CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

203093Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA203093
Supplement #:	0 (original)
Drug Name:	Elvitegravir (VITEKTA®)
Indication(s):	Treatment for HIV
Applicant:	Gilead
Date(s):	Submission date: 6/27/2012
Review Priority:	Standard
Biometrics Division:	DB IV/OB/OTS/CDER
Statistical Reviewer:	Lei Nie
Concurring Reviewers:	Guoxing Soon
Medical Division:	DAVP/OAP/OND/CDER
Clinical Team:	Fleischer, Russell D; Linda Lewis
Project Manager:	Hong, Myung-Joo P.

Keywords: active control/non-inferiority; clinical studies; confounding; double-blind; interaction; logistic regression; NDA review; odds ratio; one trial application; subgroup analysis; interaction; stratification; treatment by baseline interaction; one single trial;

Table of Contents

1 EXECUTIVE SUMMARY

Elvitegravir (EVG) is the second drug in a relatively new class of HIV/AIDS Antiretroviral Drugs: human immunodeficiency virus type 1 integrase strand transfer inhibitor (HIV-1 INSTIs). Gilead Sciences INC. concludes that EVG is noninferior to Raltegravir (RAL; Isentress®, approved by FDA in 2007), the first drug in the same class of HIV/AIDS Antiretroviral Drugs. The sponsor now submits a new application for FDA's approval of ritonavir-boosted EVG for the treatment of HIV-1 infection in antiretroviral treatment-experienced adults, administered together with other antiretroviral agents.

The reviewer concurs with the sponsor's above conclusion in general. While the overall evidence based on the analysis of the pre-specified and commonly used primary endpoint is clear, the concurrence came after the reviewer carefully assessed the following statistical issues: 1) the evidence to support EVG essentially comes from one single trial, while typically two trials are required to support a new drug; 2) two newly approved antiretroviral drugs are used in the trials for EVG as a part of some subjects' background therapy while they were not used in historical trials for RAL. This fact raises concerns about whether the predefined noninferiority margin is valid in subjects who had these new drugs in their background therapy; 3) treatment heterogeneity is observed in gender and in race without a clear explanation; 4) the trial for EVG reveals a remarkably higher rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy higher than in reference trials for RAL.

For issue 1), the reviewer concludes that the evidence from this single trial is as strong as two trials could contribute. For issue 2), by using a newly developed method, hybrid design, the reviewer is able to conclude that the evidence is sufficient to support the noninferiority of EVG relative to RAL. We requested the sponsor to address issues 3) and 4). The sponsor argued that the treatment heterogeneity observed in gender and race could be due to multiple testing and that the other important endpoint did not reveal the heterogeneity in gender or in race. We consider the arguments valid. Based on these arguments and two additional analyses that the reviewer conducted, the reviewer concludes that the potential treatment heterogeneity dose not affect the approvability as the data support EVG's superiority to placebo in men and in women, in white or nonwhite subjects. For issue 4) Gilead believes that the high discontinuation rate may be associated with change in the medical landscape in antiretroviral treatment as a few novel and potent antiretroviral agents became available since the approval of RAL. They believe that these new agents made possible for clinicians to construct effective regimens that could reestablish virologic suppression (HIV-1 RNA < 50 copies/mL) in many patients, even the most treatment-experienced patients who are failing current therapy. The reviewer did not find any data that contradict to the sponsor's explanation. See section 3.2.4 for details.

Nevertheless, we note that, among male subjects, 63% (55%) and 57% (54%) had HIV-1 RNA <50 copies/mL at week 48 (week 96) in the EVG and RAL treatment arms,

respectively. Among female subjects, 47% (39%) and 63% (52%) had HIV-1 RNA <50 copies/mL at week 48 (week 96) in the EVG and RAL treatment arms, respectively. It could be a public interest to assess whether the observed poor performance of EVG relative to RAL, -15% at week 48 and -13% at week 96 in women, is a real signal in future trials. In addition, it may not be appropriate to include (b) (4) , as the sponsor proposed, in the label. Please see Section 5.4 for details.

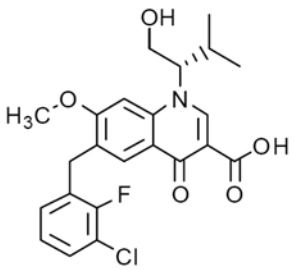
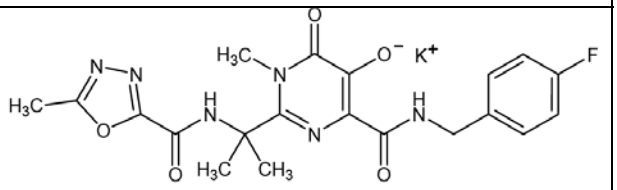
2 INTRODUCTION

2.1 Overview

More than 20 different antiretroviral drugs in 5 classes (nucleoside reverse transcriptase inhibitors [NRTIs], nonnucleoside reverse transcriptase inhibitors [NNRTIs], protease inhibitors [PIs], fusion inhibitors, and integrase strand transfer inhibitors [INSTIs]) are available for the treatment of HIV infection. However, because of poor tolerability and toxicity, because of the development of resistance to the existing treatments, treatment-experienced subjects who suffer side effects from or develop drug resistance to these existing drugs continue to have limited treatment options. Developing safe and effective therapies for treatment-experienced subjects to expand the range of treatment options remains a priority.

Elvitegravir (EVG) is an HIV-1 INSTI that inhibits the HIV-1 integrase, an HIV-1 encoded enzyme that is required for viral replication. Inhibition of integrase prevents the integration of HIV-1 DNA into host genomic DNA, blocking the formation of the HIV-1 provirus. The provirus is required for production of progeny virus, so inhibiting integration prevents propagation of the viral infection. EVG has an empirical formula of $C_{23}H_{23}ClFNO_5$ with a molecular weight of 447.9 Da. The first drug in this class, INSTI, Raltegravir (RAL; Isentress®), was approved In the United States of America (USA) Isentress on 2007. RAL has an empirical formula of $C_{20}H_{20}FKN_6O_5$ with a molecular weight of 482.5 Da. See Table 1.

Table 1: EVG vs RAL

	EVG	RAL
empirical formula	$C_{23}H_{23}ClFNO_5$	$C_{20}H_{20}FKN_6O_5$
molecular weight (Da)	447.9	482.5
Structure formula		

Reference: Label of two drugs.

