STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 3 Multicenter, Open-Label Study to Investigate the Efficacy and Safety of Sofosbuvir/Velpatasvir Fixed-Dose Combination for 12 Weeks in Subjects with Chronic HCV Infection and Compensated Cirrhosis
Name of Test Drug:	SOF/VEL FDC
Study Number:	GS-US-342-5531
Protocol Version (Date):	Original: 03 July 2019
Analysis Type:	Posttreatment Week 12 Analysis Posttreatment Week 24 Analysis
Analysis Plan Version:	1.0
Analysis Plan Date:	12 April 2021
Analysis Plan Author(s):	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TABLES

LIST OF ABBREVIATIONS

AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BMI	body mass index
CCG	eCRF Completion Guidelines
CI	confidence interval
COVID-19	Coronavirus disease 2019
CRF	case report form
CSR	clinical study report
DMC	data monitoring committee
ECG	electrocardiogram
eGFR	estimated glomerular filtration rate
ET	early termination
EOT	End of treatment
FAS	Full Analysis Set
FDC	Fixed-dose combination
FU	Follow-up
GT	Genotype
HCV	hepatitis C virus
HLGT	high-level group term
HLT	high-level term
INR	international normalized ratio
LLT	lower-level term
LLOQ	Lower limit of quantitation
LOQ	Limit of quantitation
MedDRA	Medical Dictionary for Regulatory Activities
NLP	Natural Language Processing
Peg-INF	Pegylated interferon
PK	pharmacokinetics
PT	preferred term
Q1, Q3	first quartile, third quartile
RNA	Ribonucleic acid
RR	electrocardiographic interval representing the time measurement between the R wave of one heartbeat and the R wave of the preceding heartbeat
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SE	Standard error
SOC	system organ class

SOF	sofosbuvir
SVR	Sustained Virologic Response
SVRx	sustained virologic response x weeks after cessation of treatment
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TND	Target not detected
ULN	upper limit of normal
VEL	velpatasvir
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) in the clinical study report (CSR) for Study GS-US-342-5531. This SAP is based on the study protocol dated 03 July 2019 and the electronic case report form (eCRF). The SAP will be finalized before SVR 12 database finalization. Any changes made after the finalization of the SAP will be documented in the CSR.

1.1. Study Objectives

The primary objective of this study is as follows:

- To evaluate the antiviral efficacy of therapy with sofosbuvir/velpatasvir (SOF/VEL) fixed-dose combination (FDC) for 12 weeks as measured by the proportion of subjects with sustained virologic response 12 weeks after cessation of treatment (SVR12)
- To evaluate the safety and tolerability by review of the accumulated safety data

The secondary objectives of this study are as follows:

- To determine the proportion of subjects who attain SVR at 4 and 24 weeks after cessation of treatment (SVR4 and SVR24)
- To evaluate the proportion of subjects with virologic failure
- To evaluate the emergence of viral resistance to SOF and VEL during treatment and after cessation of treatment

1.2. Study Design

This is a multicenter, open-label study evaluating efficacy and safety of SOF/VEL FDC in subjects with chronic HCV infection and compensated cirrhosis.

Subjects may have HCV of any genotype. Based on the HCV genotype (GT) distribution in Japan, it is expected that approximately 70% of subjects will be GT1, and approximately 30% of subjects will be non-GT1. Subjects may be either treatment-naive or treatment-experienced with interferon (IFN)-based treatments.

Approximately 36 subjects will be enrolled to be treated with SOF/VEL FDC for 12 weeks.

The total time to complete all study visits is approximately 40 weeks (42 weeks for those requiring an extension of the Screening period):

- 28 days (4 weeks) screening period [up to 42 days (6 weeks) for extenuating circumstances may be granted]
- 12 weeks study treatment period
- 24 weeks posttreatment period

The schedule of assessment is provided as an appendix to the SAP (Appendix 1).

1.3. Sample Size and Power

A sample size of approximately 36 compensated cirrhotic Japanese subjects in the study will provide > 80% power to demonstrate superiority to a performance threshold of 78%, assuming an expected SVR12 rate of 96% with 12 weeks of SOF/VEL therapy (based on a 2-sided exact 1-sample binomial test of superiority at a significance level of 0.05)

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee

This study does not have a data monitoring committee (DMC).

2.2. Interim Analyses

2.2.1. Posttreatment Week 12 Analysis (Primary Analysis)

The analysis for the primary endpoint SVR12 will be conducted after all subjects complete the posttreatment Week 12 visit or prematurely discontinue from study. All the safety and efficacy data through the posttreatment Week 12 visit will be cleaned, finalized, and included for the analysis.

2.3. Posttreatment Week 24 Analysis (Final Analysis)

After all subjects have completed the study, outstanding data queries have been resolved, and the database has been cleaned and finalized, the final analysis of the data will be performed.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

Analysis results will be presented using descriptive statistics. For categorical variables, the number and percentage of subjects in each category will be presented; for continuous variables, the number of subjects (n), mean, standard deviation (SD) or standard error (SE), median, first quartile (Q1), third quartile (Q3), minimum, and maximum will be presented.

By-subject listings will be presented for all subjects in the Safety Analysis Set and sorted by subject ID number, visit date, and time (if applicable). Data collected on log forms, such as AEs, will be presented in chronological order within the subject. The treatment group to which subjects were initially assigned will be used in the listings. Age, sex at birth, race, and ethnicity will be included in the listings, as space permits.

3.1. Analysis Sets

Analysis sets define the subjects to be included in an analysis. Analysis sets and their definitions are provided in this section. The number of subjects eligible for each analysis set will be provided. Subjects who were excluded from each analysis set will be summarized or provided in a by-subject listing with reasons for exclusion.

3.1.1. All Enrolled Analysis Set

All Enrolled Analysis Set includes all subjects who received a study subject identification number in the study after screening.

3.1.2. Full Analysis Set

The Full Analysis Set (FAS) includes all enrolled subjects who took at least 1 dose of study drug in this study, which is SOF/VEL FDC. This is the primary analysis set for efficacy analyses.

