GlaxoSmithKline	**Statistical Analysis Plan**
Detailed Title:	A study to explore cytomegalovirus primary infection, re-activation and re-infection in an adolescent female population.
eTrack study number and Abbreviated Title	115639 (CMV-014 EXPLO)
Scope:	All data pertaining to the above study. The primary analysis will be performed on the ATP cohort(s). Analysis on the Total cohort may be performed to complement the primary analysis.
Date of Statistical Analysis Plan	Final
Co-ordinating author:	PPD (Lead Statistician) PPD (Biostatistician)
Reviewed by:	PPD (Clinical and Epidemiology Project Lead) PPD (Director Statistics) PPD (Lead statistical analyst) PPD (Scientific writer) PPD (Regulatory Affair) PPD (SERM Physician) PPD (Public disclosure representative)
Approved by:	PPD (Clinical and Epidemiology Project Lead) PPD (Lead statistician) PPD (Lead Scientific writer)

	PPD (Lead stat Analyst)

APP 9000058193 Statistical Analysis Plan Template (Effective date: 14 April 2017)

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

No table of figures entries found.

LIST OF ABBREVIATIONS

AE	Adverse event
AESI	Adverse Events of Special Interest
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ATP	According to Protocol
CI	Confidence Interval
CRF	Case Report Form
CSR	Clinical Study Report
CTRS	Clinical Trial Registry Summary
EL.U/ml	ELISA unit per milliliter
Eli Type	Internal GSK database code for type of elimination code
ELISA	Enzyme-linked immunosorbent assay
ES	Exposed Set
GMC	Geometric mean antibody concentration
GMT	Geometric mean antibody titer
GSK	GlaxoSmithKline
IU/ml	International units per milliliter
LL	Lower Limit of the confidence interval
MedDRA	Medical Dictionary for Regulatory Activities
N.A.	Not Applicable
PD	Protocol Deviation
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SBIR	GSK Biological's Internet Randomization System
SD	Standard Deviation
SR	Study Report
SUSAR	Suspected Unexpected Serious Adverse Reactions
TFL	Tables Figures and Listings
TOC	Table of Contents
UL	Upper Limit of the confidence interval
WBR	Web-based Randomization

1. DOCUMENT HISTORY

Date	Description	Protocol Version
14-June-2017	first version	CMV-014 EXPLO (115639) Protocol Amendment 2 (19-Feb-2015)
21-Feb-2018	Second version: Description Of changes from first version are as below - The table of elimination code has been modified for more clarify by adding Elim Code3500	CMV-014 EXPLO (115639) Protocol Amendment 2 (19-Feb-2015)

2. STUDY DESIGN

This is a prospective, multi-center, multi-country study. In total, approximately 240 seropositive subjects whom anti-CMV IgG antibodies are detected in serum sample collected at Visit 1 and 160 seronegative subjects (10-17 years) whom anti-CMV immunoglobulin G (IgG) antibodies are not detected in serum sample collected at Visit 1.will be enrolled in this study.
Table 1 describes the difference between enrolment groups. Table 2 describes the time and events during the study.

Table 1 Study groups foreseen in the study

Study Groups	Number of subjects	Age (Min/Max)
S+	+/- 240 seropositive subjects	10 – 17 years
S-	Estimated approximate number of seronegative subjects(1): 160	10 – 17 years

(1) As subjects will be enrolled consecutively without previous CMV screening with a target number of 240 seropositive subjects, the number of seronegative subjects enrolled will depend on the seroprevalence of the participating countries.

Table 2 Time and Events Table – Follow-up Period (until 36 Month)

Time point							
	Year 1						
Month	**M0**	**M2**	**M4**	**M6**	**M8**	**M10**	**M12**
Site Visits	**V1**		**V2**		**V3**		**V4**
Sample Collection Visits		**SCV1**		**SCV2**		**SCV3**	
	Year 2						
Month		**M14**	**M16**	**M18**	**M20**	**M22**	**M24**
Site Visits			**V5**		**V6**		**V7**
Sample Collection Visits		**SCV4**		**SCV5**		**SCV6**	
	Year 3						
Month		**M26**	**M28**	**M30**	**M32**	**M34**	**M36**
Site Visits			**V8**		**V9**		**V10**
Sample Collection Visits		**SCV7**		**SCV8**		**SCV9**	
Samples:							
Blood	✔		✔		✔		✔
Saliva	✔	✔	✔	✔	✔	✔	✔
Urine	✔	✔	✔	✔	✔	✔	✔

In case of pregnancy resulting in live birth: Urine and or saliva from the newborn will be collected (if possible) as per standard of care in the center where the subject has delivered or through a home visiting nurse or alternatively, the subject may be asked to come back to the study center within 10 days of delivery.