Raltegravir (RAL; Isentress®) is indicated for the treatment of HIV-1 infection in adult patients in combination with other antiretroviral agents. Through proving that EVG is noninferior to RAL in a confirmatory phase 3, randomized, double blinded clinical trial, the sponsor proposes indication for the EVG tablet as a part of an antiretroviral (ARV) regimen that includes a ritonavir-boosted protease inhibitor and other antiretroviral agents

for the treatment of HIV-1 infection in antiretroviral treatment-experienced adults. EVG is also a component in the 4-drug fixed-dose combination tablet (the QUAD single-tablet regimen [STR]) which is comprised of EVG, a pharmacokinetic enhancer cobicistat (COBI), and the current standard-of-care dual NRTI/NtRTI backbone FTC/TDF (Truvada® [TVD]). The NDA for the QUAD STR (NDA 203100) was submitted to the Food and Drug Administration (FDA) on 27 October 2011, with an indication as a complete regimen for the treatment of HIV-1 infection in adults who are antiretroviral treatment-naïve and the application was approved August 2012.

The Phase 1/2 pharmacokinetic/pharmacodynamic (PK/PD) Trial GS-US-183-0101 suggested that EVG inhibits viral replication in HIV-1 infected subjects and that EVG exhibited exposures supporting once daily dosing when boosted by ritonavir (RTV). Upon this foundation, a clinical program was developed based on boosted EVG as a novel, once-daily INSTI. The dose of EVG (150 mg) was selected based on results from GS-US-183-0101, as well as a Phase 2 trial in heavily treatment-experienced HIV-1 infected subjects (GS-US-183-0105), and a Phase 1 biopharmaceutics/formulation trial (GS-US-183-0140). A series of drug-drug interaction studies containing EVG with RTV-boosted PIs informed whether dose adjustment of EVG was required to achieve target exposure levels. Results of drug interaction studies between ATV/r and EVG (GS-US-183-0108), and RTV-boosted lopinavir (LPV/r) and EVG (GS-US-183-0116), indicated higher systemic EVG exposures upon coadministration with these PIs than with EVG/r alone. Through pharmacokinetic modeling and bioequivalence simulations, an EVG dose of 85 mg was expected to provide similar exposures (AUC) and maintenance of high trough concentrations when administered with ATV/r or LPV/r, relative to EVG/r 150/100 mg; this was demonstrated in a subsequent trial (GS-US-183-0106) using ATV/r, where EVG 85 mg plus ATV/r provided bioequivalent AUC and Cmax as EVG/r 150/100 mg. The reduced dose of 85 mg EVG co-administered with ATV/r or LPV/r was further confirmed in additional studies (GS-US-183-0145 and GS-US-183-0152) to achieve comparable plasma levels of EVG to that intended from the therapeutic dose.

The purpose of the Phase 3 trial 145 was to compare the safety, tolerability, and efficacy of a regimen containing once-daily EVG or twice-daily RAL added to a background regimen (BR) in HIV-1 infected, antiretroviral treatment-experienced adults who had documented resistance, as defined by current International Antiretroviral Society- (IAS) USA Guidelines at the time of the trial, or at least 6 months experience prior to screening with 2 or more different classes of antiretroviral agents. This inclusion criterion was selected to enable the evaluation of EVG in the treatment of a wide range of treatment-experienced subjects failing any line of therapy. The BR, constructed by the investigator based on viral resistance testing, was to be composed of a fully active RTV-boosted PI (PI/r) plus a second agent. A third agent could be used if, and only if, the M184V/I reverse transcriptase (RT) mutation was present on the screening genotype report and a NRTI was used as the second agent, then either emtricitabine (Emtriva®, FTC) or lamivudine (3TC) may have been added as a third agent in the BR. The fixed-dose combination therapies Combivir®, Truvada®, or Epzicom®/Kivexa® may have been prescribed as the combined second and third agents of the BR only in the presence of the M184V/I RT mutation at screening.

This trial 145 is a unification of two trials GS-US-183-144 and GS-US-183-0145, which are identically designed. The only difference between the two trials is: trial 144 is conducted in the United States and Puerto Rico and trial 145 is conducted in Europe, Australia, Canada, and Mexico. Given the declining numbers of patients with unsuppressed viremia, 7 months after the first screening subject visit for GS-US-183-0145, Gilead unified GS-US-183-0144. In this report, trial 145 refers to the unified trial unless otherwise noted.

The primary objective of the trial 145 is to assess noninferiority of a regimen containing ritonavir-boosted elvitegravir versus raltegravir, each administered with a background regimen in HIV-1 infected, antiretroviral treatment-experienced adult subjects as determined by the proportion of subjects achieving and maintaining confirmed HIV-1 RNA < 50 copies/mL through Week 48.

The sponsor submitted safety and efficacy data from the trial 145 as the principal data of EVG for the treatment of HIV-1 infection in antiretroviral treatment-experienced adults, administered together with other antiretroviral agents. They also submitted supportive safety and efficacy data from Phase 2 Trial GS-US-183-0105 (complete) and the roll-over trial GS-US-183-0130 (ongoing and including data through Week 192) are also included. Trial GS-US-183-0130 is an uncontrolled, open label, single arm trial.

Table 2: List of all studies included in analysis

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Trial Population
Applicant defined study number	*Phase 2/3*				*e.g., critical disease or patient characteristics*
145	3	Blinded 96 weeks (or longer)	Open label 144 weeks or longer	351 per arm	HIV infected treatment experienced adults
105	2	48 weeks	0	73-75 per arm	HIV-1 infected, treatment experienced subjects
130	2	Week 48	Ongoing	192	HIV-1 infected, treatment experienced subjects

During the planning stage of the principal trial 145, the review team expressed concerns about the fact that newly approved potent drugs such as Etravirine or Maraviroc were used in trial 145 as a part of background drug in many subjects, while they were not used in the registrational trials, Benchmrk trials, of RAL. We recommended that the sponsor

considers an adaptive trial design using a two-arm trial with a four-arm two-week, lead-in phase, (see statistical review for the document with SN#182. Comments were sent to the sponsor on 6/26/2008). However the sponsor decided not to use the adaptive design. The concern is that "the study results may be uninterpretable due to the highly potent optimized background regimen and the non-inferiority margin cannot easily be estimated and may be close to zero". Given the complexity, we acknowledged that "highly potent optimized background regimen, the non-inferiority margin of 10% will be a review issue" (see statistical review for the document with SN#326. Comments were sent to the sponsor on 2/29/2011). This is one of the major issues that were assessed in the review. In the same document (i.e. comments were sent to the sponsor on 6/26/2008), we also let the sponsor know that "use the FDA snapshot algorithm for the primary analysis of efficacy rather than the TLOVR algorithm…".

We have also encountered several other issues and chief of them are

1. For regulatory approval of a new drug, the United States Code of Federal Regulations (CFR) requires "substantial evidence" from "adequate and well-controlled investigations." This requirement is interpreted in the FDA guidance as the need of "at least two adequate and well-controlled studies, each convincing on its own to establish effectiveness." The guidance also emphasizes the need of "independent substantiation of experimental results from multiple studies." In this application, there is essentially one single trial to support EVG in the targeted population.
2. Heterogeneity is observed in gender and race without clear explanation.
3. It appears that the rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy in 0145 is remarkably higher than the same rate in Benchmark trials.