3.1.3. Safety Analysis Set

The Safety Analysis Set includes all subjects who took at least 1 dose of study drug. This is the primary analysis set for safety analyses. Treatment-emergent data will be analyzed and defined as data collected from the first dose of the study drug through the date of last dose of the study drug plus 30 days.

3.2. Subject Grouping

For analyses based on the All Enrolled Analysis Set, FAS, or demographic and baseline characteristics tables, subjects will be grouped according to the HCV genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b).

3.3. Examination of Subject Subgroups

Subject subsets within each genotype will also be explored by presumed prognostic baseline characteristics for the primary efficacy endpoint, SVR12. The presumed prognostic baseline characteristics include the following:

- age (< 65 years, ≥ 65 years)
- sex (male, female)
- baseline body mass index (BMI) (< 25 kg/m^2, ≥ 25 kg/m^2)
- IL28B (CC, non-CC; with non-CC further broken down to CT, TT)
- baseline HCV RNA (< 800,000 IU/mL, ≥ 800,000 IU/mL)
- baseline alanine aminotransferase (ALT) (≤ 1.5 × upper limit of normal (ULN), > 1.5 × ULN)
- prior HCV treatment experience (treatment naive, treatment experienced)
- prior HCV treatment (pegylated interferon [Peg-IFN], Other) for treatment experienced subjects
- prior HCV treatment response (non-responder, relapse/breakthrough, early treatment discontinuation, met a virologic stopping rule, unknown) for treatment experienced subjects
- study treatment status (completed study treatment, discontinued study treatment)
- adherence to study regimen (< 80%, ≥ 80%)

3.4. Multiple Comparisons

No multiplicity adjustment will be made for testing because only one test will be performed.

3.5. Missing Data and Outliers

3.5.1. Missing Data

In general, missing data will not be imputed unless methods for handling missing data are specified. Exceptions are presented in this document.

For missing last dosing date of study drug, imputation rules are described in Section 3.7.1. The handling of missing or incomplete dates for AE onset is described in Section 7.1.5.2, and for prior and concomitant medications in Section 7.4.

For analyses of categorical HCV RNA data, missing posttreatment HCV RNA data will have the missing data imputed. Missing on-treatment HCV RNA will have missing data imputed up to the time of the last dose (for on-treatment displays). If the study day associated with the last dosing date of the study drug is greater than or equal to the lower bound of a visit window, and the value at the visit is missing, then the value will be imputed. If the study day associated with the last dosing date is less than the lower bound of a visit window, then the on-treatment value at that visit will remain missing.

If an HCV RNA data point is missing and is preceded and followed in time by values that are "< lower limit of quantification (LLOQ) target not detected (TND)," then the missing data point will be set to "< LLOQ TND." If a data point is missing and preceded and followed by values that are "< LLOQ detected," or preceded by "< LLOQ detected" and followed by "< LLOQ TND," or preceded by "< LLOQ TND" and followed by "< LLOQ detected," then the missing value will be set to "< LLOQ detected." In these situations, the data point will be termed a bracketed success; otherwise, the data point will be termed a bracketed failure (ie, ≥ LLOQ detected). If a data point is missing and is not bracketed, the missing data point will also be termed a failure (ie, ≥ LLOQ detected) except for SVR24, which will be imputed according to SVR12 status. Success for SVR12 who have no further HCV RNA measurements collected will be counted as a success for SVR24 due to the high correlation between these 2 endpoints.

For the analyses of continuous HCV RNA efficacy data, when and only when a missing HCV RNA value is imputed as < LLOQ TND or < LLOQ detected according to the imputation rule described above, the corresponding continuous value will be imputed to LLOQ − 1 IU/mL. No other imputation will be performed for continuous HCV RNA data.

3.5.2. Outliers

Outliers will be identified during the data management and data analysis process, but no sensitivity analyses will be conducted. All data will be included in the data analysis.

3.6. Data Handling Conventions and Transformations

By-subject listings will be presented for all subjects in the Safety Analysis Set and sorted by subject ID number, visit date, and time (if applicable) unless otherwise specified. Data collected on log forms, such as AEs, will be presented in chronological order within subject.

Age (in years) on the date of the first dose of study drug and sex at birth will be used for analyses and presentation in listings.

If a subject was never dosed with study drug, then the date the informed consent was signed will be used instead of the first dose date of study drug. For some countries, only birth year is collected on the CRF. In those cases, "01 January" will be used for the unknown birth day and month for the purpose of age calculation, unless age is captured on the CRF.

Non-PK data that are continuous in nature but less than the lower limit of quantitation (LOQ) or above the upper limit of quantitation will be imputed as follows:

- A value that is 1unit less than the limit of quantitation will be used for calculation of descriptive statistics if the datum is reported in the form of "< x" (where x is considered the LOQ). For example, if the values are reported as < 50 and < 5.0, values of 49 and 4.9, respectively, will be used for calculation of summary statistics. An exception for this rule is any value reported < 1. For the values reported as < 1 or < 0.1, a value of 0.9 or 0.09, respectively, will be used for calculation of summary statistics.
- A value that is 1 unit above the LOQ will be used for calculation of descriptive statistics if the datum is reported in the form of "> x" (where x is considered the LOQ). Values with decimal points will follow the same logic as above.
- The LOQ will be used for calculation of descriptive statistics if the datum is reported in the form of "≤ x" or "≥ x" (where x is considered the LOQ).

The COBAS® AmpliPrep/COBAS® TaqMan® HCV Quantitative Test, v2.0 was used to determine HCV RNA results in this study. The LLOQ of the assay is 15 IU/mL.

When the calculated HCV RNA value is within the linear range of the assay, then the result will be report as the "<< numeric value>> IU/mL." This result will be referred to in this document as the numeric result or as "≥ LLOQ detected" for categorical result.

When HCV RNA is not detected, the result is reported as "No HCV RNA detected" or "target not detected". This result will be referred to in this document as "< LLOQ target not detected" or "< LLOQ TND."

When the HCV RNA IU/mL is less than LLOQ of the assay, the result is reported as "< 15 IU/mL HCV RNA detected." This result will be referred to in this document as "< LLOQ detected."

The overall category of HCV RNA < LLOQ includes "< LLOQ TND" and "< LLOQ detected."