Outline of study procedures

Table3 List of study procedures

Age	**10-17 years**							
Epoch	**001**							
	Year 1							
Time-point (Months)	**M0**	**M2**	**M4**	**M6**	**M8**	**M10**	**M12**	
Visit/Sample Collection Visit	**V1**	**SCV1**	**V2**	**SCV2**	**V3**	**SCV3**	**V4**	
	Year 2							
Time-point (Months)		**M14**	**M16**	**M18**	**M20**	**M22**	**M24**	
Visit/Sample Collection Visit		**SCV4**	**V5**	**SCV5**	**V6**	**SCV6**	**V7**	
	Year 3							
Time-point (Months)		**M26**	**M28**	**M30**	**M32**	**M34**		**M36**
Visit/Sample Collection Visit		**SCV7**	**V8**	**SCV8**	**V9**	**SCV9**		**V10**
Informed consent	●							
Check inclusion/exclusion criteria	●							
Record demographic data	●							
Record social and behavioral data(1)	●						●	●
Medical history	●							
CMV serostatus(2) (~3.5 to 5 mL)	●							
Blood sample(3) - for humoral immunity (10 mL) - for molecular biology (4 mL)	●		●		●		●	●
Blood sample for gene expression signature (qPCR or RNA microarray; 2.5 mL)***(4)***	●						●	
Urine sample ~~(4)~~ ***(5)*** (~10 mL)	●	●	●	●	●	●	●	●
Saliva	●	●	●	●	●	●	●	●
Subject sample collection booklet distribution~~(5)~~ ***(6)***	○							
Check elimination criteria	●		●		●		●	●
Record any concomitant medication/vaccination~~(6)~~ (7)	●		●		●		●	●
Reporting of SAEs related to study participation	●		●		●		●	●
Reporting of pregnancy and outcome	●		●		●		●	●
Study conclusion								●

● is used to indicate a study procedure that requires documentation in the individual eCRF.
○ is used to indicate a study procedure that does not require documentation in the individual eCRF
~~The double line border indicates that an interim analysis will be performed on clean data each year (Month 12 and Month 24).~~
(1) Social and behavioral data will be recorded at inclusion (Visit 1 [Month 0]), at Visit 4 (Month 12), Visit 7 (Month 24) and Visit 10 (Month 36).
(2) CMV serostatus, for allocation to a study group (S+ or S-), will be determined by the local laboratory as per local practices or by a GSK designated and validated central laboratory.
(3) CMV serostatus for endpoints analysis will be determined by GSK Biologicals or designated laboratory at Visit 1 (Month 0).
(4) These blood samples will be collected until approval of protocol amendment 2. A year 1 (at M0 and M12) dataset will be used for analysis of gene expression signature in a subset of subjects
(5) Prepared from the 20 mL urine samples collected from the subject.
(6) The sample collection booklet will instruct the subjects of the study procedures to be performed at the Sample Collection Visits. The sampling can be done through self-collection, through a home-visiting nurse or the subject may be asked to come back to the study center.
(7) Only concomitant medication/vaccination related to inclusion/exclusion and elimination criteria will be recorded.
.

3. OBJECTIVES

3.1. Primary objective

- To estimate the incidence of CMV secondary infections in seropositive adolescent females.

3.2. Secondary objectives

- To estimate the incidence of CMV primary infections in seronegative adolescent females.

3.3. Tertiary objectives

- To develop assays that allow the differentiation of CMV secondary infections caused either by reinfection or by re-activation, in CMV seropositive adolescent females.
- To assess the diagnostic value of different assays to evaluate seroconversion in blood and the detection of CMV DNA in different body fluids.
- To describe the proportion of secondary infection caused either by re-infection or re-activation.
- To collect samples for immunological and biological disease-related exploratory assays.
- To explore socio-demographic or behavioral factors associated with CMV infection.
- In case of pregnancies, to document the birth prevalence of CMV congenital infections.
- To evaluate the concordance between screening testing data at local laboratories (or alternatively, central laboratories) and the GSK Biologicals or designated laboratory testing data.