Through the review, we assessed all these major issues and concluded that none of them should affect the approvability of EVG. However, we feel the numerically poor performance of EVG in women should be noted and it would be interesting to assess whether the poor performance is a real signal or not.

2.2 Data Sources

Study report:

\\Cdsesub1\evsprod\NDA203093\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\hiv\5351-stud-rep-contr\gs-us-183-0145

\\Cdsesub1\evsprod\NDA203093\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\hiv\5351-stud-rep-contr\gs-us-183-0105

Data sets analyzed

\\Cdsesub1\evsprod\NDA203093\0000\m5\datasets\gs-us-183-0145\analysis\adam\datasets\48-wk\

\\Cdsesub1\evsprod\NDA203093\0000\m5\datasets\gs-us-183-0145\analysis\adam\datasets\96-wk\

\\Cdsesub1\evsprod\NDA203093\0000\m5\datasets\gs-us-183-0105\analysis\legacy\datasets\

All data were provided electronically using data formats SDTM and software codes were submitted.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The reviewer was able to reproduce the primary analysis dataset, and in particular the primary endpoint, from the original data source. The final statistical analysis plan (SAP) was submitted and relevant analysis decisions were made prior to unblinding.

The sponsor has submitted data for planned randomization and actual randomization for trial 145, the main trial. The original submission only has the randomization date but has no time. The sponsor lately submitted the randomization implementation with date and actual time. With the submitted code, the reviewer was able to verify the randomized treatment assignments and however detected some problems with the randomization during our review. Specifically, we found discrepancies in two main parts

- part 1) between the IVRS randomization file (RAND0145.xpt) the sponsor sent via FedEx, on June 27, 2012) and the subject level analysis dataset file (~\m5\datasets\gs-us-183-0145\analysis\adam\datasets\96-wk\ADSL.xpt) in the levels of stratifications factors;
- part 2) between the actual treatment received and the treatment original randomization code would assign.

We communicated with the sponsor and sent them our programming code that the reviewed used for the discrepancies, seeking an explanation for these findings.

For part 1), the sponsor confirmed that the discrepancy is due to:

- 26 subjects switched investigational sites which led to a change in the investigator ID during the study;

- 30 subjects were incorrectly stratified during randomization due to the data entry error that occurred when sites randomized subjects to the trial and did not report the errors. For example, Subject 4390-3128 had an HIV-1 RNA value of 326,000 copies/mL at screening as reflected in ADSL.xpt, but the site entered HIV-1 RNA level as <100,000 copies/mL when performing randomization in RAND0145.xpt. Therefore the randomization is incorrect for these subjects;

- Some difference between screening background drugs and baseline background drugs.

For part 2), there are multiple sources for the discrepancies:

- the error reported in part 1);

- 5 reported randomization errors causing the inconsistency.

Despite these errors, the sponsor claimed "The stratification errors either not reported by study sites or reported by study sites as identified have no impact on statistical analysis for efficacy because Gilead used baseline HIV-1 RNA data and the second agent (NRTI or other) at baseline to re-classify subjects' strata in efficacy analysis".

On one hand, the reviewer agrees that the analysis is not impacted as we use the correct stratification factor values in the analysis. However, this does not mean the trial is free of problems. In fact, the randomization error itself is a problem. On the other hand, there is no evidence, from reading into the sponsor's reported errors, that the randomization errors have some systematic patterns to bias the study results.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

Trial 145 is a multicenter, randomized, active-controlled trial with an ongoing 96-week double-blind phase that will be followed by an optional 144-week open-label extension phase. This trial is assessing the safety and efficacy of EVG and RAL, each administered with a background regimen containing a fully active RTV-boosted PI and a second agent in HIV-1 infected, antiretroviral treatment-experienced adults.

Eligible subjects had plasma HIV-1 RNA levels ≥ 1000 copies/mL and documented resistance (as defined by current IAS-USA Guidelines) or at least 6 months experience prior to screening with 2 or more different classes of antiretroviral agents. Thus, subjects may have had resistance to one class and at least 6 months experience prior to screening with a second class of antiretroviral agents, or resistance to 2 classes of antiretroviral agents, or at least 6 months experience with the 2 classes of antiretroviral agents. Subjects

may also have resistance or at least 6 months experience prior to screening with 3 or more classes of antiretroviral agents. Prior to randomization, the components of the BR were selected by the investigator based on each subject's antiretroviral drug history and results of the screening viral resistance profile. Eligible subjects were randomized in a 1:1 ratio to one of the 2 treatment groups. Randomization was stratified by screening HIV-1 RNA level (≤ 100,000 copies/mL or > 100,000 copies/mL) and the class of the second agent (NRTI vs. other classes). After Week 96, subjects will continue to take their blinded study drug and attend visits every 8 weeks until Week 144, and every 12 weeks thereafter until treatment assignments have been unblinded, at which point subjects will be given the option to participate in an open-label extension phase of the study.

Prior to the Baseline/Day 1 visit and before randomization, the investigator chose and documented the subject's BR based on each subject's antiretroviral drug history and results of the viral resistance profile. Subjects determined to be eligible for enrollment into the trial were randomized into one of 2 treatment groups below, to receive oral EVG once daily or oral RAL twice daily plus placebo tablets once or twice daily, as appropriate to maintain the blind. The EVG dose (85 or 150 mg) received by subjects randomized to Treatment Group 1 was based on the PI/r in the BR. All subjects received 3 tablets of study drug per day (active drug and placebo), in addition to the components of the BR. Study drug was administered in a blinded fashion, continuing to be administered in a blinded fashion post-Week 96.

Subjects who were receiving darunavir/r (DRV/r), fosamprenavir/r (FPV/r), or tipranavir/r (TPV/r) as part of their BR received EVG 150 mg if randomized to Treatment Group 1. Due to known PK interactions, subjects who were taking ATV/r or lopinavir (LPV)/r as part of their BR received EVG 85 mg if randomized to Treatment Group 1. See Table **3**.

Table 3: Treatment administration

Randomization	**Ritonavir-Boosted PI**	
	ATV/r or LPV/r	**DRV/r, FPV/r, TPV/r**
Treatment Group 1 EVG/r + BR	EVG 85 mg (pentagon) once daily RAL 400 mg placebo (oval) twice daily	EVG 150 mg (triangle) once daily RAL 400 mg placebo (oval) twice daily
Treatment Group 2 RAL + BR	EVG 85 mg placebo (pentagon) once daily RAL 400 mg (oval) twice daily	EVG 150 mg placebo (triangle) once daily RAL 400 mg (oval) twice daily

ATV, atazanavir; DRV, darunavir; FPV, fosamprenavir; LPV, lopinavir; /r, boosted with ritonavir; TPV, tipranavir

Sources: study report 0145 (week 96), Table 7-1.

