Statistical Analysis Plan

Applying *Wolbachia* to Eliminate Dengue (AWED): A non-blinded cluster randomised controlled trial to assess the efficacy of *Wolbachia*-infected mosquito deployments to reduce dengue incidence in Yogyakarta, Indonesia

Table of Contents

1. Objectives

1.1. Primary Objective

To assess the efficacy of community-based deployments of *Wolbachia*-infected *Ae. aegypti* mosquitoes in reducing the incidence of symptomatic, virologically-confirmed dengue cases of any severity in Yogyakarta residents aged 3-45 years in release (intervention) areas, relative to non-release (untreated) areas.

1.2. Secondary Objectives

- To measure the efficacy of the *Wolbachia* method against each of the four DENV serotypes.
- To measure the efficacy of the *Wolbachia* method in reducing the incidence of symptomatic virologically-confirmed Zika virus and chikungunya virus infection in intervention areas, relative to untreated areas, and
- To quantify the impact of *Wolbachia* deployments on notifications of dengue haemorrhagic fever (DHF) cases to the Yogyakarta district health office

2. Study Design

2.1. Type of Study

The AWED trial is a parallel two-arm non-blinded cluster randomised controlled trial conducted in a single site in Yogyakarta City, Indonesia. The study site was subdivided into twenty-four contiguous clusters, approximately 1km^2 in size (range 0.7km^2-1.65km^2). Clusters were randomly allocated in a 1-to-1 ratio to receive *Wolbachia* deployments or no intervention, such that 12 clusters received *Wolbachia* deployments and 12 received no intervention (see **Figure 1**).

There are no buffer areas between clusters, but natural borders were used to define cluster boundaries as much as possible, to limit the spatial spread of *Wolbachia* from intervention clusters into untreated areas, and of wild-type mosquitoes in *Wolbachia*-treated clusters. Exclusion areas were minimised, and any areas within the study site where releases were not possible for reasons of logistics, public acceptance or absence of mosquito populations were pre-specified prior to randomisation and balanced between study arms. No attempt is made

to alter the routine dengue prevention and vector control activities conducted by public and private agencies throughout the study area (intervention and untreated clusters). The capacity of the disease surveillance system to detect (and thus respond to) dengue has been enhanced across the city through increased availability of diagnostic kits, which have been supplied to primary care clinics and hospitals since March 2016 by the World Mosquito Program (previously Eliminate Dengue Project) Indonesia, to support efforts to enhance the surveillance of dengue across Yogyakarta.

The impact of *Wolbachia* deployments on dengue incidence will be assessed by comparing the exposure distribution (probability of living in a *Wolbachia*-treated area) among virologically-confirmed dengue cases presenting to a network of public primary clinics (Puskesmas), against the exposure distribution among patients with febrile illness of non-arboviral aetiology presenting to the same network of clinics in the same temporal windows. Dengue cases and arbovirus-negative controls are sampled concurrently from within the population of patients presenting with febrile illness to the study clinic network, with case or control status classified retrospectively based on the results of laboratory diagnostic testing. By recruiting participants from within the population of patients presenting to clinics with febrile illness – with dengue test-positive patients classified as cases and test-negative patients classified as controls – the controls are necessarily drawn from the same source population as the cases, thus avoiding common pitfalls that can introduce selection bias [1]. In this situation, the odds ratio is an unbiased estimate of the rate ratio in the source population over the period of participant enrolment (the 'risk' period), without the need for any rare disease assumption [2,3].

