gsk GlaxoSmithKline	**Statistical Analysis Plan**
Detailed Title:	A prospective, cohort study to determine the incidence of acute febrile dengue illness and to build capacity for dengue vaccine clinical endpoint trials in South Asian communities
eTrack study number and Abbreviated Title	200274 (DPIV – 021 EXPLO)
Scope:	All data pertaining to the above study.
Date of Statistical Analysis Plan	Final 21 October 2019

APP 9000058193 Statistical Analysis Plan Template V4 (Effective date: 3 June 2019)

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

AFI	Acute Febrile Illness
CRF	Case Report Form
DENV	Dengue Virus
DOB	Date of Birth
ELISA	Enzyme-Linked Immunosorbent Assay
FN	False Negative
FP	False Positive
GMT	Geometic mean titre
ICT	Immunochromatographic
JEV	Japanese Encephalitis Virus
LCD	Laboratory Confirmed Dengue
NPV	Negative Predictive Value
NS1	Non-Structural Protein 1
PPS	Per Protocol Set
PPV	Positive Predictive Value
RT-qPCR	Reverse Transcriptase-quantitative Polymerase Chain Reaction
TFL	Tables Figures and Listings
TN	True Negative
ToC	Table of Contents
TP	True Positive

1. DOCUMENT HISTORY

Date	Description	Protocol Version
21 OCT 2019	first version	Amendment 3: 01 MAR 2017

2. STUDY DESIGN

This a prospective, community-based cohort, household-sampling study conducted in Sri Lanka. Approximately 2000 subjects aged between 6 month and 50 years from randomly selected households in the geographically-defined communities were enrolled in the study.

The study consists of 2 scheduled visits (one at enrolment and one at study close-out), weekly subject contacts, and unscheduled suspected Dengue visits in case of Acute Febrile Illness (AFI). Blood samples will be collected at each suspected dengue visit.

Figure 1   Study Design Overview

Scheduled Visits

- Visit 1 and Visit 2 will take place at home or at the hospital/clinic

- Weekly contacts will be home visits (at least every other week) or telephone contacts (TCs)
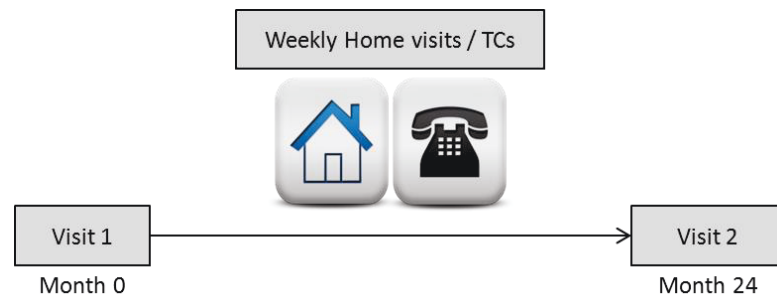
Unscheduled Visits

- Suspected Dengue First Visit (hospital/clinic)

- Optional Return Visit (hospital/clinic)

- Suspected Dengue Follow-up Visit (hospital/clinic or home)
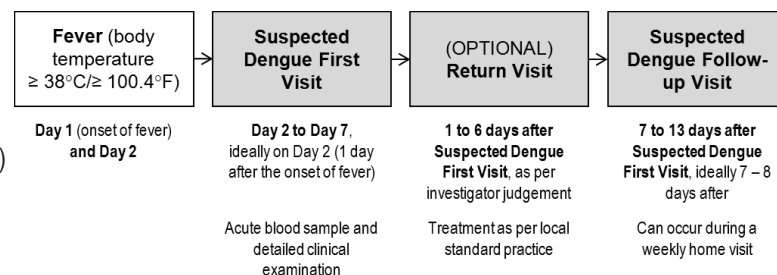
3.   OBJECTIVES/ENDPOINTS

3.1.   Primary Objective

- To determine the incidence of AFI due to Laboratory Confirmed Dengue (LCD) in the study population.

3.2.   Secondary Objectives

- To determine the incidence of AFI due to non-LCD in the study population.

- To describe the signs and symptoms of AFI due to LCD and of AFI due to non-LCD.

- To estimate the incidence of AFI due to LCD by Dengue Virus (DENV) type, , study site and age group.

3.3.   Tertiary Objectives

- To describe DENV and Japanese Encephalitis Virus (JEV) antibody profiles in subset*s* of subjects with AFI.