For numerical HCV RNA data, values below LLOQ will be set to the LLOQ – 1 IU/mL (ie, 14 HCV RNA IU/mL). HCV RNA values returned as "No HCV RNA detected" will also be set to 14 IU/mL.

For selected analyses, HCV RNA data (IU/mL) will be transformed to the logarithmic (base 10) scale ($\log_{10}$ IU/mL).

3.7. Visit Windows

3.7.1. Definition of Study Day

Study day is the day relative to the date of the first dose of study drug. Study Day 1 will be defined as the day of first dose of study drug administration.

Study day will be calculated from the date of first dose of study drug administration and derived as follows:

- For postdose study days: Assessment Date – First Dosing Date + 1
- For days prior to the first dose: Assessment Date – First Dosing Date

The last dose date for the study drug will be the end date on study drug administration eCRF for the record where the "subject permanently discontinued" flag is 'Yes'.

If there are subjects for whom the date of last study drug is unknown due to the reason that the subject was lost to follow-up and not able to be contacted, the date of last dose will be estimated using the maximum of nonmissing study drug start or stop dates, visit dates, and laboratory collection dates (posttreatment visits and unscheduled visits are not included).

3.7.2. Analysis Visit Windows

Subject visits might not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows.

In general, the baseline value will be the last nonmissing value on or prior to the first dose date of study drug.

HCV RNA, vital signs, and safety laboratory data collected up to the last dose date + 3 days are considered to be on-treatment data and HCV RNA, vital signs, and safety laboratory data collected after the last dose date + 3 days are considered posttreatment data. The analysis windows for on-treatment HCV RNA, vital signs and safety laboratory data are provided in Table 1.

Table 1. Analysis Visit Windows for On-treatment HCV RNA, Vital Signs and Safety Laboratory Data

Nominal Visit	HCV RNA and Safety Laboratory Data			Vital Signs Data		
	Nominal FU Day	Lower Limit	Upper Limit	Nominal FU Day	Lower Limit	Upper Limit
Baseline	1	(none)	1	1	(none)	1
Week 1	N/A	N/A	N/A	7	2	11
Week 2	14	2	21	14	12	21
Week 4	28	22	42	28	22	42
Week 8	56	43	70	56	43	70
Week 12	84	71	≥ 85	84	71	≥ 85

HCV RNA, vital sign, and safety laboratory data collected after the last dose date + 3 days will be assigned to the posttreatment follow-up (FU) visit. Visit windows will be calculated from the last dose date (ie FU Day = collection date minus the last dose date) as shown in Table 2.

Table 2. Analysis Visit Windows for Posttreatment HCV RNA, Vital Signs and Safety Laboratory Data

Nominal FU[a] Visit	HCV RNA			Vital Signs and Safety Laboratory Data[b]		
	Nominal FU Day	Lower Limit	Upper Limit	Nominal FU Day	Lower Limit	Upper Limit
FU-4	28	21	69	28	4	30
FU-12	84	70	146	NA	NA	NA
FU-24	168	147	210	NA	NA	NA

a FU-x visit = posttreatment Week-x follow-up visit.
b Days are calculated based on last dose date.

ECG data collected up to the last dose date + 3 days are considered to be on-treatment data. Qualitative assessments of whether the ECG is normal or abnormal will be assessed for on-treatment data based on the visit windows shown in Table 3.

Table 3. Analysis Visit Windows for On-treatment ECG Data

Nominal Visit	Nominal Day	Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 12	84	2	(none)

Note: ECGs are to be collected at screening, baseline, Week 1, and Week 12 or Early Termination. For purposes of analysis, baseline value will be the last available value prior to the first dose of study drug and end of treatment value will be the last available value on or prior to the last dose date + 3 days.

3.7.3. Selection of Data in the Event of Multiple Records in an Analysis Visit Window

Depending on the statistical analysis method, single values may be required for each analysis window. For example, change from baseline by visit usually requires a single value, whereas a time-to-event analysis would not require 1 value per analysis window.

If multiple valid, nonmissing measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- For baseline, the last available record on or prior to the date of the first dose of study drug will be selected. If there are multiple records with the same time or no time recorded on the same day, the average (arithmetic mean) will be used for the baseline value. If multiple ECG measurements occur on the same day prior to first dose of any study drug, the average will be used as baseline for continuous data, regardless of the timing of these multiple ECG measurements.

- For postbaseline values:
  - The record closest to the nominal day for that visit will be selected except for HCV RNA posttreatment follow-up visits, for which the latest record in the analysis window will be selected.
  - If there are 2 records that are equidistant from the nominal day, the later record will be selected.
  - If there is more than 1 record on the selected day, the average will be taken, unless otherwise specified.
- If multiple valid nonmissing categorical observations exist in a window, records will be selected as follows:
  - For baseline, the last available record on or prior to the date of the first dose of study drug will be selected. If there are multiple records with the same time or no time recorded on the same day, the value with the lowest severity will be selected (eg, normal will be selected over abnormal). If multiple ECG measurements occur on the same day prior to the first dose of any study drug, the value with the lowest severity will be selected regardless of the timing of these multiple ECG measurements.
  - For postbaseline visits, follow the same rules described above for postbaseline numeric observations, except that if there are multiple records on the same day, the most conservative value will be selected (eg, abnormal will be selected over normal).

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

A summary of subject enrollment will be provided for each investigator. The summary will present the number and percentage of subjects in the Safety Analysis Set. For each column, the denominator for the percentage calculation will be the total number of subjects analyzed for that column.

A summary of subject disposition will be provided by HCV genotype (GT1, GT1 broken down [1a, 1b], GT2, GT2 broken down [GT2a/2c, GT2b]) and overall. This summary will present the number of subjects screened, the number of subjects enrolled, the number of subjects enrolled but never treated, and the number and percentage of subjects in each of the categories listed below.