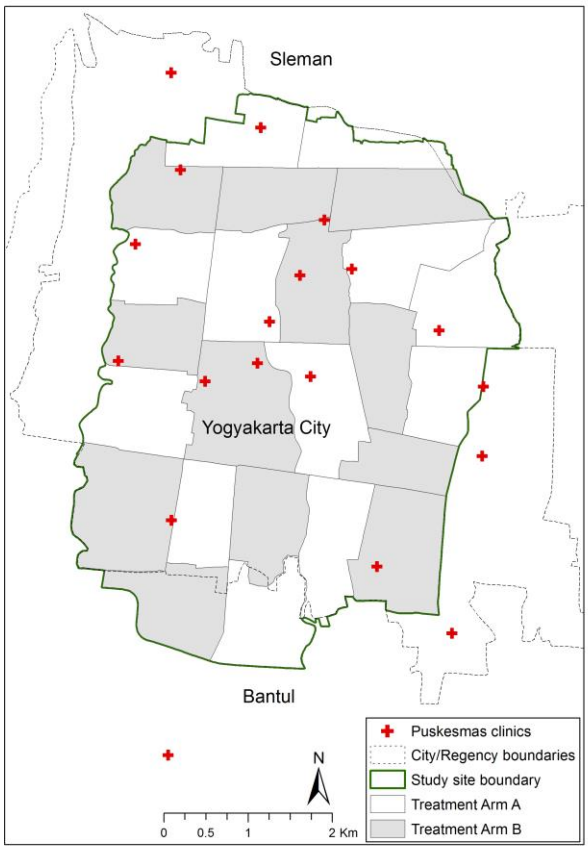

Figure 1. Map of study area, cluster boundaries, and Puskesmas clinics. The study area is outlined in green. The 12 clusters in each treatment arm are shown in grey and white. The location of the Puskesmas clinics at which trial recruitment is conducted are shown by red crosses.

2.2. Study Participants

The study population for measurement of the efficacy endpoint is the population of patients resident in the study area, presenting to the network of participating Puskesmas with febrile illness, and meeting the eligibility criteria as described in **Table 1**. Based on two years of historic data collated from the network of health clinics in the study area, it was estimated that at least 5000 patients per year present to these clinics with febrile illness (range 200-1500 per clinic per annum). We will enroll all participants presenting to any of the participating clinics who meet the eligibility criteria. Following laboratory testing and classification of participants' diagnostic status, all cases and those controls enrolled within the same calendar month as any case will be retained in the dataset for analysis.

Table 1. Participant eligibility criteria

Inclusion criteria	Exclusion criteria
1. Fever (either self-reported or objectively measured, e.g. tympanic membrane temperature ≥37.5^{o}C) with a date of onset between 1-4 days prior to the day of presentation.	1. Localising features suggestive of a specific diagnosis other than an arboviral infection, e.g. severe diarrhea, otitis, pneumonia.
2. Aged between 3-45 years old.	2. Prior enrollment in the study within the previous 4 weeks.
3. Resided in the study area every night for the 10 days preceding illness onset.	

2.3. Expected Duration of Study

The clinic-based sampling of febrile patients commenced in pilot phase in September 2017, with active enrolment in all clinics by December 2017. Enrolment will continue for up to 36 months, unless early termination is recommended by the independent data monitoring committee (IDMC).

3. Analysis Endpoints

3.1. Primary Efficacy Endpoint: Dengue

The primary outcome measure will be virologically-confirmed dengue virus infection in patients reporting febrile illness. Participants will be classified as dengue cases for the primary analysis if plasma samples collected 1-4 days after onset of fever test positive for dengue virus nucleic acid by RT-qPCR and/or dengue virus NS1 antigen (BioRad Platelia NS1 ELISA) (see **Figure 2**). A predefined exploratory analysis will evaluate hospitalised virologically-confirmed dengue cases as an outcome measure (a pragmatic proxy indicator for disease severity).

3.2. Secondary Efficacy Endpoint: DENV serotype-specific

For each participant who tests positive for dengue by RT-qPCR, the infecting serotype will be determined by DENV serotype-specific RT-PCR, and participants with a known serotype

will be included in a secondary analysis to estimate serotype-specific efficacy, as described in section 0.

3.3. Secondary Efficacy Endpoints: Chikungunya and Zika

Secondary outcome measures include chikungunya and Zika virus infection in patients reporting febrile illness. Participants will be classified as virologically-confirmed chikungunya cases if chikungunya nucleic acid is detected in plasma samples by RT-qPCR (see **Figure 2**). Participants will be classified as virologically-confirmed Zika virus cases if Zika virus nucleic acid is detected in plasma samples by RT-qPCR.

