WHO/IVB/17.07
ORIGINAL: ENGLISH

INFORMING VACCINATION PROGRAMS:

A Guide to the Design and Conduct of Dengue Serosurveys

Immunization, Vaccines and Biologicals

World Health Organization

Informing vaccination programs: a guide to the design and conduct of dengue serosurveys

ISBN 978-92-4-151258-9

Suggested citation. Informing vaccination programs: a guide to the design and conduct of dengue serosurveys. Geneva: World Health Organization; 2017.
Licence: CC BY-NC-SA 3.0 IGO.
Cataloguing-in-Publication (CIP) data. CIP data are available at *http://apps.who.int/iris*.

Printed in Geneva, Switzerland

WHO/IVB/17.07
ORIGINAL: ENGLISH

INFORMING VACCINATION PROGRAMS:

A Guide to the Design and Conduct of Dengue Serosurveys

Immunization, Vaccines and Biologicals

World Health Organization

Contents

Acknowledgements

The guidance articulated in this document is based on discussions of an ad hoc WHO Technical Working Group on Dengue Serosurveys to Target Vaccination. This document was developed on behalf of WHO by **Natalie Dean** (University of Florida).

Special thanks are due to the numerous colleagues who contributed to the discussions and preparation of this document, including the following (in alphabetical order): **Katie Anderson** (University of Minnesota), **Brad Biggerstaff** (US Centers for Disease Control and Prevention), **Oliver Brady** (London School of Hygiene and Tropical Medicine), **Marcelo N. Burattini** (University of Sao Paolo), **Neil Ferguson** (Imperial College), **Emily Gurley** (ICDDR,B), **Elizabeth Hunsperger** (US Centers for Disease Control and Prevention), **Barbara Johnson** (US Centers for Disease Control and Prevention), **Natsuko Imai** (Imperial College), **Ira Longini** (University of Florida), **Mick Mulders** (WHO), **Minal Patel** (WHO), **Isabel Rodriguez** (Johns Hopkins School of Public Health), **Diana Rojas** (Red AEDES - Colombia, University of Florida), **Hasitha Tissera** (Sri Lanka Ministry of Health), **Kirsten Vannice** (WHO), and **In-Kyu Yoon** (International Vaccine Institute).

Input was also provided during a review by the WHO Immunization and vaccines related implementation research advisory committee (IVIR-AC). The meeting report and recommendations may be found at *http://www.who.int/immunization/research/committees/ivir_ac*.

Abbreviations & acronyms

AIC	Akaike information criterion
CHIKV	chikungunya virus
CI	confidence interval
CYD-TDV	Dengvaxia® vaccine
DEFF	design effect
DENV	dengue virus
E	envelope protein
ELISA	enzyme-linked immunosorbent assay
ESS	effective sample size
FN	false negative
FOI	force of infection
FP	false positive
HPV	human papillomavirus
ICC	intracluster/intraclass correlation coefficient
IgG	immunoglobulin G
IgM	immunoglobulin M
JE	Japanese encephalitis
NPV	negative predictive value
NS1	non-structural protein 1
NT	neutralization test
OD	optical density
PPS	probability proportional to size
PPV	positive predictive value
ROC	receiver operator curve
PRNT	plaque reduction neutralization test
SAGE	WHO Strategic Advisory Group of Experts

SOP	standard operating procedure
TdaP	tetanus, diphtheria, and pertussis
TN	true negative
TP	true positive
WHO	World Health Organization
WNV	West Nile virus
YF	yellow fever
ZIKV	Zika virus

1. Introduction

Dengue virus (DENV) is a growing problem in tropical and subtropical regions of the world. It is estimated that 50 to 100 million people are infected and develop symptomatic dengue annually *(1, 2)*.

Since the end of 2015, the first dengue vaccine, a live-attenuated tetravalent vaccine (CYD-TDV, or Dengvaxia®), has been licensed in several countries in Asia and Latin America. The vaccination schedule consists of 3 injections administered at 6-month intervals with the lower limit of the indication at 9 years of age. In the Phase 3 clinical trials, vaccine efficacy was highest for recipients with pre-existing immunity to dengue as well as older trial participants, noting that immunity levels tend to increase with age. As a consequence, the World Health Organization (WHO) Strategic Advisory Group of Experts (SAGE) recommends that countries consider introduction of the dengue vaccine CYD-TDV only in populations (national or subnational) where epidemiological data indicate a high burden of disease. In defining populations to be targeted for vaccination, prior infection with any serotype of dengue virus, as measured by dengue seroprevalence, should be approximately 70% or greater in the age group targeted for vaccination, in order to maximize public health impact and cost-effectiveness *(3)*. Vaccination of populations with seroprevalence between 50% and 70% is acceptable but the impact of the vaccination program may be lower. The vaccine is not recommended when seroprevalence is below 50% in the age group targeted for vaccination because of low efficacy and potential longer-term risks of severe dengue in vaccinated seronegative individuals *(4)*. To ensure that the vaccine is optimally used for public health, countries should ensure that the age groups targeted have enough pre-existing immunity to benefit from this vaccine and minimize potential theoretical harms.

There are multiple sources of epidemiologic data that could be used as evidence of high pre-existing immunity to dengue. Ideally a country will have nationally representative surveillance data, indicating where incidence of disease is high and where a dengue vaccination program might be most useful. However, surveillance data alone can be unreliable, as clinically apparent cases represent a variable fraction of all dengue infections, typically estimated to be around 25%, and healthcare seeking for dengue can vary greatly based on access to care *(1, 2)*. Because surveillance data can be unreliable *(5)*, population-based seroprevalence studies are the best way to measure past infection with DENV.

Since most countries do not have these data already available, a well-designed dengue serosurvey is recommended to support decision-making for vaccine introduction and to identify which populations to target. A serosurvey involves collecting and testing blood specimens from a defined population to estimate the proportion positive for

DENV immunoglobulin G (IgG) antibodies as a measure of population immunity. Countries considering vaccination are advised to have available at least one recent age-stratified serosurvey (within the last 3–5 years) in a geographically relevant location and capturing the likely vaccine target age range. Investments in serosurveys to target vaccination are justifiable to support vaccine decision-making given the financial costs associated with vaccine introduction and the potential sub-optimal cost-effectiveness if the vaccine is not targeted appropriately, including a potential increased risk for severe cases. A serosurvey precisely at all locations targeted for vaccination is not required (or feasible), but an assessment of expected seroprevalence in the target resident population and age group should be done.

Countries may be interested in using blood samples collected for other purposes to assess seroprevalence. However, these samples would not typically be an appropriate substitute for a well-designed serosurvey. For example, since serosurveys would need to focus on children, samples from blood banks are unlikely to be useful as children do not usually donate blood. However, samples collected as part of representative, population-based serosurveys might be useful. Samples from sick children may not be generalizable to healthy children. Samples would also need to cover an adequate range of geographic areas and ages with enough depth to inform vaccination policy.

In this document, recommendations are provided on designing and implementing a cross-sectional serosurvey using school-based sampling to estimate age-specific DENV seroprevalence to inform a country's national dengue vaccination program. The document includes recommendations for methods for planning and conducting serosurveys, including survey design, specimen collection, laboratory testing, data analysis, and the interpretation and reporting of results.

2. Designing a dengue serosurvey

This document is primarily aimed at lead investigator(s), public health officers, and policy makers to assist in designing dengue serosurveys. In this section, recommendations are provided for defining the survey objectives, survey population, survey design, sample size, and diagnostic assays.

2.1. Survey objectives

The primary objective is to estimate age-specific seroprevalence of dengue antibodies in order to comply with the WHO recommendations that the seroprevalence of dengue antibody in the target population for vaccination is not less than 50% and is ideally greater than 70%. Ages are categorized into one-year bands with seroprevalence estimated separately for each age. Estimated age-specific seroprevalence at each age should be sufficiently precise (i.e. with a narrow confidence interval) to support the primary objective.

A secondary objective is to identify routinely available data which can be used to characterize or predict local seroprevalence. Such data sources include ecological, demographic, or dengue surveillance data. If their correlation with seroprevalence is locally validated using data from the serosurvey, these variables will be useful for guiding vaccination policy in broader geographic areas.

Additional secondary objectives may be specified, though these may increase the cost and logistical complexity of the serosurvey. Examples include informing dengue burden estimation by expanding the geographic range of the serosurvey to lower burden areas that may not be considered for dengue vaccination. The age range of the serosurvey may be expanded beyond what is minimally necessary to more reliably estimate transmission intensity and force of infection (FOI). It may be desired to better characterize spatial heterogeneity in dengue seroprevalence, thereby necessitating greater breadth of sampling. Finally, serosurveys may be expanded to measure the prevalence of other biological markers of infection, including antibodies against specific DENV serotypes circulating using plaque reduction neutralization tests (PRNTs); antibodies for other flaviviruses such as Zika virus (ZIKV), Japanese encephalitis (JE), or yellow fever (YF); antibodies for other arboviruses, such as chikungunya virus (CHIKV); or even antibodies for pathogens that are not vector-borne. Combining serosurveys for multiple diseases may be cost-efficient and improve interpretation of results where there is cross-reactivity between co-circulating viruses. Expanding the objectives of

the serosurvey to include other diseases may require an expansion of the age range or of the survey region to ensure that the results are generalizable and useful.

2.2. Age range to survey

WHO policy supports use of this vaccine in individuals aged 9–45 years. In areas with high dengue burden, policies will likely target school-aged children; therefore, recommendations are provided for serosurveys targeting children aged 5–18 years, though it is expected that most countries will select a narrower age range depending on their local epidemiological context and their available resources.

School-based sampling is the most efficient way to enroll children for serosurveys as long as more than 95% of children in the relevant age group attend school. This document assumes that this will be the primary strategy used by countries considering vaccination. If this is not the case, the survey may be expanded to include some form of community-based sampling, such as sampling children attending basic health facilities during a non-dengue season, but this document does not provide guidance on such an approach.

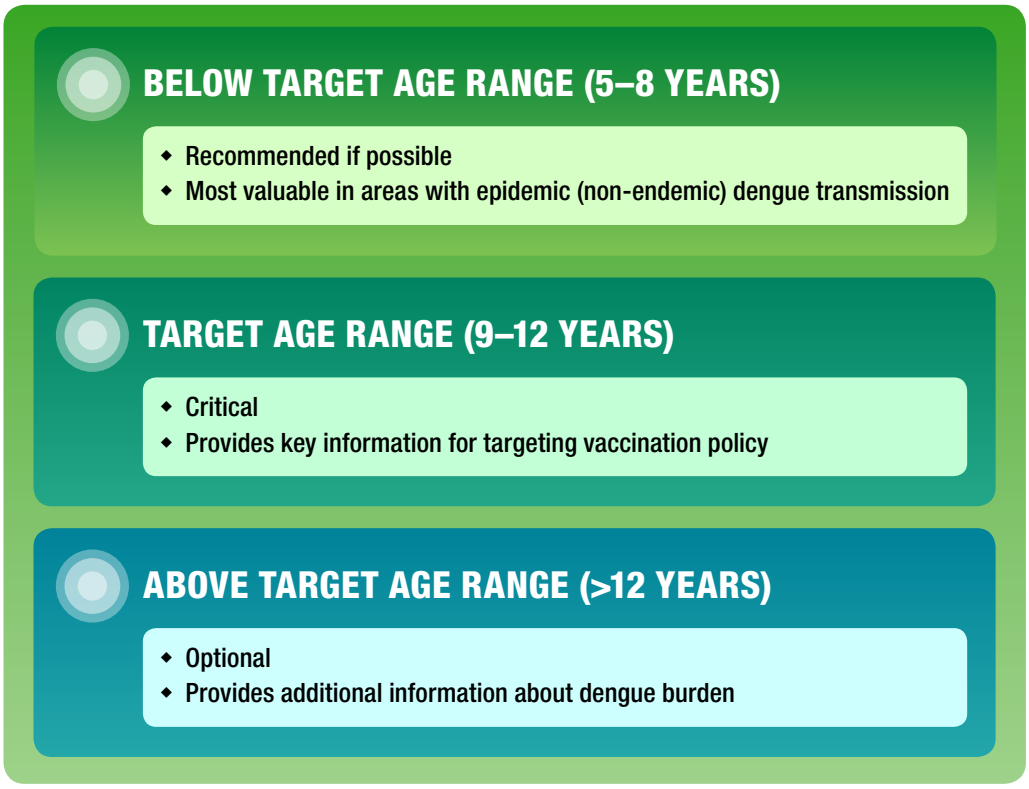
The serosurvey informs selection of the target age for vaccination. Seroprevalence typically increases with age, and countries may choose to target vaccination to the youngest age (9 years or older) for which seroprevalence exceeds the recommended 70% threshold. Other programmatic factors may influence the precise age chosen by countries for vaccination, such as alignment with existing vaccination schedules, such as for human papillomavirus (HPV) or tetanus, diphtheria, and pertussis (TdaP), as WHO permits the co-administration of these non-live vaccines with the dengue vaccine. Co-administration may be desirable to reduce programmatic costs associated with school-based vaccination programs. For practical reasons, if the vaccine is introduced into multiple areas in a country, it is expected that only one age (or school grade) should be targeted for routine vaccination.