- To assess the concordance between dengue Reverse transcriptase-quantitative Polymerase Chain Reaction (RT-qPCR) and Non-Structural protein 1 (NS1)-antigen Immuno Chroma Tographic (ICT) test and/or Enzyme-Linked Immunosorbent Assay (ELISA) in the assessment of dengue infection.

- To gather clinical data that will support the definition of clinical endpoints for future efficacy trials.

- To describe spatio-temporal clustering of LCD cases.

- To describe entomological characteristics at selected site(s).

- To estimate the prevalence and incidence of tuberculosis diagnosis in the study population.

- To describe the incidence of hospitalisation and discharge diagnosis in the study population.

- To isolate and characterise infectious agents from a subset of subjects with AFI.

- To evaluate the antibody response to infectious agents in a subset of subjects with AFI.

3.4.   Primary endpoint

- Occurrence of AFI due to LCD.

3.5.   Secondary endpoints

- Occurrence of AFI due to non-LCD.

- Occurrence and intensity of signs and symptoms of interest, during the 7-day period following the onset of each episode of AFI due to LCD and due to non-LCD.

- Occurrence of AFI due to LCD by DENV type, study site and age group.

3.6.   Tertiary endpoints

The following is a list of planned tertiary analyses. Per protocol, tertiary study endpoint assays other than dengue serology *may be* performed for a subset of samples (for example, JEV assays will only be conducted if additional studies indicate that JEV serology is needed to interpret the DENV antibody).

Refer to section 8 for information regarding tertiary analyses that will not be performed.

- Neutralising antibody titres against DENV 1-4.

- Neutralising antibody titres against JEV.

- Neutralising antibody titres against other viruses, including but not limited to, Chikungunya virus.

- Incidence of NS1-antigen (ICT test and/or ELISA)-positive AFI by DENV type.

- Concordance between dengue RT-qPCR and NS1-antigen (ICT and/or ELISA) assays.)

- Occurrence of AFI due to LCD and due to non-LCD having $\geq$ 3 days of fever (body temperature $\geq 38°C/\geq 100.4°F$).

- Occurrence of AFI due to LCD and due to non-LCD resulting in hospitalisation.

- Occurrence of combinations of signs and symptoms during the 7-day period following the onset of each episode of AFI due to LCD and due to non-LCD.

- Occurrence of AFI due to LCD within 2 weeks from an index case, in the same household or within 50 metres of the household of the index case.

- Entomological characteristics for LCD cases (e.g., vector species, density and number of breeding sites)

- Occurrence (by medical history only) of a diagnosis of tuberculosis, from birth up to study conclusion, in the study population.

- Occurrence of hospitalisation and discharge diagnosis.

- Identification and characterisation of infectious agents isolated from blood (e.g., DENV, influenza viruses, chikungunya viruses), in subjects with AFI.

4.   ANALYSIS SETS

4.1.   Definition

4.1.1.   Total Cohort

The Total cohort will include all subjects enrolled in the study.

4.1.2.   Total Cohort without subjects with code 900

The Total cohort without subjects with code 900 will include all subjects from the Total cohort excluding subjects with elimination code 900.

This cohort will be used only if a code 900 is attributed. In this case, this cohort will replace the Enrolled cohort in all statistical analyses.

4.2.   Criteria for eliminating data from Analysis Sets

Elimination codes are used to identify subjects to be eliminated from analysis. Details are provided below for each set.

4.2.1.   Elimination from Total Cohort without subjects with code 900

Code 900 (invalid informed consent or fraud data) will be used for identifying subjects eliminated from the Total cohort without subjects with code 900.

5.   STATISTICAL ANALYSES

That standard data derivation rules and statistical methods are described in section xx while the study specific data derivation rules and statistical methods are described in section xx.

5.1.   Demography/baseline characteristics

5.1.1.   Analysis of demographics/baseline characteristics planned in the protocol

Socio-demographic and patient characteristics (e.g., age at study enrolment, gender, household conditions, medical history and vaccination history) will be summarized overall and for LCD cases using descriptive statistics:

- Frequency tables will be generated for categorical variable such as gender.

- Mean, median, standard deviation will be provided for continuous data such as age.

5.2.   Primary objective

5.2.1.   Analysis of primary objective planned in the protocol

The incidence rate of first AFI due to LCD will be calculated along with 95% exact confidence interval. This analysis will be based on the total enrolled set. See section 9.2.1 and section 9.2.2 for details of incident rate and confidence interval calculation respectively.