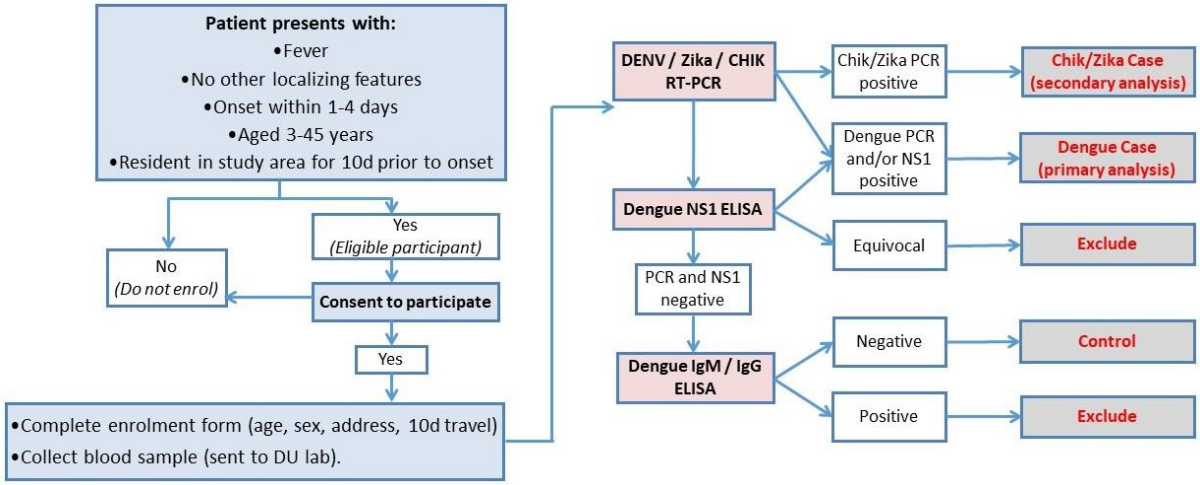

Figure 2. Flowchart of data and sample collection procedures and diagnostic algorithm.

4. Monitoring of *Wolbachia* prevalence in local *Ae. aegypti* populations

A network of BG-Sentinel adult mosquito traps (BioGents) has been in place throughout intervention and untreated clusters for the duration of the trial, evenly spaced throughout residential areas at a density of approximately 16 traps/km^2. BG traps are serviced weekly, with trapped mosquitoes screened for *Wolbachia* at weekly intervals during releases, fortnightly intervals after completion of releases, and monthly intervals since *Wolbachia* establishment (≥80% prevalence for two consecutive screening events). Mosquitoes are bio-banked in the intervening weeks when screening is not done. Trapped mosquitoes are identified by microscopy, and individual *Ae.aegypti* mosquitoes (male and female) are screened using quantitative PCR to detect the presence of *Wolbachia* and to confirm the species as *Ae. aegypti*.

5. Monitoring of unintended adverse effects of *Wolbachia* releases

In order to demonstrate that the deployment is not associated with any excess of a severe adverse outcome, we follow up all enrolled participants by telephone within 14 to 21 days post-enrolment to ascertain their health status, recorded categorically as recovered/died, and whether or not they were ever hospitalised during this illness. Any death of a study participant within 14 to 21 days of enrolment is classified as a serious adverse event (SAE).

6. Sample Size Estimation

It was initially estimated that enrolment of approximately 1000 cases plus four times as many controls would be sufficient to detect a 50% reduction in dengue incidence with 80% power. Simulations were used to estimate the power to detect a range of intervention effect sizes, assuming 12 clusters per arm, a total of 1000 true dengue cases enrolled and 4000 non-dengue controls.

A re-estimation of sample size requirements was conducted in January 2019 after one year of recruitment (**Error! Reference source not found.**A). The initial power calculation used 1000 dengue cases and 4000 non-dengue controls allocated to each cluster based on historical proportions of dengue cases and other febrile illnesses, assuming no variation in the proportion of cases by cluster. This method was found to overestimate power for small samples by not taking into account randomness in the sampling. The sample size re-estimation included power estimates for 200, 400, 600, 800 and 1000 dengue cases with 4 times as many controls allocated to each cluster by sampling from a multinomial distribution, which incorporated added randomness by allowing the proportion of cases allocated to each cluster to vary across simulations. The re-estimation found that 400 dengue cases plus four times as many controls would be sufficient to detect a 50% reduction in dengue incidence with 80% power.