- In Safety Analysis Set
- In FAS
- Completed study treatment
- Did not complete study treatment with reasons for premature discontinuation of study treatment
- Completed study follow-up
- Did not complete the study follow-up with reasons for premature discontinuation of study follow-up

Among subjects who completed study treatment and who discontinued study treatment, the number and percentage of subjects will be summarized for:

- Who had no HCV posttreatment Week 4 assessment and thereafter (No HCV FU-4 Assessment and thereafter)
- Who had HCV posttreatment Week 4 assessment but no HCV posttreatment Week 12 and thereafter (With HCV FU-4 but No FU-12 and thereafter)

If a subject did not have any HCV RNA assessment ≥ 21 days after the last dose of any study drug (ie, lower bound of FU-4 visit for HCV RNA data), the subject is categorized as having "No HCV FU-4 and thereafter." If a subject had the HCV FU-4 assessment but did not have any HCV RNA assessment ≥ 70 days after the last dose of any study drug (ie, lower bound of FU-12 visit for HCV RNA data), the subject is categorized as having "With HCV FU-4 but No FU-12 and thereafter."

In addition, the total number of subjects who were enrolled, and the number of subjects in each of the disposition categories listed above will be depicted by a flowchart.

The following by-subject listings will be provided by subject identification (ID) number in ascending order to support the above summary tables:

- Disposition for subjects who complete study treatment and study follow up
- Disposition for subjects who did not complete study treatment and/or study follow up with reasons for premature discontinuation of study treatment and/or study
- Lot number and kit ID (if applicable)

4.2. Extent of Exposure

Extent of exposure to study drug will be examined by assessing the total duration of study drug exposure and the level of adherence to the study drug regimen specified in the protocol.

4.2.1. Duration of Exposure to Study Drug

Total duration of exposure to study drug will be defined as last dose date minus first dose date plus 1, regardless of any temporary interruptions in study drug administration, and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks).

The total duration of exposure to study drug will be summarized using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) and using the number of subjects (ie, cumulative counts) and percentage of subjects exposed through the following time periods: baseline (Day 1), Week 1 (Day 7), Week 2 (Day 14), Week 4 (Day 28), Week 6 (Day 42), Week 8 (Day 56), Week 10 (Day 70), and Week 12 (Day 84). A 3-day window is applied to the last planned on-treatment visit to match with the protocol-specified visit window, ie, the number of subjects exposed through Week 12 will be calculated as the number of subjects who were exposed to study drug for at least 81 days. Summaries will be provided for the Safety Analysis Set.

4.2.2. Adherence to Study Drug

The presumed total number of tablets administered to a subject will be determined by the data collected on the drug accountability CRF using the following formula:

$$\text{Total Number of Doses Administered} = \left(\sum \text{No. of Tablets Dispensed}\right) - \left(\sum \text{No. of Tablets Returned}\right)$$

The level of adherence to the study drug regimen will be assessed based on the total amount of study drug administered relative to the total amount of study drug prescribed at baseline.

The level of adherence will be expressed in percentage using the following formula:

$$\text{Level of Adherence (\%)} = \left(\frac{\text{Total Amount of Study Drug Administered}}{\text{Total Amount of Study Drug Prescribed at baseline}}\right) \times 100$$

Note: If calculated adherence is greater than 100%, the result will be set to 100%.

In this study, the total amount of SOF/VEL (400/100 mg) prescribed for 12 weeks would require 84 tablets for each subject.

Subjects who prematurely discontinue study drug for lack of efficacy (ie, virologic failure) will have the total amount of study drug prescribed calculated up to the first date when virologic failure criteria were met. For virologic failure confirmed by 2 consecutive measurements, the date of the first measurement will be used. If study drug bottles are dispensed on or after the subject first met virologic failure criteria, these bottles will not be included in the calculation of adherence. If a bottle is dispensed and the bottle is returned empty, then the number of tablets returned will be entered as zero. If a bottle is dispensed but not returned (missing), the number of tablets taken from that bottle will be counted as zero.

Descriptive statistics for the level of adherence (n, mean, SD, median, Q1, Q3, minimum, and maximum) with the number and percentage of subjects belonging to adherence categories (eg, [< 80%, ≥ 80 to < 90%, ≥ 90%]) will be provided for the Safety Analysis Set. Categorical displays also will be provided for the number of subjects who are at least 80% adherent to their drug regimen (ie, adherence is ≥ 80% for the study drugs).

No inferential statistics will be provided for duration of exposure and adherence to study drug.

A separate by-subject listing of study drug administration and drug accountability will be provided by subject ID number (in ascending order) and visit (in chronological order).

4.3. Protocol Deviations

Subjects who did not meet the eligibility criteria for study entry but enrolled in the study, will be summarized. The summary will present the number and percentage of subjects who did not meet at least 1 eligibility criterion and the number of subjects who did not meet specific criteria based on the All Enrolled Analysis Set. A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected.

Protocol deviations occurring after subjects entered the study are documented during routine monitoring. The number and percentage of subjects with important protocol deviations by deviation category (eg, nonadherence to study drug, violation of select inclusion/exclusion criteria) will be summarized for the All Enrolled Analysis Set.

4.4. Assessment of COVID-19 Impact

This study was ongoing during the novel coronavirus (2019 nCOV [COVID-19]) pandemic which has caused a disruption in the regular visit schedules for this study. Some subjects were unable to attend onsite visits due to shelter in place guidelines, site closures, or other reasons. This section provides how to handle special situations due to COVID-19 in the analysis.

Adverse events due to COVID-19 will be included in AE analyses if applicable.

4.4.1. Study Drug or Study Discontinuation Due to COVID-19

A by-subject listing of reasons for premature study drug or study discontinuation due to COVID-19 will be created for the enrolled analysis set.

4.4.2. Protocol Deviations Due to COVID-19

Similar summary as described in protocol deviations section will be performed for important protocol deviations due to COVID-19 for the enrolled analysis set.

A by-subject listing will be provided for subjects with important protocol deviation related to COVID-19. A separate listing will be provided for subjects with non-important protocol deviation related to COVID-19.

4.4.3. Missed and Virtual Visits due to COVID-19

A by-subject listing of subjects with missed or virtual visits due to COVID-19 will be provided by subject ID number in ascending order.

Information regarding missed or virtual visits due to COVID-19 was collected as free text in the CRF comment fields. The determination of missing or virtual visits due to COVID-19 was done using Natural Language Processing (NLP) to search the CRF comment fields. A brief description of the natural language processing algorithm developed for determining missed and virtual visits based on CRF comment text is provided in Appendix 2.