With Amendment 2, the 2 identical Phase 3 EVG studies, GS-US-183-0144 and GS-US-183-0145, were unified. Initiated in July 2008, Trial GS-US-183-0144 (planned N = 700) was being conducted in the US and Puerto Rico and Trial GS-US-183-0145 (planned N = 700) was being conducted in Europe, Australia, Mexico, and Canada.

Given the declining numbers of patients with unsuppressed viremia, Gilead combined the 2 ongoing studies into a single, global Phase 3 trial (GS-US-183-0145) to enroll a total of 700 subjects. Amendment 2, described below, enabled the transfer of subjects from Trial GS-US-183-0144 into Trial GS-US-183-0145.

This amended Trial GS-US-183-0145 contained 2 randomization schemes based on geographic areas: (1) US and Puerto Rico from original Trial GS-US-183-0144 and (2) Europe, Australia, Canada, and Mexico from original Trial GS-US-183-0145. These 2 randomization schemes were equivalent to randomization strata within a trial, ie, the 2 geographic areas corresponded to 2 strata. Both Phase 3 studies were identical in design with the same 1:1 randomization ratio and there were independent unique patient identification numbers across both studies. Incorporating the randomization scheme for subjects in the US and Puerto Rico of Trial GS-US-183-0144 into Trial GS-US-183-0145 maintained the validity of randomization in the amended Trial GS-US-183-0145.

In Amendment 2, dated 18 February 2009, the protocol was updated to reflect the unification of Studies GS-US-183-0144 and GS-US-183-0145.

The external IDMC examined the safety results of the trial and also focused on logistical issues such as accrual, retention, and quality of clinical and laboratory data. Close attention was given on virologic rebound rate throughout the independent monitoring process. Blinding was preserved during the conduct of the trial and access to unblinded data was limited to designated parties. The IDMC has reviewed the progress, efficacy, and safety profile of this trial throughout trial conduct. No formal stopping rules were used by the IDMC for safety outcomes. Rather, a clinical assessment was made to determine if the nature, frequency, and severity of AEs associated with the study regimen warranted early termination of the trial in the best interest of the participants. An analysis of data for the Week 12 IDMC meeting was conducted after the first 350 subjects either completed their Week 12 visit or prematurely discontinued study drug prior to the Week 12 visit. After reviewing the data, the IDMC did not have any concerns and approved the trial to continue. Week 48 data were added after the double-blind period was extended to 96 weeks, and after a review of Week 48 data, the IDMC again approved that the trial continue unchanged.

The primary efficacy endpoint was the percentage of subjects who achieved and maintained confirmed HIV-1 RNA < 50 copies/mL through Week 48. This outcome variable was derived based on the FDA-defined TLOVR algorithm.

Trial 105 was a randomized, partially blinded, multicenter trial to assess the noninferiority of ritonavir boosted EVG relative to CPI/r (one or two marketed protease inhibitors as deemed appropriate by the investigator, dosed in combination with ritonavir unless the chosen CPI was co-formulated with ritonavir), both with a background ARV regimen. Before randomization, the components of the CPI and Antiretroviral (ARV) regimens were selected by the investigator, without input from the sponsor, and were based on each subject's ARV drug history and results of a viral resistance profile.

Subjects with plasma HIV-1 RNA levels ≥ 1,000 copies/mL with documented presence of at least one of the protease gene mutations were eligible to enter the trial. In the original protocol, the ARV portion of the regimen consisted of at least two marketed agents, not including a protease inhibitor and an NNRTI. Subjects were allowed to use T-20 (enfuvirtide or Fuzeon™, Hoffman-LaRoche & Trimeris) as part of their ARV regimen. Randomization was stratified by the use of T-20. Subjects were initially randomized (1:1:1:1) to receive CPI/r or one of the EVG/r doses: 20/100, 50/100, or 125/100 mg.

The dose of EVG was blinded, but CPI and ARV components of the regimen were given open-label. Therefore there is open label comparison between EVG and the control group.

The primary endpoint is time-weighted average changes from baseline in HIV-1 RNA (DAVG) at week 24.

Trial 130 is a Phase 2, rollover, open-label, multicenter, multiple-dose, single-arm extension trial designed to assess the safety of EVG/r, in combination with other antiretroviral agents, in treatment-experienced HIV-1 infected adults and adolescents. Subjects were eligible for this trial if they had completed a prior EVG/r treatment study without experiencing any dose-limiting toxicity; eligible subjects may or may not have received EVG in their prior study. Subjects were enrolled in this extension study regardless of their baseline HIV-1 RNA level (i.e., subjects with baseline HIV-1 RNA levels of either < 50 or ≥ 50 copies/mL were enrolled). Non virologically suppressed subjects entering this trial had, for the most part, failed prior antiretroviral regimens and had limited treatment options available; these subjects were allowed in the current extension trial even if they had been exposed to EVG in their prior study. Genotyping was not performed at baseline, so subjects who met eligibility requirements were enrolled at the discretion of the investigator. Approximately 1000 to 1500 subjects were planned to be rolled-over from multiple prior studies; however, subjects from only 2 prior studies (GS-US-183-0105 and GS-US-183-0152) were rolled-over into this extension trial. Trial GS-US-183-0130 is ongoing at the time of this Week 192 interim clinical study report. In this submission a total of 192 were enrolled and 184 were from trial 130.

The primary objective of this trial was to observe the long-term safety of EVG/r in combination with other antiretroviral agents in subjects who have completed a prior EVG/r treatment study
The primary endpoint is the percentages of subjects with HIV-1 RNA < 50 copies/mL.

Issues and Reviewer's Comments

1. In this application, there is essentially one single trial to support EVG in the targeted population.

In the application, the proposed registration dose of EVG is as follows

Dose of EVG	Dose of Coadministered Ritonavir-Boosted Protease Inhibitor
85 mg once daily	atazanavir/ritonavir 300/100 mg once daily
	lopinavir/ritonavir 400/100 mg twice daily
150 mg once daily	darunavir/ritonavir 600/100 mg twice daily
	fosamprenavir/ritonavir 700/100 mg twice daily
	tipranavir/ritonavir 500/200 mg twice daily

Trial 145 is designed to support this dose regimen. However, trial 105 is not. In trial 105, only the arm of EVG 125 mg, which enrolled about 75 subjects, is within the range of targeted dose regimen. In addition, the control group in trial 105 did not include RAL.

Trial 130 is a single arm, open label uncontrolled trial.