4. ENDPOINTS

4.1. Primary endpoints

- Occurrence of CMV secondary infections determined in all seropositive subjects on samples collected during the 4-month Site Visits until study conclusion:
  - Anti-CMV tegument protein IgG antibody concentration in serum (ELISA).
  - Number of CMV DNA copies (pp65 or other genes) in urine (qPCR).

4.2. Secondary endpoints

- Occurrence of CMV primary infection determined in all seronegative subjects on samples collected during the 4-month Site Visits until study conclusion:
  - Anti-CMV tegument protein IgG antibody concentration in serum (ELISA).

4.3. Tertiary endpoints

- Further characterization of primary and /or secondary CMV infections in a subset of subjects on selected samples of the Sample Collection Visits*.
  - Number of CMV DNA copies (pp65 or other genes) in urine (qPCR).
  - Number of CMV DNA copies (pp65 or other genes) in saliva (qPCR).

* In case of a positive sample at a 4-month Site Visit time point, testing will be done on selected samples collected during the Sample Collection Visit. The selection of samples to be tested will be based on the qPCR results obtained following the Site visits.

- Further characterization of primary and /or secondary CMV infections in a subset of subjects on all available samples of the 4-month Site Visits:
  - Anti-CMV tegument protein IgG antibody avidity index in serum (ELISA).
  - Anti-gB IgG antibody avidity index in serum (ELISA).
  - Anti-gB IgG antibody concentration in serum (ELISA).
  - Anti-CMV IgM antibody concentration in serum (ELISA).
  - Number of CMV DNA copies (pp65 or other genes) in saliva, blood (qPCR).
  - Assessment of anti-CMV neutralizing antibodies on different target cells and/or other virus strains.
- Development of assays that will allow differentiating re-infection from re-activation from primary infection in a subset of subjects
  - Characterization of CMV strains by genotyping/sequencing.
  - Exploring the CMV strain specific antibody profile using peptide microarrays.
- Assessment of the expression of host genes* using techniques such as qPCR or mRNA microarray in a subset of subjects

  * excluding genes related to hereditary characteristics of the subject.
- Evaluation of the impact of demographic, social or behavioral factors upon CMV infection.
- Occurrence of congenital CMV infection in newborns of subjects who become pregnant during the study:
  - Evidence of CMV DNA in urine and/or saliva of newborns within 10 days of delivery by using qPCR (pp65 or other genes).

5. ANALYSIS SETS

5.1. All Enrolled Set ("Total cohort" in protocol)

All screened subjects who provide informed consent and provide demographic and/or other baseline screening measurements, regardless of the subject's randomization and vaccination status in the trial, and receive a subject ID.

Demography and baseline characteristics tables as well as subject listings will be produced on the All Enrolled Set.

5.2. Per Protocol Set (PPS) ("ATP cohort" in protocol)

All subjects in the All Enrolled Set who:

- Have serology results available.
- Have no major protocol deviations leading to exclusion
- Are not excluded due to other reasons

5.3. Other Analysis Sets

All subjects in the All Enrolled Set who consent to provide exploratory assay data at relevant time points will be included in an exploratory assay subset. Subgroup analysis sets will be included based on two-fold and four-fold increases in anti-CMV tegument protein IgG antibodies in serum among CMV seropositive subjects. The analyses of the tertiary endpoints will be detailed in a separate statistical analysis plan.

5.4. Criteria for eliminating data from ATP

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

Elimination from According to Protocol (ATP)

Code 900 (invalid informed consent or fraud data) will be used for identifying subjects eliminated from PPS.

5.4.1.1. Excluded subjects

A subject will be excluded from the ATP analysis under the following conditions

Code	Condition under which the code is used
1040	Administration of concomitant vaccine(s) forbidden in the protocol (see also eligibility criteria)