Additional simulations were conducted in September 2019 to assess the potential impact on power if a number of untreated clusters were 'lost' to *Wolbachia* contamination. For the target minimum observed effect size of 50% (Relative Risk (RR)=0.5) and 400 enrolled dengue cases, contamination of 3 untreated clusters (assuming that contaminated clusters experience the full intervention effect for 1 out of the 3 years of trial recruitment) is expected

to result in a ~7% loss of power, and contamination of 6 clusters to result in a ~14% loss of power (**Error! Reference source not found.**A).

7. Statistical Analysis Method

7.1. General Considerations

This SAP was developed on the information provided in AWED Protocol version 5.1 dated 16 October 2019.

All statistical analyses will be generated using Stata version 14.0 or higher, or R (R Foundation for Statistical Computing, Austria).

A blinded data review will be conducted to assess the accuracy and completeness of the study database, prior to unblinding of the cluster intervention allocations. The appropriateness of planned statistical analyses will be assessed on a blinded set of 1000 observations comprised of exposure and demographic data from 1000 randomly selected participants combined with diagnostic results from a separate 1000 randomly selected participants. Exposure information and diagnostic results are stored in separate tables within the database. By merging exposure and outcome information from different randomly selected sets of 1000 participants we aim to avoid accidental unblinding of the data.

Blank result tables are provided in Appendices B and C.

7.2. Analysis Sets

The dataset for analysis will retain all enrolled virologically-confirmed dengue cases, and all test-negative controls that are matched to a case by calendar month of enrolment. Unmatched controls will not be used for the primary analysis.

The analysis will be performed on data acquired during the case surveillance period, that is the period commencing when *Wolbachia* is deemed to have been established throughout intervention clusters, defined as one month after completion of releases in the last cluster (i.e. 8 January 2018). Cases and controls enrolled prior to 8 January 2018 will be excluded from the analysis dataset.

7.3. Status of potential participants

The status of all potential participants that were screened for enrolment will be summarized descriptively, according to the following categories, overall and by treatment arm:

- Number screened
- Number of screened patients that met eligibility criteria
- Number of eligible patients that consented to participate
- Number of consenting participants enrolled in the trial
- Number of enrolled participants successfully followed up for safety endpoints
- Number of enrolled participants for whom a blood sample was available for diagnostic testing
- Number of enrolled participants included in datasets for ITT and PP analysis

7.4. Demographic Characteristics

Participants' age and sex will be summarized descriptively overall, and by treatment arm, diagnostic category, inclusion/exclusion from analysis, and follow-up status.

7.5. Analysis Plan for Primary Efficacy Endpoint

Intention-to-Treat Analysis

The intention-to-treat (primary) analysis will consider *Wolbachia* exposure as a binary classification based on residence in a cluster allocated to *Wolbachia* deployment or not. Residence will be defined as the primary place of residence during the 10 days prior to illness onset.

The intervention effect will be estimated from an aggregate odds ratio comparing the exposure odds (residence in a *Wolbachia*-treated cluster) among test-positive cases versus test-negative controls (for data aggregated across all clusters), using the constrained permutation distribution as the foundation for inference. The null hypothesis is that the odds of residence in a *Wolbachia*-treated cluster is the same among test-positive cases as test-negative controls. The resulting odds ratio provides an unbiased estimate of the RR providing that the key assumptions underlying the TND are upheld (i.e. that test-negative controls are allowed to include participants who may test positive for dengue at any other time during the

study period, and the distribution of non-dengue febrile illness is not associated with the intervention status). To note, since the constrained permutation distribution used for statistical inference contains only the 247 potential allocations (494 distinct randomisations) that meet all balancing criteria, the most extreme odds ratio in the distribution would carry a two-sided p-value of ~0.004 (1/494*2). Therefore, p<0.004 is the minimum threshold at which statistical significance can be evaluated in this design. An exploratory analysis will estimate the intervention effect over time, by calculating the aggregate odds ratio at 12 months and 24 months into the ITT case surveillance period based on the cumulative test-positive cases and test-negative controls enrolled up to that point in time. Efficacy of the intervention will be calculated as 100*(1-aggregate odds ratio). For clarity in reporting of study results, primacy will be given to the aggregate odds ratio approach.