The serosurvey should at minimum include ages 9 to 12, which are the likely target ages for vaccination in most highly endemic countries. Countries considering vaccination policy in children >12 years should include older children in the serosurvey.

Though children under 9 years of age will not be targeted for vaccination, sampling children aged 5 to 8 years from the same schools in the serosurvey is advised. In places where dengue transmission is epidemic and not endemic, it might be possible to have high seroprevalence in one age group but low seroprevalence in a slightly younger age group, depending on dengue transmission, seasonality, and the timing of recent outbreaks. To conserve resources, younger age groups could be under-sampled relative to the target age group of children aged 9 to 12. Including younger age groups may also provide valuable information for future evaluations of the impact of the dengue vaccine on dengue transmission. If resources are available, age groups could be sampled above the target age range to inform dengue burden estimation, though school attendance is known to decline in older children so results may only be representative of children still in school.

Fig. 1 provides a brief summary of the age groups targeted in this survey.

Fig. 1: Considerations for age groups to survey



2.3. Area to survey

An important design feature is the geographic or administrative populations for which age-specific seroprevalence will be estimated.

Dengue incidence is highly heterogeneous even at small spatial scales, exhibiting substantial temporal and geographic variability *(6)*. Large differences in annual dengue incidence may be seen in neighboring municipalities *(7)*. Spatial variations are influenced strongly by rainfall, temperature, and the degree of urbanization *(1, 8)*. See Box 1 for examples of drivers of heterogeneity in dengue incidence and seroprevalence.

Because of the heterogeneity of dengue transmission, vaccine introduction may only be appropriate in confined subnational geographic areas where seroprevalence reaches the levels recommended by WHO. Subnational vaccine introduction has been successfully used for other vaccines, including for JE and YF, which are also used in dengue-endemic countries.

Before conducting this survey, the smallest administrative unit that will be considered for vaccination policy should be specified. This could be a large administrative unit, like a province, state, or department, or a small administrative unit, like a municipality or county. Spatial heterogeneity does not follow administrative borders, which is typically how vaccination programs have been structured. Decision-makers will need to assess their local dengue epidemiology to inform the spatial scale that is scientifically justified while also administratively feasible for implementation. Vaccine policies may be easier to operationalize for a larger administrative unit, but there may be significant heterogeneity within the unit, such that vaccination may not be uniformly appropriate throughout the unit. An example is a region that is partly high altitude with no mosquito vector and partly low altitude with endemic dengue. With a smaller administrative unit, the program can more reliably target vaccination to high burden areas, though it may be more difficult to implement the policy. On the other hand, it may be easier to conduct the serosurvey if smaller administrative units are used. A more targeted approach may also require fewer vaccinations overall as only high burden areas receive vaccination.

WHO advises that a serosurvey precisely in the location of planned vaccination is not required, but an assessment of likely seroprevalence should be done. As evaluating seroprevalence in each administrative unit may not be feasible, administrative units are divided into categories based on predicted seroprevalence. These categories are referred to as strata. In the serosurvey, only a few administrative units from each stratum are randomly selected and surveyed, and serosurvey results from the selected administrative units are used to make predictions about likely seroprevalence in other administrative units within the same stratum. This approach works best when the administrative units within each stratum have similar seroprevalence.

It is recommended that three dengue burden strata be defined based on the predicted seroprevalence in each administrative unit; these categories are labeled as highest, middle, and lowest burden, and allocation to each group is based on the best information available to predict seroprevalence. The labels of highest, middle, and lowest burden do not correspond to standardized definitions of dengue burden. For this reason, suggested cutoffs for highest, middle, and lowest groupings in terms of reported cases or likely seroprevalence are not provided; these are likely to be highly context-specific and difficult to operationalize without reliable existing seroprevalence data. The goal is to provide the best grouping of administrative units based on likely seroprevalence in children in order to improve the efficiency of the serosurvey and the interpretability of its results. Some countries may also have administrative units with no dengue transmission; these administrative units may be considered as part of a fourth stratum that is excluded from the serosurvey.

A natural way to define predicted seroprevalence would be to categorize administrative units based on surveillance data, ideally age-stratified subnational estimates of cumulative incidence in school-aged children over the past five years. High quality surveillance data will include laboratory confirmation of cases *(9)*. Mean age of cases may be another useful metric for classifying administrative units within a country, with units with the lowest mean age being assigned to the highest dengue burden stratum.

Variability in dengue incidence may be masked by variability in reporting, and countries may wish to supplement surveillance data with other information about the local

context when classifying administrative units into strata. Countries may incorporate results from any earlier serosurveys that have been conducted, if those are available. Countries may also consider the predictors of dengue incidence and seroprevalence in Box 1. Countries may use epidemiological models or dengue risk maps, where available. In addition, the reliability of surveillance data may be heterogeneous, with barriers to treatment seeking and reporting potentially being driven by socioeconomic status. Local epidemiologists may help assess whether areas reporting low dengue case numbers are truly lowest dengue areas or areas where the disease is underreported. Guidelines are available for assessing completeness of routine surveillance for dengue.[1]

Based on their desired vaccination strategy, countries will next identify which strata to include in the serosurvey. Table 1 summarizes the possible approaches.

BOX 1. Drivers of spatial heterogeneity in dengue incidence and seroprevalence

- Population size, density, and degree of urbanization/vegetation
- Socioeconomic status
- Altitude
- Rainfall, temperature, and average temperature during the rainy season
- Climatic types (humid, subhumid, etc.)
- Mosquito species and density (if density is not high throughout the country)
- Mosquito host preference
- Presence of other vector control measures
- Distance/mobility from other areas of DENV transmission
- History of dengue in that area
- Circulating serotype(s)

Table 1: Potential serosurvey strategies

Dengue Burden Strata (based on surveillance data)	National vaccination	Targeted (regional) vaccination	Highly targeted (regional) vaccination	Targeted (regional) vaccination, plus burden estimation
Lowest	✔	✖	✖	✔
Middle	✔	✔	✖	✔
Highest	✔	✔	✔	✔

Countries interested in a national vaccine policy must assess all strata (highest, middle, and lowest) in the serosurvey. Age-specific seroprevalence will be estimated separately within each stratum, though results can be combined across strata to obtain a national estimate. Nonetheless, it is recommended that the stratum-specific estimates are used

1 A Tool Kit for National Dengue Burden Estimation: *http://www.who.int/denguecontrol/en/*

to guide policy decisions rather than the national estimate; this approach will help countries assess whether the national policy is appropriate in all administrative units, including units with the lowest dengue burden. In some cases, through countries may desire a national vaccination policy, the seroprevalence may be too low in certain areas to support such a policy.

Other countries may prefer a subnational (regional) vaccination policy using a more targeted approach. Given that WHO recommends a seroprevalence ideally of ≥70% in the targeted age range, it is possible that only the highest dengue burden areas will meet the seroprevalence thresholds for vaccine introduction. Countries may choose to focus serosurveys in these areas by only including the highest and middle dengue burden strata; in countries where resources are highly limited, countries may include only the highest dengue burden stratum, though there is then a risk of failing to capture areas classified as middle dengue burden by surveillance but that may have truly high seroprevalence. The lowest dengue burden stratum is excluded in both scenarios.

Finally, countries may desire a subnational policy but prefer a more inclusive serosurvey, with lowest, middle, and highest dengue burden strata represented to improve dengue burden estimation, even if these areas are unlikely to meet the seroprevalence target. Administrative units with no dengue transmission, for example where the mosquito vectors are absent, may be excluded entirely.

The recommended procedure for defining the strata is to first prepare a list of administrative units. The units should then be roughly categorized into highest, middle, and lowest dengue burden; for some countries, there may also be a no dengue burden category. The country should decide which survey approach it will take from the options described in Table 1. If the country selects an option in which not all strata are sampled, the country may then need to reclassify some administrative units so that all units to be targeted for vaccination are captured within the sampled strata. The strata included in the serosurvey should be of roughly equal size. For example, a country considering a targeted policy plans to only sample middle and highest dengue burden units. The country first identifies all lowest dengue burden administrative units that will not be considered for the vaccination policy and excludes these units. The remaining administrative units are divided into two roughly equal groups – middle and highest dengue burden – according to their likely seroprevalence.

2.4. Survey design

The appropriate serosurvey design to achieve the primary survey objective is a cross-sectional survey, i.e., measuring the seroprevalence of DENV in the study population at a single point in time. Because a census is resource-intensive, random sampling is recommended as a cost-effective method for obtaining seroprevalence estimates that are representative of the target population. It is important that sampling is conducted using true random sampling. Convenience sampling, such as selecting administrative units or schools that are easiest to sample, is expected to result in bias. This is because administrative units selected because of convenience may not be generalizable to the larger population. Possible sources of bias are outlined in Section 2.7.

To report seroprevalence estimates that are representative of the target population it will be necessary for participants to be selected using a probability sample wherein

1) every eligible respondent has a non-zero chance of being selected into the survey sample; and

2) for those respondents who are selected, the probability of selection can be calculated.

A complex survey design is recommended, such as a stratified multistage cluster sample with random selection at each stage. The definitions of these and other key terms in survey design are listed in the Glossary in Appendix 1. Cluster sampling allows individuals to be selected from only a subset of eligible administrative units and schools, thereby allowing the survey team to work in focused geographic areas. Clustering increases logistical feasibility and lowers field costs. The level of clustering required will depend on operational considerations, including the ability of a trained survey team to travel and effort required to engage with local governments. This is a multistage design because it requires sampling administrative units first, followed by schools, and followed by students (three stages).

In a stratified survey, the population is divided into non-overlapping groups, called strata, and sampling is conducted separately for each stratum. Stratified cluster sampling allows calculation and comparison of age-specific dengue seroprevalence in each dengue burden stratum. The survey should be powered to achieve sufficient precision for estimating stratum-specific mean age-specific seroprevalence. The survey does not need to be powered to achieve sufficient precision within finer groupings, such as within a particular region, or in urban as compared to rural areas. Such a survey is more difficult to design and could quickly become cost-prohibitive. Nonetheless, the survey data, once collected, can still be used to conduct a basic investigation of the impact of various factors on seroprevalence.

Useful references for designing and analyzing complex surveys are available online, including *WHO Immunization Coverage Cluster Surveys: Reference Manual (10)* and its 2015 draft update *(11)*. It may be useful to consult a sampling statistician or WHO to support the survey design and analysis.

2.4.1. Stage one: sampling administrative units

A list of administrative units in the country should be prepared along with the population size for each administrative unit and some available measure of dengue burden. Using the measure of dengue burden along with knowledge of the local epidemiological context, administrative units should be categorized into highest, middle, and lowest dengue burden strata, following the guidance in Section 2.3. Each stratum-specific list serves as the *sampling frame* for that stratum. A sampling frame lists all clusters and their relative sizes, and samples are drawn from this list.

Because the survey is stratified, a fixed number of units will be sampled from each stratum included in the survey. The number of units required will depend on the desired precision within that stratum. Units will be randomly sampled using *probability proportional to size* (PPS) sampling. In PPS sampling, larger units are more

likely to be sampled than smaller units, which improves the statistical precision of the survey. Very large units, especially areas with densely populated cities, may be so large that they are automatically selected. These are referred to as self-representing units or certainty units, and they are handled differently in the analysis. To avoid sampling units with certainty, very large units may be segmented into smaller areas prior to sampling *(11)*. Units not selected with certainty will be non-self-representing; this means that the seroprevalence results in selected units are used to represent or infer seroprevalence in units within the same stratum that were not selected.