	CDM responsibility from review of individual data listings
2010	Protocol violation linked to the inclusion/exclusion criteria including age and excluding codes mentioned below.
2040	Administration of any medication forbidden by the protocol *Responsibility of CRDL following review of individual data listings*
2050	Underlying medical condition forbidden by the protocol *Responsibility of CRDL following review of individual data listings*
2070	Concomitant infection not related to the vaccine which may influence immune response *Responsibility of CRDL following review of individual data listings*
2090	Blood samples taken but: non-compliance with blood sampling schedules (dates of BS not corresponding to adapted protocol intervals or unknown BS)
When a visit and/or a vaccination and/or a blood sample is Not Done/Performed, only the code 2100 is attributed to the subject.	
2100	Serological results not available (including lost samples, blood sample not done, unable to test, absence of parallelism). elimination code if **ALL** are missing
2500	Other: For seropositive subjects only: Serological results not available for anti-CMV tegument protein IgG at the 4 site visits (Visit 1 to Visit 4) and from at least 2 sample collection visits during Year 1
3500	Urine and saliva samples taken but: non-compliance with blood sampling schedules (dates of BS not corresponding to adapted protocol intervals or unknown BS)

5.5. Important protocol deviation not leading to elimination from per-protocol analysis set

The following important protocol deviations will be reported by groups:

- Short follow-up: subjects who missed one or more of the site visits and sample collection visits.

6. STATISTICAL ANALYSES

Note that standard data derivation rules and statistical methods are described in annex 1 and will not be repeated below.

6.1. Demography

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

Demographic (e.g. age at study entry in years, race) and baseline characteristics will be summarized using descriptive statistics.

- Frequency tables will be generated for categorical variables such as race.
- Mean, median, standard error will be provided for continuous data such as age.
- The numbers of withdrawn subjects will be tabulated according to the reason for withdrawal.

6.2. Objectives

6.2.1. Analysis of primary objective

Descriptive statistics will be provided for each primary endpoint at all available time-points and will include (not exhaustive) the following tabulations (with 95% confidence interval [CI]):

- The percentage of CMV seropositive subjects with increase of anti-CMV tegument protein IgG antibodies in serum. An increase will be defined by both a 2-fold or greater increase *and* separately a 4-fold or greater increase in the anti-tegument antibody concentration between subject visits or the last available anti-tegument antibody value if the subject missed a visit or serological results are unavailable.
- The percentage of CMV seropositive subjects with appearance or increase of CMV DNA in urine. A reactivation/reinfection event, based on DNA in the urine, will be defined as any one of the following : (1) the appearance of any detectable CMV DNA by PCR in a sample if CMV DNA was not detected in the previous sample for that subject (2) if the previous sample had CMV DNA detected, but was below the LLOQ for the assay, any value above the LLOQ will be considered a reactivation/reinfection event; and (3) if the previous sample result was above the LLOQ, an increase ≥ 0.5 $\log_{10}$ CMV DNA copies/mL in the subsequent sample will be considered an event. It is expected that CMV-positive subjects may have more than one reactivation/reinfection event during the follow up period in this study.

6.2.2. Analysis of secondary objective

Descriptive statistics will be provided for each secondary endpoint at each time point for which a sample is available and will include (not exhaustive) the following tabulations (with 95% CI):

- The percentage of CMV seronegative subjects with appearance of anti-CMV tegument protein IgG antibodies in serum.

6.2.3. Analysis of Tertiary objective

Descriptive statistics will be provided for the number of CMV DNA copies in urine at each time point for which a sample is available and the percentage of CMV seronegative subjects with presence of CMV DNA in urine will be tabulated (with 95% CI).

For all results of assays performed for the further characterization of primary and/or secondary CMV infections, descriptive statistics will be provided for each assay at all selected time points for which a sample is available.

The analysis of tertiary objectives will be described in an analysis plan amendment (before unblinding) or in an additional analysis request (after unblinding).

For further details on standard business rule for GMT computation, see annex 1.

7. ANALYSIS INTERPRETATION

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

8. CONDUCT OF ANALYSES

8.1. 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**
Analysis of epoch 001	E1_01	Study report			All tables from TFL dated
Analysis of epoch 001	E1_02	Study report			All tables need to be detailed

9. LIST OF FINAL REPORT TABLES, LISTINGS AND FIGURES++++

Listing of TFL and Mock TFLs can be found in the document “CMV 014 (115639) TFL”

10. ANNEX 1 STANDARD DATA DERIVATION RULE AND STATISTICAL METHODS

10.1. Statistical Method References

The exact two-sided 95% CIs for a proportion within a group will be the Clopper-Pearson exact CI [Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika*. 1934;26: 404-413].The standardised asymptotic two-sided 95% CI for the group difference in proportions is based on the method described in the following paper: Robert G. Newcombe, interval estimation for the difference between independent proportions: comparison of eleven methods, *Statist Med.* 1998; 17, 873-890]. The standardised asymptotic method used is the method six.