An additional group-level analysis will be performed using a cluster-level summary measure of the proportion of test-positive individuals amongst all tested individuals in each cluster. The difference in the average proportion of test positives between the intervention clusters and untreated clusters will be used to test the null hypothesis of no intervention effect using the t-test statistic but basing inference on the exact permutation distribution. These average proportions in each arm can be used to derive an estimate of the RR of dengue in treated versus untreated clusters, which is a much more intuitive effect measure, using a method described in detail elsewhere [4]. Briefly, we can substitute the estimated difference in the proportions, *d* into the formula $d = \frac{1}{1+(\frac{r}{2})(1+RR)} - \frac{RR}{RR+(\frac{r}{2})(1+RR)}$, where *r* is the simply the ratio of the total number of test negatives to the total number of test positives, which yields a quadratic equation for the unknown *RR*. Only one solution is plausible so that this then yields an estimate of *RR*, along with the appropriately transformed confidence interval (from that associated with *d*).

Per-protocol analysis

The per-protocol analysis will consider *Wolbachia* exposure as a quantitative index based on measured *Wolbachia* prevalence in local *Ae. aegypti* mosquitoes in the participant's cluster of residence, and in locations visited by the participant during the period 3-10 days prior to illness onset. The per-protocol analysis therefore allows for *Wolbachia* exposure to vary in a location over time, and also accounts for human mobility, in terms of the exposure-time that individuals spend outside their cluster of residence as reported in the travel history interview

at enrolment. This analysis can also account for the temporal matching of dengue cases and test-negative controls: risk sets of cases and controls will be defined by frequency matching enrolled confirmed dengue cases to arbovirus-negative controls enrolled in the same calendar month.

Participants are asked about their mobility during the ten days prior to illness onset using a structured interview administered at enrolment. This records the duration of time spent at home, work or school, and other locations visited during daylight hours (5am – 9pm) in the ten-day period. The geographic coordinates of those locations are derived by geo-locating them on a digital map, with the assistance of the respondent. A weighted '*Wolbachia* exposure index' (WEI) will be defined for each participant, as follows. The aggregate *Wolbachia* prevalence for each cluster will be calculated each month from all *Ae. aegypti* trapped in that cluster. For any calendar month where mosquito collection was not done, the average of the cluster-level *Wolbachia* prevalence in the one previous and one subsequent month will be used. The WEI for each participant will then be calculated by multiplying the cluster-level Wolbachia prevalence (in the calendar month of participant enrolment) at each of the locations visited, by the proportion of time spent at each location, to give a value on a continuous scale from 0 to 1. For visited locations within the quasi-experimental study area, the measured kelurahan-level *Wolbachia* prevalence from the screening event closest in time to the participant's enrolment will be used. Visited locations outside of both the AWED study area and the quasi-experimental study area will be assumed to have a *Wolbachia* prevalence of zero. The process of calculating WEI will be conducted blinded to participants' case/control status, by partitioning the travel history data from the laboratory diagnostic data, to remove any possibility of observer bias.

An additional per-protocol analysis will be conducted in which the WEI is calculated using only the cluster-level *Wolbachia* prevalence in the participant's cluster of residence (in the calendar month of participant enrolment), ignoring the participant's recent travel history. This recognises that dengue exposure risk may be higher at home versus other locations, rather than assuming an even distribution of exposure risk across daytime hours and locations visited.

Cases and controls will be classified by strata of their WEI: 0-<0.2; 0.2-<0.4; 0.4-<0.6; 0.6-<0.8; and 0.8-1. This acknowledges that the WEI is not a highly precise measure, and serves to reduce error in exposure classification. The ITT methods described above will be extended to allow for this individual level covariate using a regression approach [5], adjusted for time. A mixed effects logistic regression model will be fitted, incorporating time as random effect and with another random effect for cluster membership. Such models yield an estimate, and associated confidence interval, for the relative risk. Efficacy will then be calculated as 100*(1-RR). The WEI strata will first be included as an ordinal covariate and the slope of the WEI variable will be tested for a difference from zero. The WEI strata will additionally be included as a nominal (unordered) covariate to calculate stratum-specific IRRs (relative to the baseline 0-<0.2 stratum). This will allow examination of a 'dose response' relationship. An additional benefit of including WEI as a nominal variable is that it avoids any assumption of linearity in the dose response relationship.