5.2.2.   Additional Consideration

Additional case definition of AFI was added to the analysis. See Table 2 for details of ad-hoc case definitions. Incidence rate along with 95% exact confidence interval of AFI based on the ad-hoc case definition due to LCD will be calculated.

5.3.   Secondary objective

5.3.1.   Analysis of secondary objective planned in the protocol

The incidence rate of first AFI due to LCD by serotype (DENV 1 – 4) and incidence rate of AFI due to non-LCD with 95% exact confidence interval will be calculated for overall and age group (could you please put here the age categories?).

Description of signs and symptoms of AFI due to LCD and due to non-LCD will include the percentage of AFI presenting each sign or symptom (any intensity and grade 3) during the 7-day period from the onset of fever (body temperature 38°C/100.4°F). The sign or symptom of suspected dengue infection is summarized in Table 1. Duration of signs and symptoms will also be described for LCD and non-LCD episodes. Analysis will be done overall and then separately by age group.

Table 1   Solicited symptoms of suspected dengue infection

Signs and symptoms
body temperature and route (minimum of two measurements)
headache / irritability
eye pain
myalgia (muscle pain)
arthralgia (joint pain)
abdominal pain
nausea
vomiting
rash
any bleeding (skin, mouth, anus)
loss of appetite
Fatigue/decrease in normal activity
reduced fluid intake

5.4. Tertiary objectives

5.4.1. Analysis of tertiary objectives planned in the protocol

- Sensitivity, specificity, positive predictive value, negative predictive and overall agreement value of NS1-antigen (ICT test) in prediction of RT-qPCR test results will be computed. Cohen's Kappa will be used to evaluate the agreement between two measures. See section 9.2.3.

- Computation of incidence rate of AFI due to dengue confirmed by NS1-antigen (ICT test and/or ELISA) with 95% CI: the numerator will be the number of subjects with AFI due to dengue confirmed by NS1-antigen (ICT test and/or ELISA) during the study period. The denominator will be the total person-years at risk.

- Computation of incidence rates of AFI due to LCD having ≥ 3 consecutive calendar days of fever (body temperature $\geq 38°C/\geq 100.4°F$) and incidence of AFI due to LCD resulting in hospitalisation with 95% CI.

- Frequency of symptoms associated with AFI due to LCD having ≥ 3 consecutive calendar days of fever (body temperature $\geq 38°C/\geq 100.4°F$) and result in hospitalization will tabulated overall and by age group. Univariate and Multivariates analyses will be done to identify categories of medically relevant signs and symptoms associated with high probability of LCD (any severity, having ≥ 3 consecutive calendar days of fever (body temperature $\geq 38°C/\geq 100.4°F$) and resulting in hospitalisation) *versus* non-LCD episodes using logistic regressions. See section 9.2.4.

- Computation of the frequency and percentage of the reasons of hospitalization throughout the study among all hospitalization visits.

- Calculation of the frequency and percentage of subject with other infective agent: influenza viruses, chikungunya viruses among subject with subjects with AFI.

6. ANALYSIS INTERPRETATION

All analyses are descriptive. The use of these descriptive analyses should be limited to supportive analysis of confirmatory analyses or hypothesis generation.

7.   CONDUCT OF ANALYSES

7.1.   Sequence of analyses

Description	Analysis ID	Disclosure Purpose (CTRS=public posting, SR=study report, internal)	Dry run review needed (Y/N)	*Study Headline Summary (SHS)* requiring expedited communication to upper management (Yes/No)	Reference for TFL
Final Analysis	E1_01	SR	No	No?	DPIV-021 EXPLO (200274) TFL

8.   CHANGES FROM PLANNED ANALYSES

Due to limitations of data (no data availability) or resources, the following tertiary objectives are omitted from the analysis.

- To describe DENV and Japanese Encephalitis Virus (JEV) antibody profiles in subse*t*s of subjects with AFI.

- Neutralising antibody titres against DENV 1-4.

- Neutralising antibody titres against JEV.

- Neutralising antibody titres against other viruses, including but not limited to, Chikungunya virus.

- To describe spatio-temporal clustering of LCD cases.

- To describe entomological characteristics at selected site(s).

- To estimate the prevalence and incidence of tuberculosis diagnosis in the study population.

- To isolate and characterise infectious agents from a subset of subjects with AFI.

- To evaluate the antibody response to infectious agents in a subset of subjects with AFI.