Because the doses used in trial 105 are not really the same as the proposed dose regimen and the only relevant arm enrolled a small sample size in trial 105, and because trial 130 is uncontrolled single arm trial, the reviewer does not think that trial 130 or trial105 plays a major role, leaving basically the single trial (trial 145) to support the new drug.

2. At some time points, HIV viral load was measured by only one of following three HIV RNA assays: 1) Amplicor/standard; 2) Amplicor/ultra sensitive; and 3) Taqman. It was not clear how the analysis deals with the HIV viral load data which were measured by two (or actually three) assays.

We requested clarifications from the sponsor. They explained, for measurement of HIV-1 RNA levels, the results obtained by Amplicor/ultrasensitive assay were used first. If the result from Amplicor/ultrasensitive assay was "< 50 copies/mL", a numeric value of 49 copies/mL was used for analysis. If values were not present or above the upper limit of the assay (i.e., >100,000 copies/mL), values measured by Amplicor/standard assay were used. If values from the Amplicor/standard assay were not present or above the upper limit of the assay (i.e., > 750,000 copies/mL), values from a reflex testing based on the diluted samples were used. If HIV-1 RNA plasma levels were reported as > 100,000 copies/mL (Amplicor/ultrasensitive) or > 750,000 copies/mL (Amplicor/standard) and no reflex results were available, the numeric value of 100,001 or 750,001, respectively, was used.

The reviewer found, except rare occasions, the recorded HIV viral loads at week 48 and week 96 were measured by Amplicor/ultrasensitive assay, which are appropriate.

3. The combined randomization scheme of Studies GS-US-183-0144 and amended GS-US-183- 0145 was completed and in production in the IVR/IWR System on 24 April 2009. Subjects who were randomized in the original Study GS-US-183-0144 have subject ID numbers in the 3000's, country code as US (United States) or PR (Puerto Rico), and the randomization date prior to 24 April 2009. The interim analysis happened after the unification of two trials. The unification happened April 24, 2009. By Aug 20, 2009, 350 Subjects had week 12 measurement. Thus, the unification of two trials is unrelated to any interim analysis results.

4. Each of two trials 144 (international not including USA) and 145 (USA) was stratified by HIV RNA levels and class of the second agent in the BR, the unified trial, new 145, should be considered as stratified by HIV RNA levels, class of the second agent in the BR, and region (USA including Puerto Rico vs. non USA).

3.2.2 Statistical Methodologies

From now on, we only focus on trial 145, unless otherwise noted.

The sponsor's methodologies

The primary efficacy endpoint was the percentage of subjects who achieved and maintained confirmed HIV-1 RNA < 50 copies/mL through Week 48. This outcome variable was derived based on the FDA-defined TLOVR algorithm. Subjects were considered responders at Week 48 (i.e., achieved and maintained confirmed HIV-1 RNA < 50 copies/mL) based on the following criteria:

- Had achieved a confirmed suppression (i.e., HIV-1 RNA < 50 copies/mL at 2 consecutive visits) on or prior to the upper limit of the Week 48 visit window
- Had not experienced death or permanent discontinuation of study drug on or prior to the upper limit of the Week 48 visit window
- Had not had a confirmed rebound (i.e., HIV-1 RNA ≥50 copies/mL at 2 consecutive visits or the last available HIV-1 RNA ≥ 50 copies/mL followed by premature discontinuation of study drug) after achieving confirmed suppression on or prior to the upper limit of the Week 48 visit window
- Had not switched BRs for reasons other than toxicity management before achieving a confirmed suppression.

The primary analysis is an ITT analysis performed on all subjects who were randomized into the study, received at least 1 dose of study medication.

The analysis purpose of the primary efficacy endpoint was to assess noninferiority of EVG treatment relative to RAL treatment (in addition to the BR). Noninferiority was assessed using a conventional 95% CI approach, with a delta of 0.10. If the lower bound of the 2-sided 95% CI of the difference (EVG treatment group − RAL treatment group) in the response rate was > −10%, then it was to be concluded that EVG is noninferior to

RAL. The baseline strata (HIV-1 RNA level ≤ 100,000 copies/mL or >100,000 copies/mL) and the class of the second agent (NRTI vs. other classes) weighted difference in 2 proportions (P1 − P2) and its 95% CI were calculated based on stratum-adjusted Mantel-Haenszel proportions.

Randomization was stratified by the following 3 factors:

- Geographic areas (US and Puerto Rico vs. Others, including Australia, Canada, Europe, and Mexico)
- Screening HIV-1 RNA level (≤ 100,000 copies/mL vs. > 100,000 copies/mL)
- Class of the second agent (NRTI vs. other classes) in the BR

All efficacy analyses were stratified by baseline HIV-1 RNA level (≤ 100,000 copies/mL vs. > 100,000 copies/mL) and class of the second agent (NRTI vs. other classes). Efficacy analyses were not stratified by geographic areas because virologic response was expected to be similar across different geographic areas.

If noninferiority between EVG and RAL treatment regimens was established, the same 95% CI that was used for assessing noninferiority was to be used to evaluate superiority. If the lower bound of the 95% CI was > 0, then superiority was established. Superiority between treatment groups was also assessed using a 2-sided Cochran-Mantel-Haenszel test adjusted for baseline HIV-1 RNA level (≤ 100,000 copies/mL vs. > 100,000 copies/mL) and the class of the second agent (NRTI vs. other classes). There was only one primary comparison for the primary efficacy endpoint; therefore, no adjustment for alpha level was required.

A missing datum for a given study visit window may have been due to any of the following:

• A visit occurred in the window, but data were not collected or were unusable.
• A visit did not occur in the window.
• A subject permanently discontinued from the study before reaching the visit window.

Values for missing data were not be imputed.

Study drug adherence was computed based on pill counts taken up to the data cutoff date. The number of pills of study drug dispensed and returned was captured on the Study Drug
Accountability eCRF. Adherence (%) of study drug was calculated as follows:

100 × total number of pills taken divided by total number of pills prescribed
= 100 × sum of number of pills taken at each dispensing period divided by sum of number of pills prescribed at each dispensing period

As the secondary endpoint, the sponsor also proposed FDA-defined snapshot analysis for Week 48 virologic outcome (i.e., the percentage of subjects with HIV-1 RNA < 50

copies/mL), the visit window was defined as Day 309 to Day 364, inclusive. The same snapshot analysis is also performed for week 96 virologic outcome. All HIV-1 RNA data collected while subjects remained on randomized treatment were used in the snapshot analysis.

Gilead pre-specified the analysis windows prior to Week 48 database lock and data analysis for Trial GS-US-183-0145. Gilead defined the Week 48 snapshot analysis window from Day 309 (Week 44) to Day 364 (Week 52) and the Week 96 snapshot analysis window from Day 645 (Week 92) to Day 700 (Week 100).