10.2. Standard data derivation

GSK legacy

- SAS date derived from a character date: In case day is missing, 15 is used. In case day& month are missing, 30June is used.

- Onset day for an event (ae, medication, vaccination, …): The onset day is the number of days between the last study vaccination & the onset/start date of the event. This is 0 for an event starting on the same day as a vaccination. See SAS date derived in case the start date of the event is incomplete.

- Duration: Duration of an event is expressed in days. It is the number of days between the start & the stop dates + 1. Therefore duration is 1 day for an event starting & ending on the same day.

- Association of an event to the primary epoch: An adverse event belongs to the primary epoch, if the onset date is before and excluding Visit 5 or the last contact date, whichever is coming first.

- Not applicable

- Age: Age at the reference activity, computed as the number of units between the date of birth and the reference activity. Note that due to incomplete date, the derived age may be incorrect by 1 month when month is missing from the birthdate. This may lead to apparent inconsistency between the derived age and the eligibility criteria/the age category used for randomization.

- Conversion of weight to kg
  The following conversion rule is used:
  - Weight in Kilogramm= weight in Pounds / 2.2
  - Weight in Kilogramm =weight in oncs / 35.2

  The result is rounded to 2 decimals.

- Conversion of height to cm
  The following conversion rule is used:
  - Height in Centimetres = Height in Feet * 30.48
  - Height in Centimetres = Height in Inch * 2.54

  The result is rounded to the unit (ie no decimal).

- Conversion of temperature to °C
  The following conversion rule is used:
  - Temperature in °Celsius = ((Temperature in °Fahrenheit -32) *5)/9

  The result is rounded to 1 decimal.

- For a given subject and given immunogenicity measurement, missing or non-evaluable measurements will not be replaced. Therefore, an analysis will exclude subjects with missing or non-evaluable measurements.

- The Geometric Mean Concentrations/Titres (GMC/Ts) calculations are performed by taking the anti-log of the mean of the log titre transformations. Antibody titres below the cut-off of the assay will be given an arbitrary value of half the cut-off of the assay for the purpose of GMT calculation. The cut-off value is defined by the laboratory before the analysis and is described in the protocol.

- A seronegative subject is a subject whose antibody titre is below the cut-off value of the assay. A seropositive subject is a subject whose antibody titre is greater than or equal to the cut-off value of the assay.

- The assay cut-off is the value under which there is no quantifiable result available. For an assay with a specific 'cut_off' , numerical immuno result is derived from a character field (rawres):

  - If rawres is 'NEG' or '-' or '(-)', numeric result= cutt_off/2,
  - if rawres is 'POS' or '+' or '(+)', numeric result = cut_off,
  - if rawres is '< value' and value<=cut_off, numeric result =cut_off/2,
  - if rawres is '< value' and value>cut_off, numeric result =value,
  - if rawres is '> value' and value<cut_off, numeric result =cut_off/2,
  - if rawres is '> value' and value>=cut_off, numeric result =value,
  - if rawres is '<= value' or '>= value' and value<cut_off, numeric result =cut_off/2,
  - if rawres is '<= value' or '>= value' and value>=cut_off, numeric result =value,
  - if rawres is a value < cut_off, numeric result = cut_off/2,
  - if rawres is a value >= cut_off, numeric result = rawres,
  - if rawres is a value >= cut_off, numeric result = rawres,
  - else numeric result is left blank.

- All CI computed will be two-sided 95% CI.

- Not applicable; only SAE Listings will be provided

Number of decimals displayed:

The following decimal description from the decision rules will be used for the demography, immunogenicity and safety/reactogenicity.

Display Table	Parameters	Number of decimal digits
Demographic characteristics	Mean, median age	1
Demographic characteristics	SD (age)	1
Immunogenicity	Ratio of GMT/C	2
Reactogenicity	Mean, Min, Q1, Median, Q3, Max for duration	1
All summaries	% of count, including LL & UL of CI	1
All summaries	% of difference, including LL & UL of CI	2
All summaries	p-value	3

11. ANNEX 2: SUMMARY ON ELIMINATION CODES

Not applicable

12. ANNEX 3: STUDY SPECIFIC MOCK TFL

Mock TFLs can be found in the document “CMV 014 (115639) TFL”