9.   NON-STANDARD DATA DERIVATION RULES AND STATISTICAL METHODS

9.1.   Data derivation

9.1.1.   Case Definition

The different case definitions for febrile illness are summarized in Table 2

Table 2   Summary of case definitions

Endpoint	Case definition	Case definition for ad-hoc Analysis
AFI with LCD	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hours apart **AND** Dengue RT-qPCR result on the acute serum sample taken during the 7-day period from onset of fever is positive	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp ≥ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue RT-qPCR result on the acute serum sample taken during the 7-day period from onset of fever is positive
DENV – 1 LCD	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hours apart **AND** Dengue 1 specific RT-qPCR result during the 7-day period from fever is positive	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp ≥ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], AND Dengue 1 specific RT-qPCR result during the 7-day period from fever is positive

Endpoint	Case definition	Case definition for ad-hoc Analysis
DENV – 2 LCD	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hours apart **AND** Dengue 2 specific RT-qPCR result during the 7-day period from fever is positive	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp ≥ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue 2 specific RT-qPCR result during the 7-day period from fever is positive
DENV – 3 LCD	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, least twice 8 hours apart **AND** Dengue 3 specific RT-qPCR result during the 7-day period from fever is positive	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp ≥ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue 3 specific RT-qPCR result during the 7-day period from fever is positive
DENV – 4 LCD	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hours apart	[Fever (body temp ≥ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR**

Endpoint	Case definition	Case definition for ad-hoc Analysis
	AND Dengue 4 specific RT-qPCR result during the 7-day period from fever is positive	Fever (body temp $\geq$ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue 4 specific RT-qPCR result during the 7-day period from fever is positive
Non LCD	[Fever (body temp $\geq$ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hours apart **AND** Dengue RT-qPCR result during the 7-day period from fever is negative	[Fever (body temp $\geq$ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp $\geq$ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hours apart **AND** Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue RT-qPCR result during the 7-day period from fever is negative
Undetermined	[Fever (body temp $\geq$ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, at least twice 8 hour apart **AND** Dengue RT-qPCR result during the 7-day period from fever is invalid or missing	[Fever (body temp $\geq$ 38 C /100.4 F) on at least 2 consecutive days are measured at least twice, **OR** Fever (body temp $\geq$ 38 C /100.4 F) on at least 1 day is measured at least twice 8 hour apart

Endpoint	Case definition	Case definition for ad-hoc Analysis
		AND Clinical suspicion of dengue base on solicited symptoms], **AND** Dengue RT-qPCR result during the 7-day period from fever is invalid or missing

9.2. Statistical Method

9.2.1. Computation of incidence rate

The incidence rate of first event (IR, number of first episodes per 1000 person-years) will be calculated by dividing the number of subjects reporting at least one episode of event during the follow-up period by the total person-year at risk.

The person-time at risk for an event of interest (e.g. AFI due to LCD) will be calculated as the duration between the date of enrolment and the end of the at-risk period or the earliest of the following:

- Onset date of event of interest (e.g. first episode of AFI due to LCD)

- Date of last contact

- Date of death.

9.2.2. The method for estimating the CIs for incidence rate not accounting for clustered data

To estimate the confidence interval of the incidence rate, the exact Poisson confidence interval will be used [Clopper, 1934].

If *n* is the number of subjects presenting a given characteristic among these *Ny* subjects per year, the true incidence rate per 1000 person-years can be estimated by *(n/Ny)*1000*. Its exact *(1-α)%* confidence interval is obtained from:

$\qquad$ *CINV(α /2,2*n)/2/Ny*1000* as the lower boundary

and

$\qquad$ *CINV((1-α /2),2*(n+1))/2/Ny*1000* as the upper boundary.

where *CINV (probability, degrees of freedom)* returns the inverse of the chi-squared probability distribution and α is the type I error rate, which will be 0.05 for calculating a 95% CI.

9.2.3. Calculation of the Concordance between dengue RT-qPCR and NS1-antigen

Statistical analysis of concordance will be performed to compare a dengue case detected by RT-qPCR and NS1-antigen

Three 2x2 contingency table will be provided and the following measures of agreement will be calculated as shown in Table 3.

Table 3   Comparison of two diagnostic methods for Dengue cases

Dengue cases		RT-qPCR	
		Positive	Negative
NS1-antigen	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True Negative (TN)

- Sensitivity: TP/ (TP+FN)

- Specificity: TN/ (TN+FP)

- Positive predictive value (PPV): TP/ (TP+FP)

- Negative predictive value (NPV): TN/ (TN+FN)

- Proportion of overall agreement, which is the proportion of cases similarly classified: (TP + TN)/ (TP+FP+FN+TN).