7.6. Analysis of Secondary Efficacy Endpoints

DENV serotype-specific efficacy of* Wolbachia *deployment

In laboratory experiments, the degree to which *Wolbachia* reduces the DENV transmission potential of *Ae. aegypti* is dependent on the infecting virus serotype, with DENV1 transmission least affected [6]. A secondary analysis will estimate the serotype-specific efficacy of *Wolbachia* deployments in reducing symptomatic dengue virus infection with a known infecting serotype, for each of the four serotypes in turn, or as many as are detected in the study population. The same intention-to-treat and per-protocol analyses will be used as described for the primary endpoint above, with case populations restricted to each of the DENV serotypes in turn, and with the same control population as for the primary analysis.

Impact of Wolbachia deployment on Zika and chikungunya

There exists no baseline data on the prevalence of Zika or chikungunya infection among febrile patients presenting to primary health care clinics in Yogyakarta City, from which to estimate the expected number of cases; therefore, these secondary analyses are exploratory only and not subject to any formal sample size or power calculations. Blood samples from enrolled participants will be tested by Zika and chikungunya PCR for the purpose of defining arbovirus-negative controls for the primary analysis, as described above. These results will permit estimation of the prevalence of virologically confirmed Zika virus and chikungunya

virus infection among the study population of ambulatory febrile patients presenting to primary health care.

If ≥20 virologically confirmed Zika or chikungunya cases are detected, a secondary analysis will estimate the efficacy of *Wolbachia* deployments in reducing the incidence of symptomatic virologically confirmed Zika virus and chikungunya virus infection. The same enrolled patient population will be used to analyse all three arbovirus endpoints (dengue, Zika and chikungunya), and the same intention-to-treat and per-protocol analyses will be used as described for the primary (dengue) endpoint above. For Zika and chikungunya, the cases will be defined as enrolled participants who test positive by Zika or chikungunya PCR, respectively, and the controls will be those who test negative to all three arboviruses. Cases and controls will be matched by month of enrolment, as described above. If <20 cases of either Zika or chikungunya are detected there will be no formal analysis, only a descriptive analysis of the temporal and spatial distribution of cases.

Impact of Wolbachia deployment on notified dengue cases

The existing system for routine notification of dengue cases in Yogyakarta City is based on hospital-reporting of cases diagnosed clinically as Dengue Hemorrhagic Fever (DHF), which historically have not been accompanied by supportive laboratory testing. Since March 2016, hospitals have been encouraged to record a serological testing result, where available, on the report form, and also to report cases diagnosed clinically as Dengue Fever where there is a confirmatory NS1-positive test result. A separate reporting system, established in March 2016, collates data on the number of NS1 rapid tests performed – and number positive – in Puskesmas across the city. Both of these reporting systems include address information for notified cases.

We will collate data from these two reporting systems on a monthly basis, aggregated by kelurahan of residence, to monitor trends in reported dengue incidence across the city and by kelurahan, before, during and after *Wolbachia* deployment.

The impact of *Wolbachia* deployment on DHF case notifications will be evaluated using an interrupted time series analysis of monthly DHF notifications by kelurahan, before, during and after *Wolbachia* releases. Methods will be developed and validated *a priori* to classify

area-level *Wolbachia* exposure status in a way that aligns with the kelurahan boundaries by which dengue cases are reported. A separate statistical analysis plan will be developed for this endpoint and the results will be reported in a secondary publication, subsequent to the publication of the main trial results.