To conduct the random sampling, a procedure known as *systematic sampling* is employed *(12)*. The sampling frame is prepared as a list, and a random starting point is selected at the top and then clusters are selected that are distributed evenly along the list. Systematic sampling can be conducted with PPS sampling or with equal probability sampling. If the list is sorted on some factor, such as geographic region, before conducting sampling, the sample is more likely to be balanced with respect to that factor; this method of sorting is referred to as *implicit stratification*. This approach does not complicate the analysis.

Before each stage of sampling is conducted, it is important to assess the accessibility of all administrative units or schools in the sampling frame. Administrative units or schools that for security or logistical reasons will not participate in the survey even if selected should be excluded from the sampling frame *a priori*. Their exclusion should be noted as a limitation and possible source of bias (see Section 2.7).

2.4.2. Stage two: sampling schools

Within each *selected administrative unit*, a list of schools and school sizes needs to be prepared. These lists will serve as sampling frames for the second stage of sampling. It is not necessary to prepare a list of schools for units that were not selected. It is suggested that two schools are selected from each sampled unit. No fewer than two schools should be selected because the variance and 95% confidence interval cannot be properly calculated when only one school is selected. On the other hand, by limiting the survey to only two schools per administrative unit, it is possible to achieve better overall geographic representation with the same sample size because more distinct administrative units can be selected.

The list of schools should include all schools, both public and private. This is especially important if a large fraction of children attend private schools. Efforts should be made to include children attending private schools as these children will also be eligible for vaccination, although it is recognized this may be infeasible. It should be understood that failing to include these schools can lead to selection bias, further discussed in Section 2.7. For example, children attending private schools are typically likely to come from families with higher socioeconomic status; these children may have lower exposure to dengue if they have screens in their windows or less standing water in their neighborhoods. In practice, it may be difficult to engage private schools; excluding them should be acknowledged as a limitation of the serosurvey.

The list of schools should be sorted in some manner prior to sampling. Natural options for sorting schools would be by region or by size. By sorting the list, implicit stratification is used to achieve a more representative random sample.

Sampling of schools can be conducted using one of two approaches. The advantages and disadvantages of each approach are described, and countries are able to decide which option they prefer. The first approach is to use PPS systematic sampling so that larger schools are more likely to be selected than smaller schools; the same number of students should then be selected from each school so that equal weighting is achieved across all students in the population. The advantage of this approach is that larger schools are preferentially sampled, which may make enrolling the required sample size easier to achieve. The disadvantage is that information about school size for all schools in the administrative unit is required up front; this information may be difficult to obtain in some countries.

The second approach is to sample schools with equal probability, regardless of their size; then, entire classrooms are sampled within each school so that proportionally more students are selected from larger schools than are selected from smaller schools. The advantage of this approach is that less information is required up front for sampling schools. Furthermore, experience suggests that consenting and taking specimens from entire classrooms is easier to implement and is more acceptable to parents. The disadvantage of this approach is that it may be hard to predict how many students should be sampled from each school and to plan the overall sample size of the survey. In addition, if an entire classroom is sampled but only a few students are needed from a particular age group, it may be necessary to sample more students than are required to achieve the desired precision; it may also be harder to control the number of students enrolled in each particular age group, as age may vary within a classroom.

The procedure for school sampling will also depend upon the desired age range, as determined in Section 2.2. If the desired age range for the survey includes ages above the maximum age at any of the sampled primary/elementary schools, then another school with older students should also be sampled that will capture the age range of interest. For feasibility reasons, it is recommended that this school be from the same school system, such as the school that the primary/elementary students will attend next.

In settings where a list of schools within the administrative unit cannot be prepared at the unit level, then an additional stage of sampling will be required. This may be necessary for large administrative units with poor availability of centralized records. The highest administrative level able to construct a list of schools needs to be identified. Then, a random sample of these smaller administrative units would be conducted, only preparing a list of schools within the smaller administrative units selected. This design is less statistically efficient, meaning it will achieve inferior precision as compared to a survey with fewer stages of sampling.

Another reason why a country may choose to add an additional stage of sampling is if sampling schools that are randomly located across the selected administrative units is too difficult. This could occur if the administrative units are very large, and for logistical reasons it is hard for the survey team to travel within the unit. Alternatively,

this could occur if it is resource-intensive to engage school systems. By adding another stage of sampling, only school systems that are selected would need to be engaged, with more than one school ultimately being selected from each school system. Again, this design is less statistically efficient.

If a selected school refuses to participate, a strategy must be in place to select another school. A simple approach is to select the next school on the systematic sampling list after the school that refused; if this list was sorted prior to sampling, this school will most likely be similar in some way to the school that refused to participate. More sophisticated methods are also available *(13)*. This protocol exists to avoid convenience sampling of a replacement school, which would be inappropriate for these surveys.

2.4.3. Stage three: sampling students

From each selected school, depending on the approach selected above, either a fixed number of students are sampled from the school(s) from each age category (if PPS sampling is used), or an entire classroom is selected from each grade. It is not necessary to conduct additional sampling, such as randomly sampling the classroom within each school, as it is assumed that clustering at this stage will be negligible.

Even if the first approach is adopted, for operational reasons, a country may prefer to sample a fixed number of students by grade, expecting that the age distribution will roughly balance in the analysis. Furthermore, this approach may better align with how a vaccination policy would be implemented, as it can be difficult to isolate a particular age. Eligible students would be sampled consecutively until the target sample size is achieved for each age/grade.

Countries should follow local regulations for engagement of human subjects, including obtaining proper informed consent and assent from participants and their guardians.

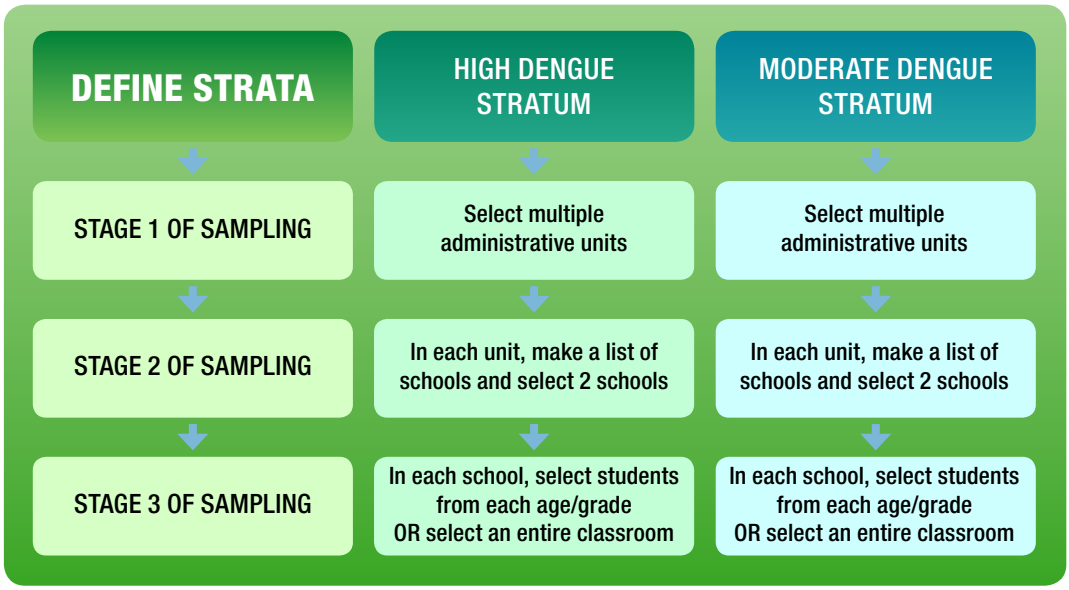
Exclusion criteria include children unwilling or unable to assent, receipt of blood or blood products within the last three months (as this may interfere with accurate assessment of antibody levels), medical contraindication to needle stick or blood draw, and children with a history of dengue vaccination. CYD-TDV is not approved for children less than 9 years of age.

Fig. 2 provides a generic flowchart for the stages of sampling.

2.5. Sample size

The sample size is the minimum total number of children needed for the serosurvey. The sample size required depends on the assumed mean seroprevalence of dengue, the design effect (a result of the sampling strategy used), and the desired precision (the width of the confidence interval around the estimated mean seroprevalence, e.g., ±6%). Because seroprevalence is estimated separately for each stratum, each stratum will have its own required sample size. Because seroprevalence is estimated separately for each age, each age will have its own required sample size within each stratum. These sample sizes may vary depending on the desired precision for that stratum and age.

Fig. 2: Generic sampling flowchart for country sampling only highest and middle dengue strata



The survey sample size depends on the *effective sample size* and the *design effect* (DEFF). The effective sample size is the number of children who would need to be sampled if you were conducting a simple random sample, which in turn depends on the desired level of precision. To achieve superior precision (narrower confidence interval), the effective sample size will need to be larger. Conservatively assuming that the age-specific seroprevalence of dengue is 50%, the effective sample size per year of age to achieve a confidence interval with width ±5% is 385, for CI ±6% is 267, and for CI ±7% is 196. Assuming a seroprevalence of 50% is conservative because, as the true seroprevalence approaches the extremes of 0% or 100%, the precision of the estimated seroprevalence will improve (narrower confidence interval).

The DEFF is an inflation factor that reflects how much larger the survey must be because it is not a simple random sample. For example, clustering of students within the same schools makes the survey statistically less efficient. If the survey's DEFF is equal to three, three times as many students are needed to achieve the same precision as a simple random sample. The effective sample size times the DEFF is called the *actual sample size*.

The survey DEFF depends on many factors, but most greatly upon the level of clustering in the population as quantified by the *intracluster/intraclass correlation coefficient* (ICC). The ICC measures the amount of variability between clusters relative to the overall variability. Dengue is highly spatially heterogeneous, as described in Section 2.3. Available literature and ad hoc calculations on school-based surveys in Thailand and Mexico suggest that clustering is very high, with an ICC of 0.10 to 0.15. While standard sample size calculations usually apply a DEFF of 2, the expected DEFF for this survey may be much higher (3–5) because of the high ICC.

To understand the relationship between survey design and sample size, below are a few general principles.

- As the desired precision increases (narrower confidence interval), the required sample size increases. For example, in younger age groups, smaller samples sizes might be needed since less precision may be required.
- As the ICC increases, the DEFF and required sample sizes increase. This is because when there is high clustering of outcomes, a clustered survey is not as efficient as a simple random sample that selects participants from throughout the population rather than from a few selected clusters.
- As the number of schools selected decreases, the DEFF and required sample size increases. This is because sampling more individuals from the same schools is not as valuable as sampling more schools. It is expected that students within the same schools (and from the same communities) have correlated outcomes, while students from different areas/schools provide more new information. In general, the number of schools should be selected so that no more than 20–30 students are selected per age in any given school. When more than 30 students are sampled per age per school, the DEFF can become very large, making the survey much less efficient.

Desired precision will likely be in the range of ±5 to ±10% for each stratum, though the exact level of precision required will be context-specific, depending upon the needs of vaccination policy makers and available resources. In general, greater precision is recommended for strata with more uncertainty about whether seroprevalence will exceed the recommended 70% threshold. Strata with seroprevalence very likely to exceed the threshold and strata with seroprevalence very likely to be below the threshold may be allocated lower precision to improve overall survey feasibility.

In countries considering a subnational vaccination policy, the greatest precision should be required from strata that are targeted for vaccination. If highest and middle strata are captured in the serosurvey, sufficient precision will be needed from both strata. If the highest stratum is expected to have high seroprevalence in the target age range, countries may prioritize resources in the middle stratum in which lower age-specific seroprevalence levels are expected. In middle dengue burden areas, the secondary objective of identifying local characteristics that are associated with high seroprevalence may also support decision-making.

If a country is considering a national vaccination policy, adequate precision will be required in all strata, though extra precision may be allocated to the lowest stratum where age-specific seroprevalence may or may not fall below recommended thresholds in the target age range.

Small countries with relatively few administrative units and/or a small eligible population may be able to adopt a more efficient survey design for finite populations, requiring a smaller overall sample size if all areas can be sampled rather than a subset.