Reviewer's Comments and additional analyses

1. We asses whether the noninferiority (NI) margin is appropriate. As we noted the sponsor before, the non-inferiority margin of 10% is a review issue. Although the NI margin is pre-defined, it usually does not take into consideration of important changes, such as population deviation in the new trials from historical trials. For example, in this trial, trial 145 basically enrolled subjects who have a Baseline Phenotypic Sensitivity Scores (PSS) between 1 or 2, while the reference historical trials, BENCHMRK, enrolled subjects with any PSS values (0,1,2,3,>3).
2. We obtain the primary endpoint using a different algorithm. The primary efficacy endpoint, proportion of subjects responding by the FDA Time to Loss of Virologic Response (TLOVR) algorithm at Week 96, is appropriate. However, FDA reviewers now prefer a simpler endpoint, the proportion of subjects responding at Week 96, referred as Snapshot approach.
3. We obtain the primary endpoint using the standard analysis window. The sponsor's analysis window for snapshot analysis is different than what is used in FDA's snapshot guidance (see Table 4).

Table 4: Analysis Windows for the Week 48 and Week 96 by FDA and the sponsor.

	FDA Snapshot Guidance		**GS-US-183-0145 Study Snapshot Analysis Specifications**	
Visit	Window (through end of study week)	Window (Days)	Window (through end of study week)	Window (Days)
24	18-30	126-209	NA	NA
48	42-54	294-377	44-52	309-364, inclusive
96	90-102	630-713	92-100	645-700 inclusive

4. In this application, there is essentially one single trial to support EVG in the targeted population. We assess whether this is sufficient as two trials were typically required to support a new drug.
5. We performed an alternative stratified analysis. The trial 145 was a combination of the old trials 144 and 145, with 144 conducted in US and 145 outside of US,

therefore the geographic region becomes a natural stratification factor in the combined trial 145. The current stratum-adjusted analysis the sponsor performed only include the original stratification factors: baseline viral load and the second agent in the background drug because they hypothesized "virologic response was expected to be similar across different geographic areas" but this hypothesis needs not to be true. See Section 3.2.4.

6. Assessment of newly approved drugs that were not used in Benchmrk trials. Two "newly" approved drugs were used in trial 145 as a part of BR but they were not sued in Benchmrk trials, the reference trial to establish noninferiority. We examine whether there is a way to evaluate the efficacy of EVG.
7. Heterogeneity of treatment effect of EVG relative to RAL. We found heterogeneity of treatment effect in gender, race, and possibly in region. We investigate whether they are "real" discovery or chance finding. Most importantly, we examine whether they affect the approvability of the drug or support some restrictions of the drug in some subgroups.
8. It appears that the rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy in 0145 is remarkably higher than the same rate in Benchmark trials. We would like to find possible explanations for this phenomenon.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

A total of 1335 subjects were screened. Among all, a total of 603 subjects were screen failures. Eight subjects were not randomized into the trial, although they met the eligibility criteria.

Of the 724 subjects randomized, 361 in the EVG group and 363 in the RAL group, 12 subjects never received study drug (EVG 7 subjects and RAL 5 subjects). Thus, 354 and 358 subjects received at least 1 dose of study drug in the EVG and RAL groups, respectively.
Among the 724 randomized subjects, 40% of subjects in the EVG arm discontinued study drug and 41% of subjects in the RAL arm discontinued study drug. Overall, the most common reasons for discontinuation of study drug or trial were: subject noncompliance (11% in the EVG arm and 9% in the RAL arm); withdrew consent (8% in the EVG arm and 5% in the RAL arm)

Reasons for discontinuation of study drug were generally similar in the 2 groups; however, there were fewer subject deaths in the EVG group compared with the RAL group (EVG 1 subjects, 0.3%; RAL 9 subjects, 2.5%), and there were more discontinuations because of withdrawal of consent in the EVG group compared with the RAL group (EVG 30 subjects, 8.5%; RAL 17 subjects, 4.7%). Adverse events resulted in study drug discontinuation for 11 subjects (3.1%) in the EVG group and for 15 subjects (4.2%) in the RAL group; lack of efficacy (5% in the EVG arm and 6% in the RAL arm). AEs (3% in the EVG arm and 4% in the RAL arm). Reasons for discontinuation appear

to be similar in two groups except the number of the death is much higher in the RAL arm.

Table 5: patient disposition at week 96

Subject Disposition (trial 145)	EVG	RAL
Subjects Randomized	361	363
Subjects Randomized and Never Dosed	7	5
Subjects Discontinued Study Drug	146 (40.4%)	150 (41.3%)
Subject Non-Compliance	38 (10.5%)	34 (9.4%)
Lost to Follow-Up	30 (8.1%)	31 (8.5%)
Withdrew Consent	30 (8.3%)	17 (4.7%)
Lack of Efficacy	17 (4.7%)	21 (5.8%)
Protocol Violation	11 (3.1%)	14 (3.9%)
Adverse Event	11 (3.1%)	15 (4.1%)
Death	1(0.3%)	9 (2.5%)
Investigator's Discretion	5 (1.4%)	8 (2.2%)
Pregnancy	3 (0.8%)	1 (0.3%)

Source: Study report of 145 (week 96) and review's analysis. Two analyses are consistent with minor differences.

Subjects enrolled at Dr. F. Marquez's site (Site 4390; 3 EVG and 7 RAL subjects) were excluded from the ITT and the PP analysis sets (but included in the safety analysis set) due to failure to comply with the signed investigator agreement. Important protocol deviations at this site included the following: (1) subjects were given the study drug without other active antiretroviral drugs; and (2) subjects who met protocol-defined inclusion and exclusion criteria could not be verified by the source documentation. A letter to Dr. F. Marquez dated 27 March 2009 that describes the number of significant findings noted during a site audit is appended to this report. All 10 subjects at Site 4390 prematurely discontinued study drug; 9 subjects because of protocol deviations and 1 subject (in the EVG group) withdrew consent.

Consequently, the ITT population refers to all randomized subjects who received at least one dose of treatment; however exclude the 10 subjects from the site 4390. That is, each of two arm has 351 subjects in the ITT. As we explained before, it is appropriate to exclude subjects enrolled at Dr. F. Marquez's site.