- Cohen's kappa coefficient. The magnitude of the kappa coefficient represents the proportion of agreement greater than that expected by chance.

9.2.4.   Logistic Regression

Univariate and multivariable logistics regression models will be presented. Variables with univariate p-value less than 0.1 will be included in the multivariable model. A step-wise selection using 0.1 as criteria of entering and removing will be used for selection variables in the multivariate analysis.

10. ANNEXES

10.1. Business rules for standard data derivations and statistical methods

This section contains GSK Vaccines' standard rules for data display and derivation for clinical and epidemiological studies. These rules will be applied along with those detailed in section XX (additional study-specific rules).

10.1.1. Handling of missing data

10.1.1.1. Dates

When partially completed dates (i.e. with missing day or month) are used in calculations, the following standard rules will be applied:

- A missing day will be replaced by 15

- A missing day and month will be replaced by June 30th.

10.1.1.2. Laboratory data

Missing laboratory results (including immunological data) will not be replaced.

10.1.1.3. Daily recording of solicited symptoms

10.1.1.3.1. Studies with electronic diaries

For studies using electronic diaries for the collection of solicited symptoms, symptoms will be considered present only when a daily recording of grade 1 or more is present.

10.1.2. Data derivation

10.1.2.1. Age at enrolment

Age of enrolment will be calculated as the difference between the date of enrolment and date of birth, in years, and will be expressed in years. It will be calculated as the number of complete calendar years between the date of birth and the date of vaccination. For example:

DOB = 10SEP1983, Date of enrolment = 09SEP2018 -> Age = 34 years

DOB = 10SEP1983, Date of enrolment = 10SEP2018 -> Age = 35 years

10.1.2.2. Age at time of AFI

The subject age at time AFI will be calculated as the differences between the date of the onset of fever and date of birth in years. It will be calculated as the number of complete calendar years between the date of birth and the date of vaccination. For example:

DOB = 10SEP1983, Date of AFI = 09SEP2018 -> Age = 34 years

DOB = 10SEP1983, Date of AFI = 10SEP2018 -> Age = 35 years

Age at time AFI will also be categorize as following subgroups: < 5 years of age, 5 to 11 years of age, $\geq$ 12 to 17 years of age, and $\geq$ 18 years of age.

10.1.2.3. Person Year

The follow up time at risk for estimation of incidence of AFI due to LCD, non-LCD and each serotype of DENV will calculated as the duration between date of enrolment and date of the first occurrence of AFI or the date of last contact whichever comes first. The person year will be calculated by dividing the duration in days between date of enrolment and date of first occurrence of AFI or the date of last contact whichever comes first by 365.25. For example:

Date of enrolment = 09JAN2018, Date of AFI = 11SEP2018 - > 245 Days/365.25 -> 0.67 PY

The total person-year at risk will be the sum of the follow-up times in years for all individuals.

10.1.2.4. Counting rules for occurrences of events (AFI or symptoms)

When the occurrences of solicited adverse events are summarized, each event recorded as having occurred during a specific period will be counted as only one occurrence regardless of the number of days on which it occurs.

10.1.3.   Display of decimals

10.1.3.1.   Percentages

Percentages and their corresponding confidence limits will be displayed with:

- no decimals when there are fewer than 50 subjects in each tabulated group

- one decimal when there are at least 50 subjects in at least one tabulated group

    - Exceptions will be made for percentages that are not 0% or 100% but appear as 0% or 100% due to rounding. For these specific cases the number of decimals will be increased until the displayed value is no longer 0% or 100%. Examples are given in Table 4.

Table 4   Display of percentages

n/N	Displayed percentage
10/45	22%
1/45	2%
10/55	18.2%
1/55	1.8%
1/300	0.3%
1/3000	0.03%
1/30000	0.003%
299/300	99.7%
2999/3000	99.97%
29999/30000	99.997%

- The display of additional decimals for values close to 0% or 100% will be applied only to point estimates and not confidence limits, which can be rounded and displayed as 0% or 100%.

- Values of exactly 0% or 100% will be presented with no decimals regardless of the number of subjects per tabulated group.

10.1.3.2.   Demographic/baseline characteristics statistics

The mean, median, and standard deviation for continuous baseline characteristics will be presented with one decimal.

The minimum and maximum values and quartile values (if required) will be presented with the same number of decimals as the observed values.

10.2.   TFL and/or TFL ToC

The TFL can be found in eTMF folder section 11.1.1.

11.   REFERENCE

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. Biometrika. 1934;26:404-413