Further details on sample size calculation are available in Appendix 2. Two worked examples of determining sample size using the spreadsheet are provided in Section 2.8. The required sample size is calculated for each age group and for each stratum, and

the overall sample size for the survey will sum across all ages and all strata. To minimize costs, strata not targeted for vaccination may be assigned lower precision or be excluded entirely. Similarly, there is an incentive to limit the sampling of older ages if these ages will not be considered for vaccination policy. On the other hand, if engaging the schools is significantly more resource-intensive than sampling additional students within the same school, it may be preferable to sample a broad age range within a school that is already being sampled.

It may be useful to contact a survey statistician or WHO for additional guidance on preparing sample size calculations.

2.6. Diagnostic assays and validation

Sera specimens sampled in a DENV serosurvey should be tested with indirect IgG enzyme-linked immunosorbent assay (ELISA) to assess presence or absence of dengue antibodies. Most commercially available diagnostic IgG ELISAs that are adjusted to measure past dengue exposure tend to have high sensitivity but suffer from low specificity due to cross-reactivity with other flaviviruses and certain flavivirus vaccines; nonetheless, these tests are still considered informative for targeting vaccination at a population level.

Anti-DENV IgG antibodies from infected individuals primarily target the viral structural gene products, particularly the envelope protein (E). Because the E protein is antigenically similar within viruses corresponding to the flavivirus family, anti-DENV IgG antibodies often display cross reactive epitopes, thus requiring neutralizing antibody confirmatory diagnostic testing to determine IgG antibody specificity. The PRNT is the most specific serological test for the determination of type-specific antibodies to an infecting virus and can be used to confirm the infecting flavivirus from a convalescent serum sample *(14)*. This biological assay is based on the specific interaction of virus and antibody in vitro which ultimately results in antibody-mediated inactivation of the virus such that it is no longer able to infect and replicate in cell culture. The PRNT result is expressed as the end-point titer of neutralizing antibodies from the serum to a specific virus and may be suggestive of the level of immune protection against the infecting virus. These two tests (ELISA and PRNT) provide complementary results since one test is a biochemical assay (ELISA) measuring antibody binding affinity to DENV antigen and the other is a biological assay (PRNT) measuring antibody neutralization capacity of an infecting virus. Since PRNT measures the neutralization capacity of the antibody response whereas the IgG ELISA is measuring total anti-DENV antibodies, the IgG ELISA is more sensitive and less specific since it also detects non-neutralizing and cross neutralizing antibodies to other flaviviruses within a serum specimen.

The presence of anti-DENV IgG antibodies is an indication of a long term acquired immunity resulting from a past flavivirus infection. This long term anti-DENV IgG immunity can be detected up to 60 years after the initial DENV infection *(15)*. There are two methods used to measure anti-DENV IgG; direct and indirect ELISA. The direct IgG ELISA method is generally less sensitive than the indirect ELISA and is less often used because it requires purified viral antigen. The indirect IgG ELISA is

more sensitive because it captures E antigen using a DENV E capture antibody to immobilize the viral antigen to a solid surface. This step is especially important since it captures and concentrates unpurified antigen which is often less immunogenic *(16)*. The individual's serum sample is added to the captured viral antigen and then detected with a secondary anti-human IgG-conjugated antibody. The DENV IgG ELISA provides either a qualitative or semi-quantitative result *(16, 17)*.

There are two types of commercially available DENV IgG ELISAs that were developed to either measure past or recent DENV infections, thus adjustments of the cut-off value for detection of DENV IgG were set according to these two criteria. Typically the ELISAs that measure recent dengue adjust the cut-off value at a high titer IgG reflecting those titers observed in most secondary dengue infections. Thus end-users should be cautioned to determine which type of commercial anti-DENV IgG ELISA are being used since the ones that measure recent dengue infections will not detect primary dengue infections. For example, in a study detecting past DENV exposure in US travelers by Marrero-Santos et al. *(18)*, the sensitivity and specificity of the Focus Diagnostic ELISA was 100% and 24%, respectively when compared to the PRNT. Interestingly this study population had previous YF vaccination, which yielded a false-positive rate of 52.8% when using the manufacturer's suggested cut-off value and was reduced to 6.7% when the cut-off value was adjusted accordingly. The results of this study indicated that the manufacturer's suggested index cut-off value for the ELISA caused substantial misclassification that increased the prevalence of prior DENV exposure among US travelers when compared to a more specific test, PRNT. This misclassification rate in dengue non-endemic regions may be due to the positive and negative predictive values (PPV and NPV) of the test, based on the DENV prevalence in the population. For example, a seroprevalence study conducted in Key West, Florida following a DENV1 outbreak in 2009 indicated that the anti-DENV IgG ELISA, falsely elevated the prevalence of prior dengue. This may be explained by the previous exposure of this population to other known circulating flaviviruses (St. Louis Encephalitis virus and/or West Nile virus (WNV)) or previous YF and/or JE vaccinations within this population *(19)*. Conversely, in a dengue endemic area such as Puerto Rico, the prevalence of DENV infection is 95%; hence, most of the anti-DENV IgG ELISA-positive results would also be PRNT-positive since DENV is the predominant circulating flavivirus in Puerto Rico and WNV transmission is only sporadically introduced and not maintained.

Other assays besides IgG ELISA may be considered for testing specimens, but they should only be used to replace IgG ELISA if they have been similarly validated against NTs and demonstrate acceptable sensitivity and specificity. Measurement of anti-DENV IgG in salivary samples may be an attractive option to simplify specimen collection procedures and/or increase survey participation, but available assays for salivary samples have poor sensitivity *(20)* and have not yet been sufficiently validated. Similarly, rapid tests are attractive because they can be done on site, but these tests have generally been evaluated in the context of detecting acute infections by non-structural protein 1 (NS1) and IgM antibodies, including differentiating primary from secondary acute infections, for which purpose they have highly variable sensitivity and specificity *(18, 21–23)*. Data on the performance of rapid tests for serological characterization of past exposures by IgG are lacking and their use for estimating seroprevalence is therefore discouraged at this time.

Country-specific contexts, including the presence of other circulating flaviviruses such as ZIKV, JE, YF, and WNV, and vaccination programs for JE and YF, must be considered when interpreting dengue seroprevalence results measured with DENV IgG ELISA. Co-circulation of other flaviviruses or expansive vaccination programs can result in substantial misclassification of prior DENV exposure. This may be especially important in areas with recent introduction of ZIKV as the extent of cross-reactivity with Zika in a dengue endemic population is not known. Countries are encouraged to prepare an assessment of relevant co-circulating viruses and vaccinations to assist in this interpretation. Nonetheless, surveillance data are likely to be unreliable for determining the relative proportions of these viruses because many flaviviruses have non-specific clinical presentations and high proportions of cases may present as asymptomatic.

It is strongly recommended that a subset of specimens are retested using neutralization tests (NTs), such as PRNTs. PRNTs and other neutralization assays measure neutralizing antibodies from serum to a specific virus, and they are more specific than ELISA because they do not detect the same cross-reactivity. The value of retesting is that the indirect IgG ELISA cutoffs can then be tailored to the local context, especially if the specificity is very poor with the standard cutoff. Local validation of IgG ELISA improves interpretability of results. Instructions on how to modify the cutoffs using the NT validation set are provided in Section 2.6.1. NTs are resource-intensive and require highly trained staff. Testing can be conducted after the specimen collection has been completed.

Pros of NTs:

- Can confirm results of anti-DENV IgG ELISA.
- Can locally evaluate the sensitivity and specificity of IgG ELISA, and, if necessary, modify the cutoff for seropositivity to improve assay accuracy.
- Determine probable prior flavivirus infections.
- Measures level of protective or cross-reactive neutralizing antibodies.
- These baseline measures may be important when assessing the efficacy of CYD-TDV post-implementation.

For countries seeking to identify a laboratory with capacity to conduct NTs, contact WHO for assistance.

2.6.1. Procedure to adjust assay cutoff

The decision to adjust a cutoff value of an IgG ELISA should be based on an analysis of sensitivities and specificities using a Receiver Operator Curve (ROC), comparing the IgG ELISA with a confirmatory NT in a subsample of appropriate size. The goal is to optimize the IgG ELISA cutoff for positivity to improve sensitivity and specificity of the test. The new cutoff is set by re-testing a subset of specimens with NT, assuming the neutralization test is a gold standard.

The size of the subsample may depend on feasibility and availability of NTs, but the following formulae are provided to guide the subsample size calculation. A random sample of 100 to 200 should be adequate in most settings. Ideally specimens should be representative of the larger participant population, randomly sampled from the collected specimens rather than conveniently sampled from a particular school.

Table 2: Sensitivity and specificity calculations

	PRNT positive	PRNT negative	
ELISA positive	True positive (TP)	False positive (FP)	Positive predictive value (PPV) of ELISA = TP / (TP + FP)
ELISA negative	False negative (FN)	True negative (TN)	Negative predictive value (NPV) of ELISA = TN / (TN + FN)
	Sensitivity of ELISA = TP / (TP + FN)	Specificity of ELISA = TN / (TN + FP)	

Sample size to evaluate cut-off values for ELISAs will depend on the desired minimum level of specificity (or sensitivity), a precision for this specificity (or sensitivity), and the assumed seroprevalence by NT. To establish a minimum level of specificity (or sensitivity) of 70% with a confidence interval no wider than ±10%, a formula for the sample size (n) is provided below.

$$n \geq (1.96)^2 \frac{(0.70)(1-0.70)}{(0.10)^2} = 80.7$$

In this example, a sample of at least 81 NT negative specimens (true negatives by NT) would be needed to establish a minimum level of specificity of 70% with confidence intervals within ±10%. To establish a higher level of specificity, a smaller sample size would be required because variability is smallest for proportions close to the extremes of 0% or 100%. Similar calculations to determine the minimum number of true positives by PRNT to establish sensitivity could be conducted, if desired.

If the true seroprevalence in the study population is estimated to be 50%, then to obtain at least n = 81 NT negatives, one will need to sample approximately 162 participants (= 81 / 0.50) individuals as the other 50% would be NT positive. If the true seroprevalence were 70%, then only 30% of specimens would be NT negative; in this case, a larger number of specimens (270 = 81 / 0.30) would be required to obtain n = 81 NT negative specimens.

Once a subsample has been tested with both NT and IgG ELISA, sensitivity and specificity of IgG ELISA can be calculated for a series of assay cutoffs using the formulae provided in Table 2. The sensitivity and 1-specificity can be plotted in an ROC curve to inform selection of the optimal cutoff. Traditionally the optimal cutoff is the point closest to the upper left-hand corner of the plot (or the point that maximizes

sensitivity plus specificity). If you lack experience with this type of calibration, it is advised that you consult an expert.

For an example of modifying a cutoff for a particular population, refer to Marrero-Santos et al. *(18)*. The study evaluated the use of anti-DENV indirect IgG ELISA for determining prior dengue exposure in U.S. travelers. The study included 591 pre-travel specimens from U.S. residents who had traveled to dengue-endemic countries. Specimens were tested with IgG ELISA and classified using the manufacturer's cut-off; specimens were re-tested with PRNT. Of 71 PRNT positive specimens, 71 were ELISA positive (sensitivity = 100%). Of 50 PRNT negative specimens, only 12 were ELISA negative (specificity = 24%). In this population of U.S. travelers, there were excessive false positive results using the manufacturer's ELISA cutoff; the authors noted that false positive rates were highest for individuals who had received YF and/or JE vaccination. The authors evaluated different ELISA cutoffs and identified a higher cutoff with superior specificity of 95.7% and corresponding sensitivity of 85.3%.

2.7. Sources of bias

There are a number of possible sources of bias in these serosurveys. Some are outlined below, and each should be considered within the country context. Efforts should be made to minimize their impact during the design, implementation, and analysis stages.

Selection bias could result if administrative units, schools, or children were excluded from the opportunity to enter the sample, or if the process used to select them was not truly random. For example, selecting schools that are easiest to sample (convenience sampling) or excluding private schools from the sampling frame could induce bias if there is a systematic difference in seroprevalence between the included and excluded schools.

Administrative units or schools that are excluded a priori for security or logistical reasons may affect the generalizability of the serosurvey results. This is important if these areas would ultimately be targeted for vaccination as part of the policy because the results would not be fully representative. If these areas would not be targeted for vaccination for the same security or logistical reasons that prevented their inclusion in the serosurvey, then the serosurvey results are at least representative for the population being vaccinated.