7.7. Monitoring of Safety Endpoints

The safety endpoints of hospitalisation and death will be summarised by treatment arm. Any difference in the distribution of these two safety endpoints between treatment arms will be evaluated from an aggregate odds ratio comparing the exposure odds (residence in a *Wolbachia*-treated cluster) among those with versus without the endpoint (for data aggregated across all clusters), using the constrained permutation distribution as the foundation for inference, and from the relative risk of hospitalisation in the intervention versus untreated clusters, derived from a comparison between treatment arms of the mean proportion of hospitalised participants among total participants in each cluster. These analyses will be repeated among VCD cases only, to compare the distribution of hospitalisations of VCD cases between treatment arms.

7.8. Interim Analysis

The trial protocol states that an interim analysis will be conducted at the mid-point of the study, i.e. after enrolment of 500 dengue cases with an initial target sample size of 1000. Re-estimation of statistical power conducted in January 2019 showed that the trial has 80% power to detect a reduction in dengue incidence greater than or equal to 50%, for a minimum sample of 400 virologically-confirmed dengue cases. This finding demonstrates that the trial is likely to be adequately powered even though it will not reach the original target of 1000 dengue cases prior to its revised completion date in August 2020. The original plan of conducting an interim analysis after enrolment of 500 dengue cases is therefore no longer appropriate. The IDMC and Trial Steering Committee decided in November 2019 that no interim analysis will be done for this study.

8. Differences between protocol and SAP

Differences between the approved protocol (version 5.1) and the SAP are listed in the table below.

SAP section, page	Text in SAP	Difference from protocol
Figure 2, page 7	All blood samples are tested by RT-PCR and NS1	Only samples that are PCR negative for dengue, chikungunya and Zika are subsequently tested using NS1.
7.5, page 12	This records the duration of time spent at home, work or school, and other locations visited during daylight hours (5am – 9pm) in the ten-day period.	Protocol stated '…and up to three other most-visited locations…', but in practice all locations visited for ≥1 hour were recorded.
7.2, page 9	The same analysis dataset will be used for ITT and PP analysis, restricted to cases and controls enrolled from one month after the completion of releases (i.e. 8 Jan 2018).	In the protocol, the PP analysis dataset includes all cases and controls enrolled from the start of full clinic enrollment. In practice controls from Dec 2017 would be excluded due to no cases, so the only difference from ITT would be inclusion of participants enrolled 1–7 Jan 2018. For simplicity, align PP dataset with ITT dataset.
7.2, page 9	The dataset for analysis will retain all enrolled virologically-confirmed dengue cases, and all test-negative controls that are matched to a case by calendar month of enrolment.	The protocol states that cases and controls will be matched on calendar month of illness onset.
7.5, page 11	This analysis can also account for the temporal matching of dengue cases and test-negative controls: risk sets of	

	cases and controls will be defined by frequency matching enrolled confirmed dengue cases to arbovirus-negative controls enrolled in the same calendar month.	
7.6, page 14	Cases and controls will be matched by month of enrolment, as described above.	
7.6, page 14	For the analysis of Zika and chikungunya secondary endpoints, added a caveat that if <20 cases of either disease are detected then no formal analysis will be undertaken, only a descriptive analysis of the temporal and spatial distribution of cases.	
7.8, page 15	This finding demonstrates that the trial is likely to be adequately powered even though it will not reach the original target of 1000 dengue cases prior to its revised completion date in August 2020.	The protocol states that the revised completion date is November 2020.

9. References

1. De Serres G, Skowronski DM, Wu XW, Ambrose CS. The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. Euro Surveill 2013;18.
2. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. Am J Epidemiol 1982;116:547-53.
3. Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. Int J Epidemiol 2012;41:1480-9.
4. Jewell NP, Dufault S, Cutcher Z, Simmons CP, Anders KL. Analysis of cluster-randomized test-negative designs: cluster-level methods. Biostatistics 2019;20:332-46.
5. Small DS, Ten Have TR, Rosenbaum PR. Randomization Inference in a Group–Randomized Trial of Treatments for Depression. Journal of the American Statistical Association 2008;103:271-9.
6. Ferguson NM, Kien DT, Clapham H, et al. Modeling the impact on virus transmission of Wolbachia-mediated blocking of dengue virus infection of Aedes aegypti. Sci Transl Med 2015;7:279ra37.