Table 6: Demographic

Characteristic	EVG (N=351)	RAL (N=351)	Total (N=702)
Age (Years)			
N	351	351	702
Mean (SD)	44 (9.0)	45 (9.2)	45 (9.1)
Median	44	45	45
Q1, Q3	38, 50	40, 51	39, 50
Min, Max	20, 78	19, 74	19, 78
Sex			
Male	292 (83.2%)	284 (80.9%)	576 (82.1%)
Female	59 (16.8%)	67 (19.1%)	126 (17.9%)
Race			
White	211 (60.1%)	226 (64.4%)	437 (62.3%)
Black or African American	125 (35.6%)	113 (32.2%)	238 (33.9%)
Asian	9 (2.6%)	5 (1.4%)	14 (2.0%)
American Indian or Alaska Native	2 (0.6%)	3 (0.9%)	5 (0.7%)
Native Hawaiian or Other Pacific Islander	1 (0.3%)	0	1 (0.1%)
Other	3 (0.9%)	4 (1.1%)	7 (1.0%)
Ethnicity			
Hispanic or Latino	79 (22.5%)	73 (20.8%)	152 (21.7%)
Not Hispanic or Latino	271 (77.2%)	277 (78.9%)	548 (78.1%)
Not Reported	1 (0.3%)	1 (0.3%)	2 (0.3%)
Region			
USA*	224 (63.8%)	215 (61.3%)	439 (62.5%)
Non USA*	127(36.2%)	136 (38.7%)	263(37.5%)

Source: Table 8.4: Study 145 week 48 report and reviewer's analysis (USA* does not include Puerto Rico)

Table 7 : Baseline characteristic

Characteristic	**EVG (N=351)**	**RAL (N=351)**	**Total (N=702)**
Baseline HIV-1 RNA (log10 copies/mL)			
N	351	351	702
Mean (SD)	4.26 (0.971)	4.27 (0.944)	4.26 (0.957)
Median	4.35	4.42	4.39
Q1, Q3	3.66, 5.03	3.60, 5.02	3.64, 5.02
Min, Max	1.69, 6.63	1.69, 6.10	1.69, 6.63
Baseline HIV-1 RNA Category			
Baseline HIV-1 RNA level <= 100,000 copies/mL	261 (74.4%)	261 (74.4%)	522 (74.4%)
Baseline HIV-1 RNA level > 100,000 copies/mL	90 (25.6%)	90 (25.6%)	180 (25.6%)
Baseline CD4 (cells/mm3)			
N	340	341	681
Mean (SD)	259.3 (204.44)	264.0 (207.92)	261.7 (206.05)
Median	227.0	215.0	222.0
Q1, Q3	100.0, 371.0	111.0, 381.0	106.0, 379.0
Min, Max	2.0, 1374.0	1.0, 1497.0	1.0, 1497.0
HIV Status			
Asymptomatic	170 (48.4%)	168 (47.9%)	338 (48.1%)
Symptomatic HIV Infections	51 (14.5%)	54 (15.4%)	105 (15.0%)
AIDS	126 (35.9%)	125 (35.6%)	251 (35.8%)
Baseline Genotypic Sensitivity Score Category			
0	4 (1.1%)	1 (0.3%)	5 (0.7%)
1	50 (14.3%)	53 (15.1%)	103 (14.7%)
2	284 (81.1%)	291 (82.9%)	575 (82.0%)
3	12 (3.4%)	6 (1.7%)	18 (2.6%)
Type of PI in Background Regimen			
Atazanavir	61 (17.4%)	51 (14.5%)	112 (15.9%)
Darunavir	202 (57.5%)	207 (58.8%)	409 (58.2%)
Fosamprenavir	14 (4.0%)	19 (5.4%)	33 (4.7%)
Kaletra	68 (19.4%)	68 (19.3%)	136 (19.3%)
Tipranavir	6 (1.7%)	7 (2.0%)	13 (1.8%)
Type of NRTI in Background Regimen			
Abacavir	5 (1.4%)	12 (3.4%)	17 (2.4%)
Combivir	6 (1.7%)	5 (1.4%)	11 (1.5%)
Didanosine	1 (0.3%)	5 (1.4%)	6 (0.8%)
Emtricitabine	2 (0.6%)	2 (0.6%)	4 (0.6%)
Epzicom	4 (1.1%)	8 (2.2%)	12 (1.7%)
Lamivudine	11 (3.1%)	13 (3.6%)	24 (3.4%)
Tenofovir DF	163 (46.0%)	171 (47.8%)	334 (46.9%)
Truvada	91 (25.7%)	67 (18.7%)	158 (22.2%)
Zidovudine	3 (0.8%)	3 (1.7%)	9 (1.3%)

Source: Study report 145, Week 48. Table 8.

Reviewer's comments:

Subjects enrolled at Dr. F. Marquez's site (Site 4390; 3 EVG and 7 RAL subjects) were excluded from the ITT. The reviewer considers this conservative approach against EVG appropriate. The alternative approach that including these subjects shall benefit EVG more than benefit RAL as all 3 EVG subjects and 7 RAL subjects are not virologic success.

3.2.4 Results and Conclusions

The sponsor's results and conclusions

The analysis of the primary efficacy endpoint was evaluated in the Week 48 and was evaluated and in the Week 96. The percentage of subjects achieving and maintaining confirmed HIV-1 RNA< 50 copies/mL at Week 48 (TLOVR analysis, ITT) was similar in the EVG and RAL treatment groups: 59% of subjects (207 of 351 subjects) in the EVG group and 58% of subjects (203 of 351 subjects) in the RAL group were classified as responders. The stratum-adjusted difference between treatment groups (EVG − RAL) was 1.1%, and the 95% CI was −6.0% to 8.2%. The lower bound of the 2-sided 95% CI of the stratum-weighted difference (EVG − RAL) in the response rate was −6%, which is greater than the prespecified noninferiority margin of −10%, and confirmed the robustness of the Week 48 outcome that EVG is noninferior to RAL.
The percentage of subjects achieving and maintaining confirmed HIV-1 RNA < 50 copies/mL at Week 96 (TLOVR analysis, ITT) was similar in the EVG and RAL groups: 47.6% of subjects (167 of 351 subjects) in the EVG group and 45.0% of subjects (158 of 351 subjects) in the RAL group were classified as responders. The stratum-adjusted difference between groups (EVG − RAL) was 2.6%, and the 95% CI was −4.6% to 9.9%.

The percentage of subjects with HIV-1 RNA < 50 copies/mL at Week 48 (snapshot analysis,
ITT) was similar in the EVG and RAL treatment groups: 59.8% of subjects (210 of 351 subjects) in the EVG group and 57.5% of subjects (202 of 351 subjects) in the RAL group were classified as a virologic success. The stratum-adjusted difference between treatment groups (EVG − RAL) was 2.2%, and the 95% CI was −5.0% to 9.3%. The percentage of subjects with HIV-1 RNA < 50 copies/mL at Week 96 (snapshot analysis, ITT) was similar in the EVG and RAL treatment groups: 52.4% of subjects (184 of 351 subjects) in the EVG group and 53.0% of subjects (186 of 351 subjects) in the RAL group were classified as a virologic success. The stratum-adjusted difference between treatment groups (EVG − RAL) was −0.5%, and the 95% CI was −7.9% to 6.8%. The response rates observed in the snapshot analysis were slightly higher than those observed in the TLOVR analysis.