Non-response bias could result if some of the randomly selected schools either refused to respond or were not available to respond at the time of the survey. This type of bias would be incurred if schools willing to engage in the survey are systematically different from schools unwilling or unable to engage, in terms of exposure to dengue.

Information bias could result if a measuring process or instrument had a bias, for example if some of the lab assay kits or laboratories or technicians had a systematic bias. Similarly, misclassification (false positives) should be expected if there are co-circulating flaviviruses in the country. Retesting some samples with NTs can help to locally determine the sensitivity and specificity of IgG ELISA. NT validation is necessary so that the seroprevalence of dengue is not over-estimated.

2.8. Examples of sampling procedures

Below are two examples of hypothetical serosurvey designs.

BOX 2. Latin American country example

Country X in Latin America has 7 regions which are further subdivided into departments and then into municipalities. The Aedes mosquito vector is present in 800 municipalities. The distribution of dengue cases is highly heterogeneous, with only 20 municipalities capturing 50% of all reported dengue cases. These 20 municipalities are generally large cities and are not concentrated in any single region. Only 60 municipalities capture 70% of all reported dengue cases, with the remaining municipalities where the mosquito vector is present reporting very low numbers of cases of dengue.

The MOH identifies that it would like to use a highly targeted strategy to identify the age group of children suitable for vaccination living in some or all of these 60 municipalities. Surveillance data suggest that the remaining 740 municipalities have relatively lower risk of dengue, and they are not being considered for vaccination at this time. These 60 dengue burden municipalities are categorized into highest and middle dengue burden strata; the 20 highest risk municipalities are classified as highest, and the remaining 40 are classified as middle.

The target age range is 9 to 12 years old, and they will select a single age for their vaccination policy using serosurvey results. The target age will ideally align with routine HPV vaccination in 4th to 5th graders. The survey will also include 5 to 8 year olds because dengue transmission may be heterogeneous over time in some areas of the country. Lower precision is considered necessary for 5 to 8 year olds as compared to 9 to 12 year olds.

The list of highest dengue burden municipalities is sorted by region. Four municipalities comprised of large cities are so large that they would be automatically selected by the sampling procedure; each municipality is subdivided into smaller areas to prevent automatic selection. Ten highest dengue burden municipalities are sampled using probability proportional to size systematic sampling. The list of middle dengue burden municipalities is sorted by region, and thirteen middle dengue burden municipalities are sampled using the same procedure. From each selected municipality, a list of schools is generated, including private schools. Two schools are sampled from each selected municipality using probability proportional to size systematic sampling.

Schools are engaged, and consent is obtained from parents to take specimens from children aged 5 to 12. A fixed number of specimens are collected from each age at each school. A flowchart with sample size calculations is provided in Fig. 3. Parents are asked about child's history of dengue, Zika, YF vaccination, and socioeconomic status.

Country X does not conduct PRNT testing at the national laboratory. 150 specimens are randomly selected for validation testing. These are shipped to a country with testing capacity or a WHO-recommended laboratory. Using the PRNT results and IgG ELISA results from these specimens, it is determined that the specificity of the test is too low (<70%) using the standard cutoff in this population. Different cutoffs are considered and plotted in an ROC curve, and a slightly higher cutoff is selected that optimizes sensitivity and specificity. This cutoff is applied to classification of all specimens with IgG ELISA.

Limitations of this design include that some private schools may be unwilling to participate (non-response error). It is possible that some of the remaining 740 municipalities may have sufficiently high seroprevalence, though reporting may be poor. These areas will not be considered for vaccination.

BOX 3. South East Asian country example

Country Y in South East Asia has 5 regions subdivided into 65 provinces. Though there is spatial heterogeneity in seroprevalence, overall the seroprevalence of dengue is typically high in South East Asia due to hyper-endemic circulation of all four serotypes.

The MOH is considering a national vaccination policy for dengue, and believes that the mean seroprevalence in 9 year olds will exceed the 70% threshold. Nonetheless, recent seroprevalence studies conducted as research in Country Y suggest that seroprevalence may be declining in children. The target age range is 9 to 11 year olds. The country is also interested in improving dengue burden estimation, so they expand the age range to include 7 to 8 year olds and 12 to 13 year olds who attend the same schools as the target age range.

The MOH identifies province as the administrative unit of choice, since it will be easiest to operationalize at this level. Using available historical seroprevalence data and detailed surveillance data, they stratify provinces based on expected seroprevalence into highest, middle, and lowest strata. The strata are roughly equal in size. As the country desires a national vaccination policy, all provinces must be represented in the survey. The data will be analyzed in two ways; 1) all strata are considered together to provide a national estimate of age-specific seroprevalence; and 2) each stratum is analyzed separately, primarily to ensure that seroprevalence is sufficiently high in lowest and middle dengue burden strata.

The country desires the greatest precision from the lowest dengue burden stratum because it is unclear that seroprevalence will pass the threshold in these provinces. The country allows lower precision from the middle and highest dengue burden strata to reduce the sample size required. A flowchart with sample size calculations is provided in Fig. 4. The country selected 11 provinces from the lowest dengue burden stratum, 7 provinces from the middle dengue burden stratum, and 7 provinces from the highest dengue burden stratum. Before sampling, the capital city was subdivided into smaller areas to prevent automatic selection into the survey. In each selected province, a list (sampling frame) of schools is constructed. The sampling frame does not include private schools because of the expected difficulty in engaging these schools. The country is unable to determine the size of the school at the province level, so two schools are sampled from each selected province using equal probability sampling.

At each school selected, one classroom per grade is sampled to capture students aged 7 to 13 years old. Classroom size varies across schools. At some schools, the oldest students are 11 year olds, so a neighboring school is selected that enrolls students through age 13. Parents are asked about the child's history of dengue, JE, and JE vaccination. These data elements may be used to conduct a stratified analysis and inform the interpretation of the results.

The country has in-house capacity for PRNT testing. Anticipating poor IgG ELISA specificity due JE vaccination, 200 specimens are randomly selected for validation testing with PRNT.
As expected, it is necessary to select a higher cutoff to improve the specificity of the IgG ELISA, as determined by using an ROC curve analysis.

Limitations of this design include the exclusion of private schools.

Fig. 3: Sample size calculations example for Country X. Precision may vary across age groups to conserve resources where lower precision may be acceptable

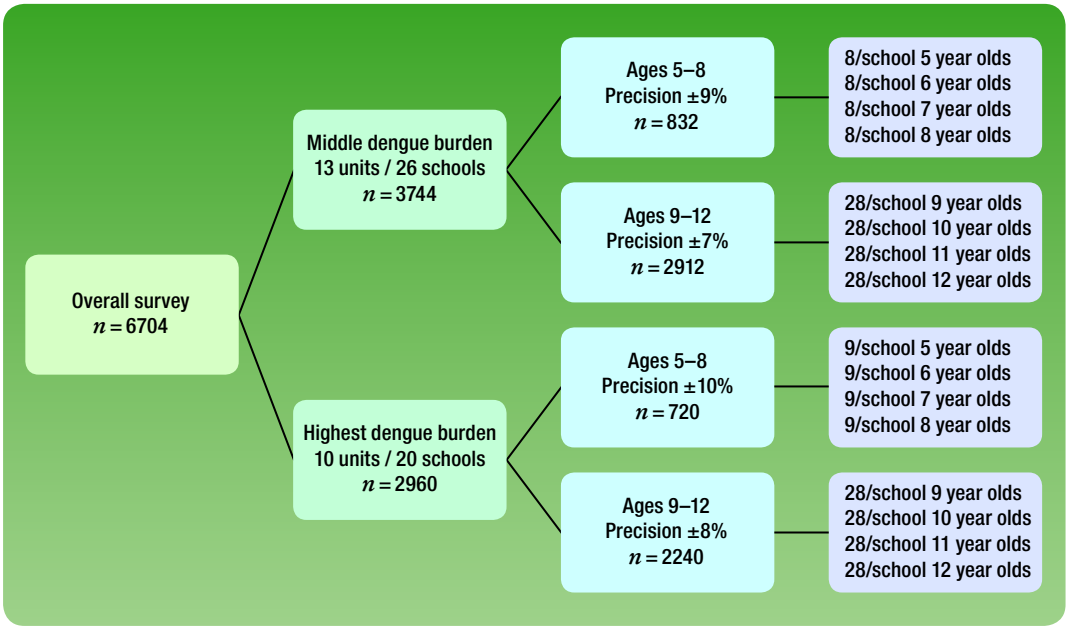
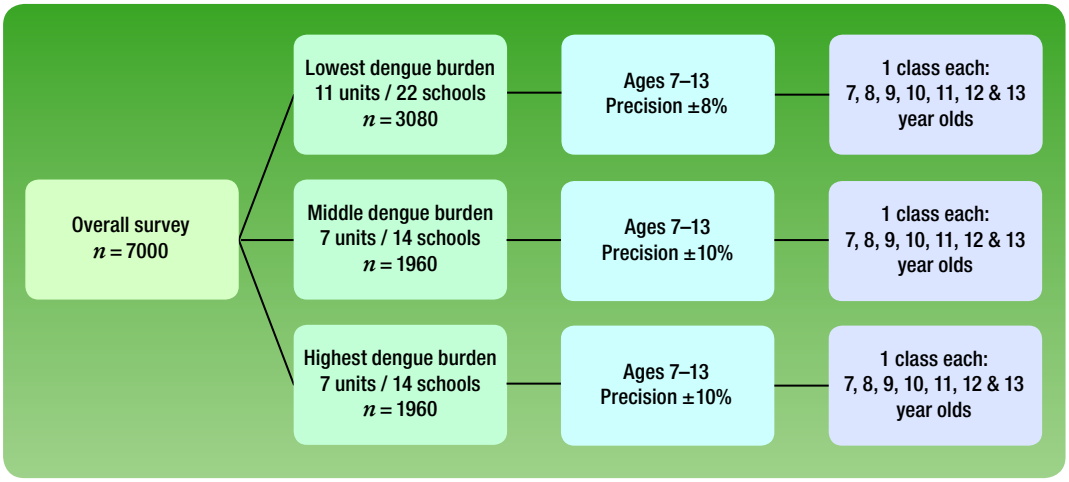


Fig. 4: Sample size calculations example for Country Y, assuming an average class size of 20 students



2.9. Regression modeling

The sample sizes above must be calculated independently for each age group and stratum because seroprevalence estimates are calculated independently for each age group and stratum. This is not the most efficient approach for analyzing the data. There is additional statistical information from students attending the same schools but of different ages, as their outcomes are expected to be correlated. Regression models can be applied that utilize all available information to improve precision in age-specific

seroprevalence *(24)*. Precision may be further improved if available surveillance and epidemiologic data are explanatory predictors and reduce residual heterogeneity; this could include reported cumulative incidence of dengue per administrative unit or the variables listed in Box 1.

If readily available surveillance and epidemiologic data are good predictors of age-specific seroprevalence, such a model could support countries making assessments of likely seroprevalence in areas not selected for the survey but where vaccination is desired. These model results also support validation of the use of routine surveillance data for dengue burden estimation by calibrating the model with local seroprevalence.

Regression modeling is presented as an optional analytical strategy for countries. It can be conducted in addition to standard approaches which estimate seroprevalence separately by age and stratum. More detail on regression modeling is provided in Section 4.4. Regression models may be especially useful if the precision is worse than anticipated, which can occur if there is more heterogeneity in seroprevalence than previously assumed. Modeling may improve precision around seroprevalence estimates and aid in interpretation. Even if countries do not initially plan to run models, they should collect basic demographic information about sites/population in case such models are later desired.

3. Field and laboratory procedures

3.1. Preparing for the field

Conducting such high-quality surveys means that appropriate expertise and resources need to be made available. For large serosurveys, an in-country assessment of available resources and personnel and laboratory capacity should be completed prior to finalizing the serosurvey planning. In some cases, conducting a small pilot study to assess the feasibility of a larger-scale serosurvey may be useful and informative.