5. BASELINE CHARACTERISTICS

5.1. Demographics

Subject demographic variables (i.e. age, sex, race, and ethnicity) will be summarized by HCV genotype (GT1, GT1 broken down [1a, 1b], GT2, GT2 broken down [GT2a/2c, GT2b], non-GT1) and overall using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) for age and BMI (kg/m^2), and using the numbers and percentages of subjects for age categories (< 65 years, ≥ 65 years), BMI categories (< 25 kg/m^2, ≥ 25 kg/m^2), sex, race, and ethnicity. Age is calculated in years at the date of initial study drug administration. If a subject did not receive study drug after enrollment, the subject's age will be calculated from the date that the subject signed the informed consent form. The summary of demographic data will be provided for the Safety Analysis Set.

A by-subject demographic listing, which includes the date the informed consent was signed, will be provided by subject ID number in ascending order.

5.2. Other Baseline Characteristics

Other baseline characteristics include:

- body mass index (BMI; in kg/m^2) as a continuous variable and as categories (< 25 kg/m^2, ≥ 25kg/m^2)
- HCV genotype/subgenotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b)
- IL28B (CC, non-CC; with non-CC further broken down to CT, TT)
- baseline HCV RNA (log_{10} IU/mL) as a continuous variable and as categories (< 800,000 IU/mL, ≥ 800,000 IU/mL)
- baseline ALT (U/L) as a continuous variable and as categories (≤ 1.5 × ULN, > 1.5 × ULN)
- prior HCV treatment experience (treatment naive, treatment experienced)
- prior HCV treatment response (non-responder, relapse/breakthrough, early treatment discontinuation, met a virologic stopping rule, unknown) for treatment experienced subjects
- prior HCV treatment for treatment experienced subjects (Peg-IFN, Other)
- estimated glomerular filtration rate (eGFR) using the Cockcroft-Gault equation

  $eGFR_{CG}$ (mL/min) = [(140 – age (yrs)) × weight (kg) × (0.85 if female)] / (serum creatinine (mg/dL) × 72), where weight is total body mass in kilograms.

These baseline characteristics will be summarized by HCV genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b) and the study overall using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) for continuous variables and using the numbers and percentages of subjects for categorical variables. The summary of baseline characteristics will be provided for the Safety Analysis Set.

A by-subject listing of other baseline characteristics will be provided by subject ID number in ascending order. The type of assay used to determine the HCV genotype will also be displayed in this listing.

A separate by-subject data listing for cirrhosis determination will be provided for safety analysis set.

A separate by-subject data listing for prior HCV treatment and response will be provided for all subjects. The listing will display the prior HCV treatment experience for all subjects as well as the prior HCV regimen, the treatment duration, and the prior HCV treatment response for treatment experienced subjects.

5.3. Medical History

General medical history data will be collected at screening and a by-subject listing will be provided for the Safety Analysis Set.

6. EFFICACY ANALYSES

6.1. Primary Efficacy Endpoint

6.1.1. Definition of the Primary Efficacy Endpoint

The primary efficacy endpoint is SVR12 defined as HCV RNA < LLOQ 12 weeks after discontinuation of study drug in the FAS. The COBAS® AmpliPrep/COBAS® TaqMan® HCV Quantitative Test, v2.0 will be used to measure HCV RNA. The LLOQ for this assay is 15 IU/mL.

6.1.2. Statistical Hypothesis for the Primary Efficacy Endpoint

In the primary efficacy analysis, the SVR12 rate in FAS will be compared to the pre-specified efficacy threshold of 78% using a 2-sided exact 1-sample binomial test at the 0.05 significance level. The null (H0) and alternative (H1) hypotheses used to assess the superiority of SOF/VEL relative to the performance goal of 78% as follows:

H0: SVR12 rate = 78%

H1: SVR12 rate ≠ 78%

In addition, a point estimate with a 2-sided 95% exact confidence interval (CI) using the binomial distribution (Clopper-Pearson method) {Clopper 1934} will be constructed for the SVR12 rate.

6.1.3. Primary Analysis of the Primary Efficacy Endpoint

The 2-sided 1-sample binomial test will be used to test the statistical hypothesis described above. The 2-sided 95% exact CI based on the Clopper-Pearson method will be provided for the SVR12 rate.

6.1.4. Subgroup Analysis of the Primary Efficacy Endpoint

The point estimates and the 2-sided 95% exact CIs of the SVR12 rates will be displayed by genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b) for each subgroup outlined in Section 3.3.

The SVR rates will be summarized by categories of early viral response to explore possible early on-treatment predictors of SVR12. The relationship between SVR12 and study drug interruption may also be explored.

A Forest plot will graphically present the point estimates and the 2-sided 95% exact CIs of the SVR12 rates by HCV genotype for each of the subgroups.

6.2. Secondary Efficacy Endpoints

6.2.1. Definition of Secondary Efficacy Endpoints

Secondary efficacy endpoints include the following:

- The proportion of subjects who attain HCV RNA < LLOQ at 4 and 24 weeks after cessation of treatment (SVR4 and SVR24)
- The proportion of subjects with virologic failure as the following:
  - On-treatment virologic failure
    - HCV RNA ≥ LLOQ after having previously had HCV RNA < LLOQ, while on treatment, confirmed with 2 consecutive values (note, second confirmation value can be posttreatment), or last available on-treatment measurement with no subsequent follow up values (ie, breakthrough)
    - > 1 $\log_{10}$ IL/mL increase in HCV RNA from nadir while on treatment, confirmed with 2 consecutive values (note, second conformation value can be posttreatment), or last available on-treatment measurement with no subsequent follow up values (ie, rebound)
    - HCV RNA persistently ≥ LLOQ through 8 weeks of treatment (ie, nonresponse)
  - Relapse
    - HCV RNA ≥ LLOQ during the posttreatment period having achieved HCV RNA < LLOQ at EOT, confirmed with 2 consecutive values or last available posttreatment measurement

6.2.2. Analysis Methods for Secondary Efficacy Endpoints

For analyses of SVR4 and SVR24, subjects will be assigned a value at each visit based on the analysis visit windows specified in Section 3.7.2. Missing values will be imputed based on the categorical imputation rules described in Section 3.5.1. The 2-sided 95% exact CI based on Clopper-Pearson method will be provided for the proportion of subjects who attain SVR4 and SVR24 by HCV genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b). The overall category for "HCV RNA < LLOQ" will be split into the following 2 subcategories: "< LLOQ TND" for subjects with target not detected and "< LLOQ detected" for subjects with < LLOQ in tabular displays.

