of EAs within districts; an average intra-cluster correlation coefficient (ICC) of 0.05 was assumed (ρ=0.05).
6. For comparison, we consider three approaches to distribute the total sample sizes (n) given above (3) across each district. These are:
   1. proportional allocation of the samples across the districts
   2. equal sample size allocation for each district
   3. a compromise sample allocation that falls between the proportional and equal allocations.

In the first approach (6.1) related to proportional allocation, the design effect (DE) which is due to stratification by district can be given by



where n is total sample size, nh is sample size for the hth district (strata) and Wh is the strata weight. This proportional allocation scheme provides more precise estimates for higher administrative levels (e.g. LGAs) while precision can be poor at lower administrative levels (e.g. Districts). On the other hand, the equal allocation (6.2) above can provide better precision for lower administrative levels (e.g. Districts) but decreased precision for higher administrative levels (e.g. LGAs). Thus the compromise approach (6.3) has been proposed where relatively higher precision at both lower and higher administrative level is considered of similar importance. One can compute such a compromise based on an allocation proposed by Kish (1976, 1988) given by ; such allocation increases the sample sizes, for example, for the small districts considerably over the proportionate allocation, but not as much as the equal allocation. The design effect for this unequal weighting allocation is also smaller compared with the equal sample size allocation.

1. Next, the number of HHs to be sampled within each EA (average cluster size) was taken as 15. This can be obtained from a simple cost model as an optimum number of households (m) to select per sampled EA as given by the equation below:



where C is the ratio of the cost of sampling an EA to the cost of adding a household. Here for the average ρ=0.05 as specified in (5) above and taking C=12 gives the average HH size m=15.

1. Given the average ρ=0.05 and average HH size m=15, the design effect due to clustering of EAs within districts can be calculated as



which gives *DEcl*=1.70. This design effect needs to be taken into account in determining the precision of district level estimates.

1. The design effect for national/LGA estimates needs to combine the design effects for clustering and the disproportionate allocation across districts. This may be obtained as *DE*=*DEst\* DEcl*.
2. For each scenario we calculate the coefficient of variation (CV) and the precision for estimating 95% CI. Hence, the selection of a specific sample size can be based on achieving a desired CV or precision for estimating EPI coverage proportions (in addition to budget).

Table 1: Distribution of EAs and respective district weights for PPS sampling

|  |  |  |  |
| --- | --- | --- | --- |
| LGA | District | No. of EAs | Weight |
| Banjul | Banjul South | 18 | 0.004392 |
|  | Banjul Central | 31 | 0.007565 |
|  | Banjul North | 25 | 0.006101 |
| Kanifing | Bakau | 58 | 0.014153 |
|  | Jeshwang | 139 | 0.033919 |
|  | Serrekunda Central | 158 | 0.038555 |
|  | Serrekunda East | 227 | 0.055393 |
|  | Serrekunda West | 191 | 0.046608 |
| Brikama | Kombo North | 704 | 0.171791 |
|  | Kombo South | 226 | 0.055149 |
|  | Kombo Central | 300 | 0.073206 |
|  | Kombo East | 93 | 0.022694 |
|  | Foni Bintang | 39 | 0.009517 |
|  | Foni Kansalla | 30 | 0.007321 |
|  | Foni Brefet | 36 | 0.008785 |
|  | Foni Bundali | 20 | 0.00488 |
|  | Foni Jarrol | 18 | 0.004392 |
| Mansakonko | Jarra West | 59 | 0.014397 |
|  | Kiang West | 41 | 0.010005 |
|  | Kiang Cental | 25 | 0.006101 |
|  | Kiang East | 19 | 0.004636 |
|  | Jarra Central | 21 | 0.005125 |
|  | Jarra East | 39 | 0.009517 |
| Kerewan | Lower Niumi | 125 | 0.030503 |
|  | Lower Badibu | 41 | 0.010005 |
|  | Illiasa | 109 | 0.026598 |
|  | Upper Niumi | 74 | 0.018058 |
|  | Jokadu | 52 | 0.012689 |
|  | Central Badibu | 43 | 0.010493 |
|  | Sabach Sanjal | 49 | 0.011957 |
| Kuntaur | Lower Saloum | 36 | 0.008785 |
|  | Niani | 73 | 0.017814 |
|  | Upper Saloum | 46 | 0.011225 |
|  | Nianija | 24 | 0.005857 |
|  | Sami | 58 | 0.014153 |
| Janjanbureh | Lower Fuladu West | 95 | 0.023182 |
|  | Upper Fuladu West | 104 | 0.025378 |
|  | Janjanbureh | 7 | 0.001708 |
|  | Niamina Dankunku | 15 | 0.00366 |
|  | Niamina West | 20 | 0.00488 |
|  | Niamina East | 56 | 0.013665 |
| Basse | Jimara | 99 | 0.024158 |
|  | Basse | 119 | 0.029039 |
|  | Tumana | 85 | 0.020742 |
|  | Kantora | 86 | 0.020986 |
|  | Wuli East | 55 | 0.013421 |
|  | Sandu | 60 | 0.014641 |
|  | Wuli West | 50 | 0.012201 |
| **Total** |  | 4098 | 1 |