3.2. Human resources

Once the decision has been made to conduct a serosurvey, a planning and implementation team should be assembled. Dengue serosurveys should be guided by epidemiology and laboratory scientists experienced in survey design, planning, implementation, training, data collection and analysis, specimen collection and laboratory procedures. Additional serosurvey personnel can include statisticians, staff responsible for survey participant enrolment and specimen collection, laboratory technicians, supervisory staff, data managers, coordinators and others as needed. The size and composition of the team will depend on the potential size and complexity of the serosurvey. Serosurveys can be lengthy, resource-intensive activities and adequate resources should be made available to ensure they are well-managed and closely coordinated.

The survey team can include a national team with external help or a national team alone. External help may include: verifying the study design; assistance with statistical issues including sampling size calculations; developing or reviewing the survey protocol; training field staff; piloting the protocol; providing guidance on laboratory issues; conducting laboratory analysis. WHO can assist with connecting national teams to external experts.

Training for field and laboratory staff needs to be provided prior to the start of data collection and should include pilot testing of the data collection tools. Pretesting, translating, back-translating of the questionnaires can provide clarity of what information needs to be collected and therefore, increase the data quality. Training should also include practice of the specimen collection standard operating procedures (SOPs), including the specimen labelling and transporting requirements.

3.3. Data collection

Generally, data in the field are collected using paper forms or hand-held electronic devices. To reduce data entry errors from paper-based collection, double entry of data from paper questionnaires is recommended. This allows for a crosscheck of the two electronic databases to identify discrepancies that can be resolved by going back to the paper forms to create a single final database that can be used for analysis. If data are collected electronically, then clear SOPs should be provided for the maintenance and care of the electronic devices as well as for transmitting the data to the study data management team.

Data collection must be systematic and closely supervised to limit potential bias. Survey instruments and data collection technologies should be piloted with sufficient time to make changes where there are ambiguities or errors. Stratum IDs, cluster IDs, and school IDs should be pre-printed on survey forms to avoid confusion due to legibility or mistakes. Specimen labels should be pre-printed and stuck on specimen tubes.

The specific types of data to be collected are outlined below.

3.3.1. Administrative unit-level surveillance and epidemiological data

To support categorization of administrative units into strata, and to support assessment of epidemiological predictors of seroprevalence informally or with regression modeling, administrative unit-level surveillance and epidemiological data should be collected and stored. Surveillance data includes measures of dengue disease burden, such as cumulative incidence of dengue in children per unit. Basic measures that can be used in a formal (regression-modeling) or informal (comparing group means) analysis of predictors of age-specific seroprevalence should also be collected. Some data, such as average temperature during the rainy season, can be collected by a simple web search. Suggestions for variables that could be collected and analyzed as predictors of seroprevalence are provided in Box 1.

3.3.2. Survey sampling data

All sampling frames used to conduct sampling should be retained. This data is important for conducting proper survey weighting. This includes sampling frames constructed at the stratum level with lists of administrative units, as well as sampling frames constructed at the administrative level with lists of schools. Sampling frames include both the names of the units/schools, their respective sizes, if PPS sampling was used, and details about the systematic sampling procedure used to conduct the sampling. Clear descriptions of administrative units or schools that were excluded a priori for logistical or security issues should be retained. Sufficiently detailed records should be kept to note schools that were sampled but were excluded for other reasons.

3.3.3. School data

From each sampled school, the survey team should collect the sizes of schools, whether the school is public or private, where the school is located, and the grade range of school. If entire classrooms are sampled, information should be collected about the number of classrooms per grade.

3.3.4. Questionnaire

Prior to collecting specimens from students, a questionnaire should be provided to parents along with the appropriate consent form. The following elements are recommended for inclusion in the questionnaire.

For all study participants, name of father and/or mother, sex, date of birth, grade of child, and school ID should be collected. A unique identification number should be given to each participant in the study and used to match questionnaire data with the laboratory specimen.

Countries may decide what other variables to include on the questionnaire that are appropriate for the local context. Options include self-reported dengue in the child (including severity of disease), socioeconomic information, and whether the child has received dengue, JE, or YF vaccine before. For the JE vaccine, this could include vaccine type and number and timing of doses, if this information is available. Information may also be collected on area of residence (home address, location of home village, or neighborhood) and number of years child has lived at this residence. If available, information may also be collected on household-level and neighborhood-level risk factors, including larval and household indices, levels of screening, and information on local vector control.

Completed forms should be checked for accuracy, legibility, and completeness and signed by team supervisors. This quality control check ideally will be done before the team leaves the school to allow for corrections and completions by the team. Records should be kept regarding the number of consent forms that are not returned and the number of consent forms that are returned with consent denied.

3.3.5. Laboratory data

A unique identification number should be given to each participant in the study and used to label their blood sample and link to their other data.

For each specimen, information should be retained on the date of blood draw, the date ELISA is run (in case the sample degrades in quality), and the raw optical density (OD) value of the test. If the specimen is selected for additional testing with NT, information should be retained on the date of NT testing and raw test results.

3.4. Specimen collection and processing

Trained field investigators will collect the minimum amount (e.g. 2–5 mL depending on age) of whole blood necessary for IgG and NT assays from the subjects in a tiger top or red top vacutainer. Blood is collected by venipuncture and the serum is separated from the cellular components by centrifugation or by use of serum separator tubes. Blood specimens must be processed within 24–48 hours of collection. If the serum cannot be separated at the collection site, whole blood must be kept at 4 °C to 8 °C and shipped to the central laboratory within 24 to 48 hours. Serum should be transferred to externally threaded screw top plastic, pre-labelled cryovials and the serum specimens can be stored at –20 °C. Once frozen, serum should be shipped on dry ice.

Blood specimens must be labelled with the subject identification number, name, date of sample collection, and date of birth. The use of several identifiers in this manner acts as a fail-safe in the event that one is not read or transcribed clearly. Labels should be affixed to the collection tubes during sampling collection. Specimens should be stored in a cold box with four frozen ice packs immediately after collection.

The laboratory should consider pre-labelling tubes or supplying pre-printed adhesive labels with a unique identifier for each specimen. Hand written labels are strongly discouraged. The laboratory should supply an Excel spreadsheet with the assigned specimen numbers so that a line list can be prepared at the collection site.

3.5. Laboratory testing algorithm

Qualitative ELISA classifies each specimen as positive, negative or equivocal based on a predetermined or calculated cutoff value for OD. Some ELISA have been calibrated to give a quantitative result and the OD is converted into international units (IUs) or milli-international units (mIU), though quantitative results can be similarly classified as positive, negative, or equivocal based on a cutoff.

It is recommended that the appropriate cutoff value be calculated using results from a subset of specimens in the survey that are re-tested with NT, as per Section 2.6. A sample size should be determined as described in Section 2.6.1, and an appropriate cutoff value consistent with NT is obtained. Once the test is properly calibrated for the population then the IgG ELISA alone can be used for classifying all specimens in the survey without further testing using NT. Though IgG ELISA testing may be performed at multiple laboratories, the validation procedure would be performed once, likely at a central laboratory, using specimens collected throughout the country.

Summary of procedures for IgG ELISA calibration:

1) Collect sera specimens from the age-stratified serosurvey population.
2) Determine sample size of study for evaluation of IgG ELISA cutoff.
3) Test this subset with PRNT with 90% reduction.
4) Measure sensitivity and specificity of different IgG ELISA cutoffs and create ROC curve.
5) Adjust cutoff value of IgG ELISA based on results of ROC.
6) Apply adjusted IgG ELISA cutoff values when classifying all specimens in the survey.

The PRNT has been used as a confirmatory assay when cross reactivity is observed in the standard antibody binding assays (IgM and IgG ELISA). The challenge of this assay is the requirement for standardization of all reagents and between all technical staff performing the assay. In an attempt to standardize the PRNT worldwide, subject matter experts from around the world published guidelines for PRNT standardization *(25, 26)*. It will likely not be practical for most countries to establish PRNT (or other NT assay) capability anew for the purposes of performing DENV

serosurveys; therefore, it will likely be necessary to partner with other laboratories in the region for NT testing. Laboratories must standardize the plaque reduction to 90% for application to DENV serosurveys; using 90% reduction in plaques provides high specificity of the test conversely the test becomes less sensitive.

If a specimen yields equivocal or indeterminate results initially and after subsequent testing, it is recommended that equivocal results are treated as a third category, distinct from seropositive and seronegative results.

The laboratory SOP must include procedures for collecting, storing, analyzing, and reporting of the data obtained from the serologic assays. Laboratories need to be fully aware that data management requirements for routine testing of a small number of diagnostic samples per week are very different to requirements for testing a large number of serosurvey samples. All OD readings together with the interpretation must be saved for each specimen and entered into the data set.

3.6. Ethical considerations

As serosurveys involve the collection of biological specimens and possible storage of specimens for future testing, they require review and clearance by one or more ethical review committees. Local and national ethical requirements need to be adhered to, and partner and funding agencies often require separate ethical clearances.

All randomly selected children who meet the eligibility criteria will be invited to participate. Written parental consent must be obtained from the parents of all participants where required in accordance with local and national ethical guidelines. Informed assent may also be required from older children in some countries. The forms for informed consent and assent should provide the expected benefits and potential risks of participation and the procedures in place to maintain data confidentiality, clarify the right to non-participation without fear of reprisals, and provide contact information for the Ministry of Health and/or the principal investigator of the survey.

It will be necessary to decide how blood specimens will be treated after the survey. Options include storing samples for future use or discarding them. Generally, serosurveys are designed a priori with objectives and methods defined prior to the start of data and specimen collection. However, sometimes archived specimens can provide an opportunity to conduct a serosurvey using existing specimens from previous studies. If considering using archived specimens, it is critical to ensure that the original protocol and consent forms included the possibility for additional future testing. Before conducting a serosurvey using archived specimens, an assessment and inventory should be completed to confirm the completeness of the original sample set, legible labeling, and quality of the stored specimens including the history of specimen storage maintained according to the specific cold chain requirements. In addition to confirming the integrity of the archived specimens, it is critical to have a thorough understanding of the initial study protocol, including the sampling methods that were used to collect the archived specimens.

4. Survey analysis

Data analysis must be appropriate and statistically valid. In this section a general statistical analysis plan is outlined, as well as recommendations for reporting and disseminating results. This includes guidance on interpreting results to inform national dengue vaccination policy.

4.1. Descriptive epidemiology

It is recommended to start with basic reporting of the dataset. A good summary would include the following:

1) Summarize the demographics of the survey sample
    a) Number of administrative units, schools, students participating in survey
    b) Include detail by stratum, by sex, and by age group
    c) Numbers of individuals refusing participation

2) Summarize the laboratory analysis
    a) How many specimens were collected at which schools?
    b) How many specimens were analyzed at which labs?
    c) How many runs had indications of invalid results and had to be re-analyzed?
    d) What is the proportion of inadequate or equivocals?
    e) What portion were re-tested with NT?
    f) How many NT positives, negatives, and indeterminates were there?
    g) Report the estimated IgG ELISA sensitivity and specificity (with confidence intervals) at each of a series of cutoff values.
    h) Which cutoff was selected as optimal?
    i) How many IgG ELISA positives, negatives, and indeterminates were there using this cutoff?

3) Summarize the analysis dataset
    a) Overall Summary. Briefly describes the study, sources of the data, the time period and manner in which it was collected and contact information for the client in case eventual codebook readers have detailed questions.

b) List of variables. A simple uncluttered list of the variable names and labels for quick reading and electronic parsing.

c) Full Dataset Summary. Summarize each variable in the dataset, documenting variable name, label, type, length, and then summarizing the variable in one of several fixed formats:

   - For categorical variables, a frequency table with data values, formatted labels, and a count of the number and percent of observations that take on that value in the dataset
   - For continuous variables, a univariate summary including minimum, maximum, median, mean, standard deviation, standard error, the number of observations that are missing, or that use special missing values (e.g., refused, don't know, questionnaire item skipped appropriately)
   - For dates, an indication of the first and last dates in the dataset (to detect outliers)
   - For open-ended questions, one can either list the variable and the number of missing and non-missing responses, or it can document every unique verbatim answer in the dataset (often in a separate section for each open-ended response)

d) Notes. Provides any helpful information about the dataset including special documentation of data quality flags, problematic periods of data collection, formulae for calculating derived variables, known problems with individual variables or citations to literature that describes derived variables and validated scales or scores calculated from raw survey responses.