For the virologic failure analysis, a summary table of the number and percentage of subjects with SVR12, virologic failure, and Other will be created. This summary will be performed broken down by HCV genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b). All subjects who achieve SVR12 will be categorized as SVR12. Virologic failure will be descriptively summarized as "on-treatment virologic failure" and relapse (which will be broken down by study drug completed yes/no).

Subjects who do not achieve SVR12 and do not meet criteria for virologic failure will be categorized as "Other." The denominator for relapse will be the number of subjects who had HCV RNA < LLOQ on their last observed on-treatment HCV RNA measurement; otherwise, the denominator will be the number of subjects in the FAS. Virologic outcomes will also be provided by prior HCV treatment experience on FAS overall and by HCV genotype.

A concordance table between SVR12 and SVR24 will be provided by HCV genotype (GT1 with GT1 broken down as GT1a and GT1b; non-GT1 with non-GT1 broken down as GT2, GT2a/2c and GT2b) and overall on FAS who had observed HCV RNA measurements at both posttreatment Week 12 and posttreatment Week 24 visits.

In addition, a summary table of the number and percentage of subjects with HCV RNA < LLOQ and ≥ LLOQ at the posttreatment follow-up visit (observed and imputed, with reasons for imputed) will be provided for each posttreatment follow-up visit overall and by HCV genotype. The 2-sided 95% Clopper-Pearson exact CIs will be presented for the overall proportion of subjects with HCV RNA < LLOQ.

CCI [redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

6.4. Changes from Protocol-specified Efficacy Analyses

There are no planned changes from protocol-specified efficacy analyses.

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

Clinical and laboratory adverse events (AEs) will be coded using the current version of MedDRA System organ class (SOC), high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lower-level term (LLT) will be provided in the AE dataset.

7.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, or 4 according to toxicity criteria specified in the protocol. The severity grade of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. the most severe will be considered (for sorting purpose only) in data presentation.

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE case report form (CRF) to the question of "Related to Study Treatment." Events for which the investigator did not record the relationship to study drug will be considered to be related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing from that captured on the CRF.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if the AEs met the definitions of SAEs that were specified in the study protocol. SAEs captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Drug Safety and Public Health Department before data finalization.

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as 1 or both of the following:

- Any AEs with an onset date on or after the study drug start date and no later than 30 days after permanent discontinuation of study drug
- Any AEs leading to premature discontinuation of study drug

7.1.5.2. Incomplete Dates

If the onset date of AE is incomplete, then the month and year (or year alone if month is not recorded) of onset determine whether an AE is treatment emergent, as long as the AE stop date is not prior to the first dose date of study drug. The event is considered treatment emergent if both of the following 2 criteria are met:

- The AE onset and end dates are the same as or after the month and year (or year) of the first dose date of study drug
- The AE onset date is the same as or before the month and year (or year) of 30th day after the date of the last dose of study drug

An AE with a completely missing onset and stop dates, or with the onset date missing and a stop date later than the first dose date of study drug, will be considered to be treatment-emergent.

7.1.6. Summaries of Adverse Events and Deaths

A brief high-level summary of TEAEs will be provided by the number and percentage of subjects who had the following: any AE, any AE of Grade 3 or above, any AE of Grade 2 or above, any treatment-related AE, any treatment-related AE of Grade 3 or above, any treatment-related AE of Grade 2 or above, any SAE, any treatment-related SAE, any AE that led to premature discontinuation of the study drug, any AE that led to interruption of the study drug. All deaths (including those that are treatment emergent and those that are not treatment emergent) observed during the study will also be summarized and included in this table.

Adverse event summaries will provide the number and percentage of subjects with TEAEs by SOC and PT on the Safety Analysis Set as follows:

- All AEs
- AEs of Grade 3 or above
- AEs of Grade 2 or above
- All treatment-related AEs
- Treatment-related AEs of Grade 3 or above
- Treatment-related AEs of Grade 2 or above
- All SAEs (including death)
- All treatment-related SAEs
- AEs leading to premature discontinuation of the study drug
- AEs leading to interruption of the study drug

Multiple events will be counted once only per subject in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC and then by PT in order of descending incidence of the pooled treatment groups within each SOC. In summaries by severity grade, the most severe grade will be used for those AEs that occurred more than once in an individual subject during the study.

In addition to the above summary tables, TEAEs will also be summarized by PT only, in order of descending incidence within the pooled treatment group for:

- AEs that occurred in at least 5% of subjects
- AEs of Grade 3 or above
- All treatment-related AEs
- All SAEs
- AEs leading to premature discontinuation of study drug
- AEs leading to interruption of the study drug

In addition to the summaries described above, data listings will be provided for the following:

- All AEs
- AEs of Grade 3 or above
- SAEs
- Deaths
- AEs leading to premature discontinuation of the study drug
- AEs leading to interruption of study drug

7.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set and will include data collected up to last dose of any study drug plus 30 days for subjects who have permanently discontinued study drug. The analysis will be based on values reported in conventional units. When values are below the LOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics. For example, if "< 0.2" was recorded, a value of 0.1 will be used for the purpose of calculating summary statistics; if "< 0.1" was recorded, a value of 0.09 will be used for the purpose of calculating summary statistics.

A by-subject listing for laboratory test results will be provided by subject ID number and visit in chronological order for hematology, serum chemistry, and urinalysis separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher on the Gilead Grading Scale for Severity of AEs and Laboratory Abnormalities will be flagged in the data listings, as appropriate.

No inferential statistics will be generated.