4.2. Survey weights and design specification

To make appropriate population level estimates of mean seroprevalence and to estimate meaningful seroprevalence confidence intervals, it will be necessary to use estimation methods that incorporate *survey weights* and that account properly for the complex nature of the survey sample, including stratification and clustering. It is inappropriate to use standard methods to estimate the mean and variance because these methods assume that all observations are independent. In complex surveys, correlation of outcomes between students in the same schools or schools within the same administrative units must be accounted for when the variance is estimated. Furthermore, not all observations may be weighted equally in a complex survey, meaning that mean seroprevalence may not be equal to the simple proportion of seropositive specimens. Additional details about survey weight calculation and analysis are provided in Appendix 2, Section 2.2.

Use software that accounts properly for the complex survey sample and that incorporates the weights into the calculations. Several modern software packages handle these calculations correctly. Examples include EpiInfo, Stata, R, SAS, and SPSS. The estimation should be conducted using a set of commands saved in a program file. For example, the SVY commands in Stata include an option to use interactive menu-driven

options to specify the survey design; if this user-friendly option is used, the Stata commands that it generates should be copied into a do file and retained. Saving the analysis program will facilitate later modifications and independent review to reproduce and verify the results.

4.3. Seroprevalence estimates

After the dataset has been described and checked, the next step is to estimate stratum- and age-specific seroprevalence and other population-level parameters.

Estimate the quantities described in the analysis plan, including age-specific dengue seroprevalence for each age group in each stratum, and if appropriate, for all strata combined. Equivocal or indeterminate results should be treated as missing data for the purposes of estimating seroprevalence. Each estimated quantity will yield a point estimate and a 95% confidence interval.[2] Summarize these quantities in output tables and figures. For example, for each stratum, countries may plot seroprevalence by age with error bars. It is advisable to also have the tables and figures generated by a saved set of commands rather than have an analyst copy results from one window by hand and paste them into the table. With the saved commands, the tables are likely to be reproducible and not have copy/paste errors.

Countries may undertake an informal analysis to identify administrative unit-level predictors of seroprevalence, such as surveillance or weather data. To assess the impact of certain predictors, countries can categorize units into groups and compare seroprevalence means across groups. For example, countries may compare seroprevalence across urban and rural administrative units within each stratum to look for systematic predictors of seroprevalence that could inform vaccination policy.

Countries may also examine the impact of individual-level predictors, such as self-reported history of dengue or prior flavivirus vaccination, where data are collected.

4.4. Regression modeling (optional)

Regression modeling can be used to improve precision in estimated age-specific seroprevalence and to identify predictors of seroprevalence, such as metrics derived from surveillance data. The model would likely take the form of a logistic regression model where the outcome is yes/no seropositive; indeterminate results would be treated as a missing data category. The model must properly account for survey weighting, multistage clustering, and stratification *(27)*, which can be achieved with the survey functions in many common statistical software packages.

The simplest model would include a categorical variable for age. Such a model, if implemented properly with the correct correlation structure, would improve precision of

2 Note that the confidence interval for a proportion will be symmetric only when the point estimate is near 50% but will become more and more skewed as the point estimate approaches either 0% or 100% – a skewed asymmetric confidence interval is appropriate for an estimated proportion, with the longer side or tail of the distribution occurring on the side of the interval nearest 50%.

age-specific seroprevalence by exploiting dependence between participants of different ages but from the same schools. A more advanced model could add an available measure of dengue burden as a predictor, as well as any relevant epidemiological measures; examples provided in Box 1 include average temperature during the rainy season and population density. The fitted model can be used to identify the best predictors of seroprevalence and the characteristics of areas that should be targeted for vaccination. Separate models with different numbers and combinations of potential explanatory covariates should be fit to available seroprevalence datasets. The model with the lowest Akaike information criterion (AIC) should be selected for the final extrapolation of predictions to other populations. Further prospective cross-validation can be used to examine the reliability of the model for predicting seroprevalence in areas not captured by the survey.

Other types of models besides logistic regression could be applied. Catalytic epidemic models are well-suited for modeling seroprevalence as a function of age *(28)*. The drawback of these models is that they must be fit properly accounting for the complex structure of the survey design, so software may not be readily available.

4.5. Force of infection estimation (optional)

Estimating FOI may be a desired secondary objective for some countries, and FOI is estimable from age-stratified cross-sectional serosurveys. FOI is a measure of transmission intensity *(29–31)*. Due to the epidemic nature of dengue transmission and sampling variability, age-stratified serosurveys can show considerable fluctuations with age. Therefore the FOI estimated from such surveys across age groups provides an average annual transmission intensity, which gives a better indication of general trends. Models can be examined to assess if transmission is likely to be epidemic or endemic. If the model fit is poor, then the area is likely not in endemic equilibrium.

FOI models can provide estimates of the expected seroprevalence at a specific age, or can be used to estimate which age will reach the seroprevalence target of 70% in the population.[3] Both results can be directly derived from the FOI model.

The serosurvey sample sizes should be sufficiently large to achieve desirable precision, assuming that the age range of sampled children is adequately wide (e.g. the full school age range of 5–18 years). If the age range is narrow, i.e. restricted to the target age range, then the data may not support precise estimation of FOI; results may have unsatisfactorily wide confidence intervals.

Limitations include that FOI can only be estimated for the period corresponding to when the population was born and conceivably first exposed (e.g., 5 to 14 years ago for a survey capturing these ages). If the minimum age is 5, data cannot be used to estimate changes in FOI within the last 5 years. Furthermore, to derive proper inference, estimation must account for the complex structure of the survey design, including clustering, stratification, and survey weights.

3 For an example, see the Global Dengue Transmission Map: *https://mrcdata.dide.ic.ac.uk/_dengue/dengue.php*

4.6. Interpretation, report writing, and dissemination

Presentation of serosurvey results should present a very clear description of sampling methods, as well as of laboratory assays and the criteria for immunity used, such as cut-off values for seropositivity. The report should provide background on dengue transmission within the country. It is recommended that reports follow the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) criteria *(32)*; these criteria specify elements that should be transparently reported defining the objectives and funding, data inputs, data analysis, and results and discussion. Authorities should make the results publicly available wherever possible. Results should be disseminated/shared with communities that participate in the survey. The report should adequately convey the limitations of the survey, the potential sources of bias, and the efforts used to minimize these sources of bias.

It is important to remember that seroprevalence estimates will be associated with a 95% confidence interval, and therefore the use of point-estimates of seroprevalence should be interpreted considering this uncertainty. Interpretation may depend on other factors, such as the presence of other circulating flaviviruses. The estimated sensitivity and specificity of the IgG ELISA at the selected cutoff should be reported where NT validation is conducted.

Due to the seroprevalence thresholds, the WHO recommendations imply that the optimal age to target for vaccination will vary based on the transmission setting. For places where vaccination is suitable, the minimum age needs to be selected. This is a complex decision, based on many context-specific factors. The survey methods outlined in this report are designed to align with the seroprevalence recommendations for the introduction of the CYD-TDV vaccine, as outlined by WHO. Though the survey will likely not provide enough information to make precise assessments within each administrative unit, the survey is designed to achieve sufficient precision within each stratum. If significant variability across administrative units exists within each stratum, countries may undertake informal or formal analyses to identify predictors of seroprevalence that may explain some of this variability and help guide vaccination policy. It is possible that seroprevalence data could be used to locally validate surveillance data for identifying highest dengue burden areas; if achievable, future dengue serosurveys could be smaller or more targeted. Decisions about vaccine introduction in many areas requires triangulation of different data inputs and local factors, including the seroprevalence survey results, and the ultimate decisions and strategies across countries are likely to differ.

Serosurvey results are cross-sectional, which means that the results are representative of a particular point in time. The relevance of a historical serosurvey must be considered in the context of whether surveillance data reflect stable transmission or highly variably transmission over the recent years. For example, low seroprevalence in children 5 to 8 years with a high seroprevalence in children 9 years old may suggest that the survey should be repeated frequently to assess changes in transmission dynamics.

References

1) Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496:504–7. doi:10.1038/nature12060.

2) Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. Lancet Infec Dis. 2016;16:712–23. doi:10.1016/S1473-3099(16)00026-8.

3) WHO. Dengue vaccine: WHO position paper – July 2016. Wkly Epidemiol Rec. 2016;91:349–64. doi:10.1016/j.actatropica.2012.04.013.

4) Flasche S, Jit M, Rodriguez-Barraquer I, Coudeville L, Recker M, Koelle K, et al. The Long-Term Safety, Public Health Impact, and Cost-Effectiveness of Routine Vaccination with a Recombinant, Live-Attenuated Dengue Vaccine (Dengvaxia): A Model Comparison Study. PLoS Med. 2016;13:1–19. doi:10.1371/journal.pmed.1002181.

5) Undurraga EA, Halasa YA, Shepard DS. Use of Expansion Factors to Estimate the Burden of Dengue in Southeast Asia: A Systematic Analysis. PLoS Negl Trop Dis. 2013;7:e2056. doi:10.1371/journal.pntd.0002056.

6) Limkittikul K, Brett J, L'Azou M. Epidemiological Trends of Dengue Disease in Thailand (2000–2011): A Systematic Literature Review. PLoS Negl Trop Dis. 2014;8:e3241. doi:10.1371/journal.pntd.0003241.

7) Restrepo AC, Baker P, Clements ACA. National spatial and temporal patterns of notified dengue cases, Colombia 2007–2010. Trop Med Int Health. 2014;19:863–71. doi:10.1111/tmi.12325.

8) Koopman JS, Prevots DR, Marin MAV, Dantes HG, Aquino MLZ, Longini IM, et al. Determinants and Predictors of Dengue Infection in Mexico. Am J Epidemiol. 1991;133:1168–78.

9) World Health Organization and the Special Programme for Research and Training in Tropical Diseases. Dengue: guidelines for diagnosis, treatment, prevention and control – New edition; 2009.

10) World Health Organization. Immunization coverage cluster survey – Reference manual; 2005.

11) World Health Organization. World Health Organization Vaccination Coverage Cluster Surveys: Reference Manual; 2015.

12) World Health Organization. Steps in applying Probability Proportional to Size (PPS) and calculating Basic Probability Weights n.d. *http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/prevalence_survey/psws_probability_prop_size_bierrenbach.pdf* (accessed January 15, 2017).

13) ASTDD. Guidance on Selecting a Sample for a School-Based Oral Health Survey; 2013.

14) Calisher C, Karabatsos N, Dalrymple JM, Shope RE, Porterfield JS, Westaway EG, et al. Antigenic Relationships between Flaviviruses as Determined by Cross-neutralization Tests with Polyclonal Antisera. J Gen Virol. 1989;70:37–43.

15) Imrie A, Meeks J, Gurary A, Suhkbaater M, Truong TT, Cropp CB, et al. Antibody to Dengue 1 Detected More than 60 Years after Infection. Viral Immunol. 2007;20:672–5.

16) Johnson AJ, Martin DA, Karabatsos N, Roehrig JT. Detection of Anti-Arboviral Immunoglobulin G by Using a Monoclonal Antibody-Based Capture Enzyme-Linked Immunosorbent Assay. J Clin Microbiol. 2000;38:1827–31.

17) Miagostovich M, Nogueira R, dos Santos F, Schatzmayr H, Araujo E, Vorndam V. Evaluation of an IgG enzyme-linked immunosorbet assay for dengue diagnosis. J Clin Virol. 1999;14:183–9.

18) Marrero-Santos KM, Beltran M, Carrion-Lebron J, Sanchez-Vegas C, Hamer DH, Barnett ED, et al. Optimization of the Cutoff Value for a Commerial Anti-Dengue Virus IgG Immunoassay. Clin Vaccine Immunol. 2013;20:358–62.

19) Radke EG, Gregory CJ, Kintziger KW, Sauber-Schatz EK, Hunsperger EA, Gallagher GR, et al. Dengue Outbreak in Key West, Florida, USA, 2009. Emerg Infect Dis. 2012;18:135–7.

20) Andries A-C, Duong V, Ong S, Ros S, Sakuntabhai A, Horwood P, et al. Evaluation of the performances of six commercial kits designed for dengue NS1 and anti-dengue IgM, IgG and IgA detection in urine and saliva clinical specimens. BMC Infect Dis. 2016;16:201.