7.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) will be provided for ALT, AST, total bilirubin, alkaline phosphatase, white blood cell (WBC) counts, neutrophils, lymphocytes, hemoglobin, platelets, reticulocytes, and international normalized ratio (INR) as follows:

- Baseline values
- Values at each postbaseline visit
- Change from baseline at each postbaseline visit

A baseline laboratory value will be defined as the final assessment performed on or prior to the date/time of first dose of study drug. Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum will be displayed to reported number of digits, SD to reported number of digits plus 1.

Median (Q1, Q3) of the observed values for ALT, AST, total bilirubin, alkaline phosphatase, hemoglobin, reticulocytes, WBC, neutrophils, lymphocytes, and platelets will be plotted using a line plot by visit.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.7.3.

The number of subjects with hemoglobin < 10 g/dL and < 8.5 g/dL at any postbaseline visits (up to 30 days after the last dose of any study drug) will be summarized.

7.2.2. Graded Laboratory Values

The Gilead Grading Scale for Severity of AEs and Laboratory Abnormalities will be used for assigning toxicity grades to laboratory results for analysis as Grade 0, Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe), or Grade 4 (potentially life threatening). Grade 0 includes all values that do not meet criteria for an abnormality of at least Grade 1. Some laboratory tests have laboratory toxicity criteria for both increased and decreased levels; analyses for each direction (ie, increased, decreased) will be presented separately.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, up to and including the date of last dose of study drug plus 30 days for subjects who permanently discontinued study drug.

If the relevant baseline laboratory value is missing, then any abnormality of at least Grade 1 will be considered treatment emergent.

7.2.2.2. Summaries of Laboratory Abnormalities

Laboratory data that are categorical will be summarized using the number and percentage of subjects in the study with the given response at baseline and each scheduled postbaseline visit.

The following summaries (number and percentage of subjects) for treatment-emergent laboratory abnormalities will be provided by analyte; subjects will be categorized according to the most severe postbaseline abnormality grade for a given analyte:

- Graded laboratory abnormalities
- Grade 3 or above laboratory abnormalities

For all summaries of laboratory abnormalities, the denominator is the number of subjects with nonmissing postbaseline values up to 30 days after last dose for the laboratory parameter of interest.

A by-subject listing of treatment-emergent Grade 3 or above laboratory abnormalities will be provided by subject ID number and visit in chronological order. This listing will include all test results that were collected throughout the study for the analyte of interest, with all applicable severity grades or abnormal flags displayed.

7.3. Body Weight, Height, and Vital Signs

Vital signs (systolic and diastolic blood pressure [mmHg] and pulse [beats/min]) at each visit, and change from baseline at each visit will be summarized for the Safety Analysis Set using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum). The baseline value will be defined as the last available value collected on or prior to the date/time of first dose of study drug. Change from baseline to a postbaseline visit will be defined as the postbaseline value minus the baseline value.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.7.3). No inferential statistics will be generated.

A by-subject listing of vital signs (systolic and diastolic blood pressure [mmHg], pulse [beats/min], respiration [breaths/min], and body temperature [°C]) will be provided by subject ID number and visit in chronological order. In the same manner, a separate by-subject listing of body weight, height, and BMI will be provided.

7.4. Prior and Concomitant Medications

Medications collected at screening and during the study will be coded using the current version of the World Health Organization (WHO) Drug dictionary. The medications will be categorized as prior, concomitant, or both using the following definitions:

- Prior medications: any medications taken prior to the initial study drug dosing date
- Concomitant medications: any medications initially taken on or after the initial study drug dosing date and within the study drug's treatment period (including study drug's therapeutic reach)
- Prior and concomitant medications: any medications taken both prior to and on or after the initial study drug dosing date and within the study drug's treatment period (including study drug's therapeutic reach); or any medications taken prior to the Baseline visit date with a stop date of "continuing"

Concomitant medications will be summarized by preferred name using the number and percentage of subjects for each treatment group. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary of concomitant medications will be ordered by descending frequency of preferred names. For drugs with the same frequency, sorting will be done alphabetically.

Summaries will be based on the Safety Analysis Set. No inferential statistics will be generated.

For purposes of analysis, any medication with a stop date that is on or prior to the initial study drug dosing date or start date that is after the last study drug dosing date will be excluded from a concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the initial study drug dosing date will be excluded from the concomitant medication summary. If a partial start date is entered, then any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing dates will be included in the concomitant medication summary.

All prior and concomitant medications (other than per-protocol study drugs) will be provided in a by-subject listing sorted by subject ID number and medication administration date in chronological order.

7.5. Electrocardiogram Results

A shift table of the investigators' assessments of ECG results at each visit compared with baseline values will be presented using the following categories: normal; abnormal, not clinically significant; abnormal, clinically significant; missing. The number and percentage of subjects in each cross-classification group of the shift table will be presented. Subjects with a missing value at baseline or postbaseline will not be included in the denominator for percentage calculation. No inferential statistics will be provided.

A by-subject listing for ECG assessment results will be provided by subject ID number and visit in chronological order.

7.6. Other Safety Measures

A data listing will be provided for subjects who become pregnant during the study.

7.7. Changes From Protocol-Specified Safety Analyses

There are no deviations from the protocol-specified safety analyses.

8. REFERENCES

Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika 1934;26 (4):404-13.

9. SOFTWARE

SAS® Software Version 9.4. SAS Institute Inc., Cary, NC, USA.

10. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision

11. APPENDICES

Appendix 1. Schedule of Assessments

	Screening	Day 1[a]	Treatment Week (±3 days)					Posttreatment Week (±5 days)		
			1	2	4	8	12/ET[b]	4	12	24
Clinical Assessments										
Informed Consent	X									
Determine Eligibility	X	X								
Medical History	X									
Physical Examination	X	X					X			
Height	X									
Weight	X	X					X		X	X
Vital Signs[c]	X	X	X	X	X	X	X	X	X	X
12-Lead ECG[d]	X	X	X				X			
Adverse Events and Concomitant Medications[e]	X	X	X	X	X	X	X	X		
Pregnancy Prevention Counseling		X					X	X		
Imaging for HCC[f]	X									
Review of Study Drug Adherence and Drug Accountability[g]			X	X	X	X	X			
Study Drug Dispensing[h]		X			X					
Laboratory Assessments										
Hematology, Chemistry	X	X		X	X	X	X	X		
Coagulation (Prothrombin activation %, PT, aPTT and INR)	X	X					X			
Urinalysis	X									

	Screening	Day 1[a]	Treatment Week (±3 days)					Posttreatment Week (±5 days)		
			1	2	4	8	12/ET[b]	4	12	24
HCV RNA	X	X		X	X	X	X	X	X	X
HCV resistance samples		X		X	X	X	X	X	X	X
HBV DNA Sample[i]				X	X	X	X	X		
Serum or Urine Pregnancy Test[j]	X	X			X	X	X	X		
Serum FSH[k]	X									
HCV Genotyping	X									
IL28B Genotype		X								
HCV Ab, HIV Ab, HBsAg, HBsAb, HBcAb	X									
HbA1c	X									
FibroTest®	X									

a Day 1 assessments must be performed prior to dosing.
b ET = early termination.
c Vital signs include resting blood pressure, pulse, respiratory rate and temperature.
d Subjects will be required to rest in a supine position for ≥ 5 minutes prior to making a recording. The investigator (or qualified designee) should review the ECG traces recorded in real time for clinically significant abnormalities.
e Adverse events and Concomitant Medications will be collected up to 30 days after the last dose of all study drugs.
f Liver imaging (eg, ultrasound or CT scan, at the discretion of the investigator) should be performed to exclude the presence of hepatocellular carcinoma (HCC) within 4 months of Day 1 for subjects with cirrhosis, and within 6 months of Day 1 for subjects without cirrhosis.
g Study drugs will be reconciled at every post- Day 1 visit by the investigator in order to monitor the subject's adherence with the study drugs. Subjects must be instructed to bring back all bottles of study drugs in the original container at every post- Day 1 visit through the end of treatment.
h The IWRS will provide direction on the specifics of each subject's study drug dispensing.
i Reflex testing done only when ALT > 2x Day 1 value in subjects who are HBsAb or HBcAb positive at Screening.
j All females of childbearing potential will have a serum pregnancy test at Screening. Urine pregnancy testing will occur at Day 1 and every 4 weeks through posttreatment Week 4 and at ET if applicable. In the event of a positive urine pregnancy result, subjects will be instructed to return to the clinic as soon as possible for a serum pregnancy test.
k Women of any age with amenorrhea of ≥ 12 months (see Protocol Appendix 4).

Appendix 2. Determining Missing and Virtual Visits due to COVID-19

This appendix describes the site collection of COVID-19 data as it pertains to missed/virtual visits and the data processing algorithm used to determine which visits were missing and which visits were virtual.

Data collection

A COVID-19 supplement to the eCRF Completion Guidelines (CCG) was provided by data management to instruct clinical trial sites with respect to data entry expectations pertaining to scenarios related to the COVID-19 pandemic. If a visit was missed, sites should enter "Visit missed due to COVID-19." If a visit which was to be conducted in-person was conducted virtually, sites should enter "Virtual visit due to COVID-19."

Determination of Missed and Virtual visits

Natural Language Processing (NLP) was used to search the CRF comment fields to identify instances of "COVID-19" (or synonyms, see Table X1) and "Virtual" (or synonyms, see Table X1). The search terms are maintained in a global lookup and can be modified and/or corrected to tune the NLP model. For each comment field the following algorithm was applied:

STEP 1: Eliminate extraneous text from each comment field, e.g. "and", "or", "for", etc. This is done using the list of extraneous terms given in Table X2.

STEP 2: Check each of the remaining comment text strings against the "COVID-19" terms and "Virtual" terms with the Levenshtein distance, using SAS function COMPGED (Computes a generalized edit distance using the Levenshtein operations to compute/summarize the degree of difference between two text strings):

i) If Levenshtein distance < 149 for any of the "COVID-19" terms then COVIDFL = 1, else COVIDFL = 0

ii) If Levenshtein distance < 149 for any of the "Virtual" terms then VIRTFL = 1, else VIRTFL = 0

STEP 3: For any comments with COVIDFL = 1, assign "Missed visit" or "Virtual visit as follows:

i) IF COVIDFL =1 and the visit date is missing then result is 'Missed Visit'

ii) IF COVIDFL =1 and VIRTFL = 1 then result is = 'Virtual Visit'

iii) Otherwise result is missing

Examples of search terms for “COVID-19” and “Virtual” used to identify missed and virtual visits.

Table X1.

Search terms for “COVID-19”	Search terms for “Virtual”
COVID19	VIRTUAL
CORONA	TELEMED
CORONAVIRUS	TELEHEALTH
PANDEMIC	TELEPHONE
OUTBREAK	REMOTE
CRISIS	TELEMEDICINE
LOCKDOWN	TELECONSULTATION
QUARANTINE	TELEPHONICALLY
SHELTER	PHONE
	HOME VISIT
	ZOOM
	SKYPE

Table X2.

Examples of extraneous text terms to eliminate from the comment fields.

a	down	in	she'd	until
about	during	into	she'll	up
above	each	is	she's	very
after	few	it	should	was
again	for	its	so	we
against	from	it's	some	we'd
all	further	itself	such	we'll
am	had	i've	than	were
an	has	let's	that	we're
and	have	me	that's	we've
any	having	more	the	what
are	he	most	their	what's
as	he'd	my	theirs	when
at	he'll	myself	them	when's
be	her	nor	themselves	where
because	here	of	then	where's
been	here's	on	there	which
before	hers	once	there's	while
being	herself	only	these	who
below	he's	or	they	whom
between	him	other	they'd	who's
both	himself	ought	they'll	why
but	his	our	they're	why's
by	how	ours	they've	with
could	how's	ourselves	this	would
did	i	out	those	you
do	i'd	over	through	you'd
does	if	own	to	you'll
doing	i'll	same	too	your
down	i'm	she	under	you're
	you've	yourself	yourselves	yours

GS-US-342-5531-SAP-V1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Biostatistics eSigned	13-Apr-2021 23:15:05
PPD	Clinical Research eSigned	14-Apr-2021 05:55:17
PPD	Regulatory Affairs eSigned	14-Apr-2021 15:49:49