21) Fry SR, Meyer M, Semple MG, Simmons CP, Sekaran SD, Huang JX, et al. The diagnostic sensitivity of Dengue Rapid test assays is significantly enhanced by using a combined Antigen and Antibody testing approach. PLoS Negl Trop Dis. 2011;5:1–8. doi:10.1371/journal.pntd.0001199.

22) Peeling RW, Artsob H, Pelegrino JL, Buchy P, Cardosa MJ, Devi S, et al. Evaluation of diagnostic tests: dengue. Nat Rev Microbiol. 2010;8:S30–7. doi:10.1038/nrmicro2459.

23) Hunsperger EA, Yoksan S, Buchy P, Nguyen VC, Sekaran SD, Enria DA, et al. Evaluation of Commercially Available Diagnostic Tests for the Detection of Dengue Virus NS1 Antigen and Anti-Dengue Virus IgM Antibody. PLoS Negl Trop Dis. 2014;8:e3171. doi:10.1371/journal.pntd.0003171.

24) Ghosh M, Rao JNK. Small Area Estimation: An Appraisal. Stat Sci. 1994;9:55–93.

25) Roehrig JT, Hombach J, Barrett AD. Guidelines for Plaque-Reduction Neutralization Testing of Human Antibodies to Dengue Viruses. Viral Immunol. 2008;21:123–32.

26) World Health Organization. Guidelines for plaque neutralization testing of human antibodies to dengue viruses; 2007.

27) Lohr SL. Sampling: Design and Analysis. 2nd ed. Brooks/Cole, Cengage Learning; 2010; doi:10.1017/CBO9781107415324.004.

28) Muench H. Catalytic Models in Epidemiology. Cambridge, MA: Harvard University Press, 1959.

29) Imai N, Dorigatti I, Cauchemez S, Ferguson NM. Estimating Dengue Transmission Intensity from Sero-Prevalence Surveys in Multiple Countries. 2015;1–19. doi:10.1371/journal.pntd.0003719.

30) Rodriguez-Barraquer I, Buathong R, Iamsirithaworn S, Nisalak A, Lessler J, Jarman RG, et al. Revisiting Rayong: Shifting seroprofiles of dengue in Thailand and their implications for transmission and control. Am J Epidemiol. 2014;179:353–60. doi:10.1093/aje/kwt256.

31) Rodríguez-Barraquer I, Solomon SS, Kuganantham P, Srikrishnan AK, Vasudevan CK, Iqbal SH, et al. The Hidden Burden of Dengue and Chikungunya in Chennai, India. PLoS Negl Trop Dis. 2015;9:e0003906. doi:10.1371/journal.pntd.0003906.

32) Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: The GATHER statement. Lancet. 2016;388:19–23. doi:10.1016/S0140-6736(16)30388-9.

Appendix 1
Glossary of terms

actual sample size	The required sample size for the survey. It is equal to the effective sample size (ESS) multiplied by the design effect (DEFF).
cluster	The population is aggregated into clusters, such as administrative units, or schools. When cluster sampling is used, a random sample of these clusters must be selected for the survey. Clusters that were not sampled are not included. It is expected that there will be a dependency between participants in the same cluster, such as students within the same school. This tends to decrease the precision of the survey and increase the required sample size.
design effect (DEFF)	A measure of variability due to selecting survey subjects by any method other than simple random sampling. It is defined as the ratio of the variance with the chosen type of sampling to the variance that would have been achieved with the same sample size and simple random sampling. Usually, cluster surveys have a design effect greater than one, meaning the variability is higher than for simple random sampling.
effective sample size (ESS)	The effective sample size is the number of simple random sample respondents that would yield the same magnitude of uncertainty as that achieved in the complex sample survey. When a survey uses a complex sampling design (stratified or clustered, or both stratified and clustered), the magnitude of sampling variability associated with its results (that is, the width of the 95% confidence interval) is usually different than the magnitude that would have been achieved with a simple random sample using the same number of respondents. The effective sample size is the complex survey sample size divided by the design effect.
force of infection (FOI)	The rate at which susceptible individuals acquire an infectious disease.
implicit stratification	A method in which the sampling frame is ordered on some factor, like region, prior to systematic sampling. This improves the likelihood that the random sample will be balanced with respect to this factor.
intracluster/ intraclass correlation coefficient (ICC)	A measure of within-cluster correlation of survey responses. In most survey outcomes of interest, ICC varies from 0 to 1. The ICC is an important component of the survey design effect (DEFF). Smaller values of ICC yield smaller values of DEFF and vice versa.

multistage	There are multiple stages of sampling. Dengue serosurveys following this guidance are expected to have at least three stages: (1) sample administrative units, (2) sample schools within selected units, and (3) sample students within selected schools.
PPS sampling	Clusters are sampled with probability proportional to size (PPS), meaning that if Cluster A is twice the size of Cluster B, Cluster A has twice the probability of being sampled as compared to Cluster B.
precision	Refers to the variability (standard error) of the estimate of mean seroprevalence, as measured by the half-width of the 95% confidence interval. If the estimated seroprevalence is 80% ±7%, the precision is ±7%. When precision is increased, the confidence intervals are narrower. Increased precision requires greater sample sizes.
sampling frame	The list of clusters that is used to conduct sampling. This could be a list of administrative units within a stratum, or a list of schools within an administrative unit. If PPS sampling is used, the sampling frame must also include the relative sizes of each of the clusters.
seropositivity	The presence in serum of an antibody above a predefined cutoff of a specific infectious disease or pathogen.
seroprevalence	Percentage of a population positive for a specific antigen or antibody.
serosurvey	The collection and testing of specimens from a defined population over a specified period of time to estimate the prevalence of antibody levels above a predefined cutoff against a given etiologic agent as a direct measure of immunity.
serosurveillance	Serosurveys conducted routinely, periodically or through ongoing collection and testing of specimens to assess seroprevalence.
stratification (or explicit stratification)	The population is divided into non-overlapping groups (strata). Sampling is conducted separately in each stratum. The mean in each stratum can be estimated, or the means can be combined to derive an overall population estimate. The stratum-specific sample size is defined to achieve desired precision within each stratum.
survey weight	A value that indicates how much each record or case will count in the statistical analysis. Each record in a survey dataset might be accompanied by one or more survey weights, to indicate how many population level eligible respondents are represented by the respondent in the sample.
systematic sampling	A method for conducting random sampling in which the sampling frame is prepared as a list. A random number is selected to start the sampling, and administrative units are selected evenly spaced throughout the list. Systematic sampling can be conducted with probability proportional to size or with equal probability sampling.

Appendix 2
Statistical appendix

2.1. Sample size calculation

Many resources are available online to guide sample size calculation for clustered surveys, including Appendix B1 in *WHO Vaccination Coverage Cluster Surveys: Reference Manual (11)*. Though this resource is for household surveys, the general steps are very similar. Provided below are a few tables and formulae specific to the dengue serosurvey setting for additional guidance.

The first step is to calculate the effective sample size (ESS). The effective sample size is a function of the desired confidence interval half-width L and the assumed mean seroprevalence of 50%. The general formula is:

$$ESS = \frac{1.96^2 \times 0.50^2}{L^2} = \frac{0.9604}{L^2}$$

Table 3 contains ESS for a range of desired confidence interval half-widths. (Note: rounding up is always recommended when calculating sample size).

Table 3: Effective sample size

L	ESS
±5%	385
±6%	267
±7%	196
±8%	151
±9%	119
±10%	97

Once the effective sample size is calculated, it has to be adjusted to reflect the *design effect*. The design effect for clustering takes the following basic form, where m is the number of students sampled per age from a given school and ICC is the intracluster/intraclass correlation coefficient:

$$DEFF = 1 + (m - 1) \times ICC$$

This gives us a naïve estimate of DEFF as it assumes a simple two-stage design, where in fact three stages of sampling were used. This is then an approximation for the sample size.

Table 4 contains DEFF for a range of numbers of student per school and three possible values of ICC.

The actual sample size for a particular age and stratum is determined by multiplying the effective sample size and the design effect. For example, if precision of ±6% is required in 9-year-olds from the middle burden stratum, and 20 9-year-olds are sampled from each school, conservatively assuming an ICC of 0.15, then 1,028 9-year olds are required. The math is included below.

$$267 \times 3.85 = 1{,}027.95$$

As 20 9-year olds are sampled per school, this corresponds to roughly (1028 / 20=) 51 to 52 schools. If this is not feasible, the desired precision can be adjusted downwards, which will bring down the effective sample size, actual sample size, and required number of schools. Alternatively, the number of students per school could be increased; this would increase the actual sample size, but would drop the required number of schools.

A sample size calculator is provided to support this process. The user specifies the assumed seroprevalence 0.50 in Cell C3 and the assumed *ICC* in Cell C4. The user then can edit the green Cells C8 and below (C10, C12, etc.) to define the precision in each age category. Precision can be set to 0 if that age will not be captured by the survey. After completing all green cells, the user can examine the large table. Column E specifies the number of schools to be sampled in that stratum. Columns F through S then define the number of students required per school to achieve the desired precision as specified in Column C. If #N/A is returned, it means there is no survey design that can achieve the desired precision with so few schools.

This calculator corresponds to a design in which schools are sampled PPS, but an equal number of students per school are selected. If schools are selected with equal probability, then the calculator reflects roughly the average minimum number of students that need to be selected per school.

These sample sizes do not account for non-response as eligible students should be consecutively sampled until the target number is achieved. If the classroom strategy is used in which entire classrooms are selected, then the actual sample size should be divided by the assumed participation rate. These sample sizes may be further inflated to account for expected laboratory failure. If laboratory failure is expected to be minimal (<5%), it is not necessary to make further adjustments.

Table 4: Design effect

m	*ICC* = 0.10	*ICC* = 0.125	*ICC* = 0.15
5	1.4	1.500	1.60
6	1.5	1.625	1.75
7	1.6	1.750	1.90
8	1.7	1.875	2.05
9	1.8	2.000	2.20
10	1.9	2.125	2.35
11	2.0	2.250	2.50
12	2.1	2.375	2.65
13	2.2	2.500	2.80
14	2.3	2.625	2.95
15	2.4	2.750	3.10
16	2.5	2.875	3.25
17	2.6	3.000	3.40
18	2.7	3.125	3.55
19	2.8	3.250	3.70
20	2.9	3.375	3.85
21	3.0	3.500	4.00
22	3.1	3.625	4.15
23	3.2	3.750	4.30
24	3.3	3.875	4.45
25	3.4	4.000	4.60
26	3.5	4.125	4.75
27	3.6	4.250	4.90
28	3.7	4.375	5.05
29	3.8	4.500	5.20
30	3.9	4.625	5.35

2.2. Calculation of survey weights

Each individual in the survey is assigned a single survey weight. The survey weights are calculated as the inverse probability that the individual is selected into the survey. To calculate the survey weight for an individual within a particular administrative unit and attending a particular school, one must calculate the probability that this administrative unit was selected, that this school was selected, and that this student was selected.

Survey weights are composed by multiplying the inverse probability of being sampled at each stage of sampling. Furthermore, if a subset of classrooms are sampled from each school (e.g., 1 out of 5 classrooms in a grade), an additional inverse probability would be incorporated from this stage. There may also need to be weights to account for participant non-response or laboratory failure.

Additional guidance about the calculation of survey weights is available in many survey design resources, including Appendix J of *WHO Vaccination Coverage Cluster Surveys: Reference Manual (11).*

2.3. Specification of survey design

Besides specifying the survey weight for each participant, to conduct the proper analysis using statistical software, it is necessary to identify strata and cluster membership. There are nice point-and-click options using the Dropdown menus in Stata's survey data analysis (SVY) package. Examples are provided for how these menus would be completed.

With generic survey design, as described in Fig. 2.

Stage	Sampling Units	Strata
1	Administrative units	Dengue burden
2	Schools	(no explicit stratification)
3	Students	(no explicit stratification)

If an additional stage of sampling was required because it was not possible to generate a list of schools at the administrative unit level:

Stage	Sampling Units	Strata
1	Administrative units	Dengue burden
2	Smaller administrative unit	(no explicit stratification)
3	Schools	(no explicit stratification)
4	Students	(no explicit stratification)

Department of Immunization, Vaccines and Biologicals

World Health Organization
20 Avenue Appia
CH-1211 Geneva 27, Switzerland
vaccines@who.int
http://www.who.int/immunization/en/