The reviewer was able to confirm these results with minor differences. For example, using their snapshot data at week 96, the reviewer obtained a stratum-adjusted difference

between treatment groups (EVG − RAL) was −0.55%, and the 95% CI was −7.8% to 6.7%. However, the difference could be due to slightly difference of classification of the second agents (while they reported a total of 144 non NRTI, in Table 25.2 (the study report for trial 145), as the second agent in the background therapy, the reviewer found 137) and possibly implementing stratum-adjusted approaches. The minor difference is negligible.

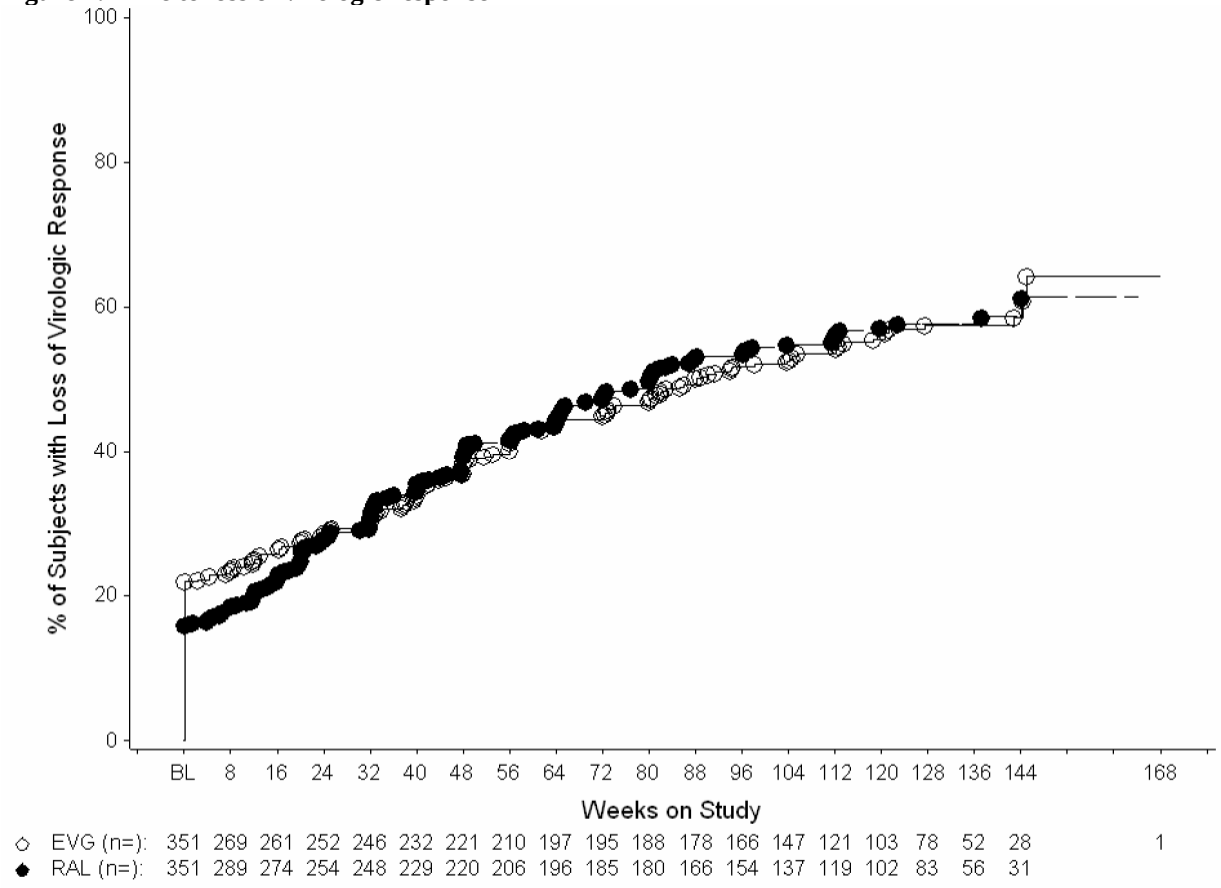
Table 8: Snapshot analysis for trial 145

	Week 48		Week 96	
	EVG (n=351)	RAL (n=351)	EVG (n=351)	RAL (n=351)
HIV-1 RNA < 50 copies/mL	**210 (59.8%)**	**202 (57.5%)**	**184 (52.4%)**	**186 (53.0%)**
Virologic Failure	**115 (32.8%)**	**112 (31.9%)**	**125 (35.6%)**	**109 (31.1%)**
HIV-1 RNA >= 50 copies/mL	56 (16.0%)	66 (18.8%)	40 (11.4%)	31 (8.8%)
Discontinued Study Drug Due to Lack of Efficacy	9 (2.6%)	7 (2.0%)	15 (4.3%)	19 (5.4%)
Discontinued Study Drug Due to Other Reasons and Last Available HIV-1 RNA >= 50 copies/mL	49 (14.0%)	37 (10.5%)	67 (19.1%)	57 (16.2%)
HIV-1 RNA >= 50 copies/mL at Background Regimen Switch and HIV-1 RNA < 50 copies/mL at Week 48	1 (0.3%)	2 (0.6%)	3 (0.9%)	2 (0.6%)
No Virologic Data in Week 48 Window	**26 (7.4%)**	**37 (10.5%)**	**42 (12.0%)**	**56 (16.0%)**
Discontinued Study Drug Due to AE or Death	8 (2.3%)	18 (5.1%)	10 (2.8%)	24 (6.8%)
Discontinued Study Drug Due to Other Reasons and Last Available HIV-1 RNA < 50 copies/mL	14 (4.0%)	17 (4.8%)	28 (8.0%)	30 (8.5%)
Missing Data during Window but on Study Drug	4 (1.1%)	2 (0.6%)	4 (1.1%)	2 (0.6%)

Source: Table 9-3 and Table 9-4, study report of Trial 145.

Time to loss of virologic response with HIV-1 RNA cutoff at 50 copies/mL is depicted graphically in Figure 1. The KM curves separated early, as a higher percentage of subjects in the EVG group compared with the RAL group had loss of virologic response due to never being suppressed and were, therefore, assumed to have failed at Day 1. In contrast, a higher percentage of subjects in the RAL group compared with the EVG group experienced virologic rebound; these subjects failed at the time when the rebound occurred. Taken together, the percentages of subjects with loss of virologic response (due to never being suppressed or rebound) were similar between the EVG and RAL treatment groups. At Week 96, the KM estimates for the percentages of subjects with loss of virologic response were 52% for the EVG group and 55% for the RAL group.

Figure 1: Time to loss of virologic response

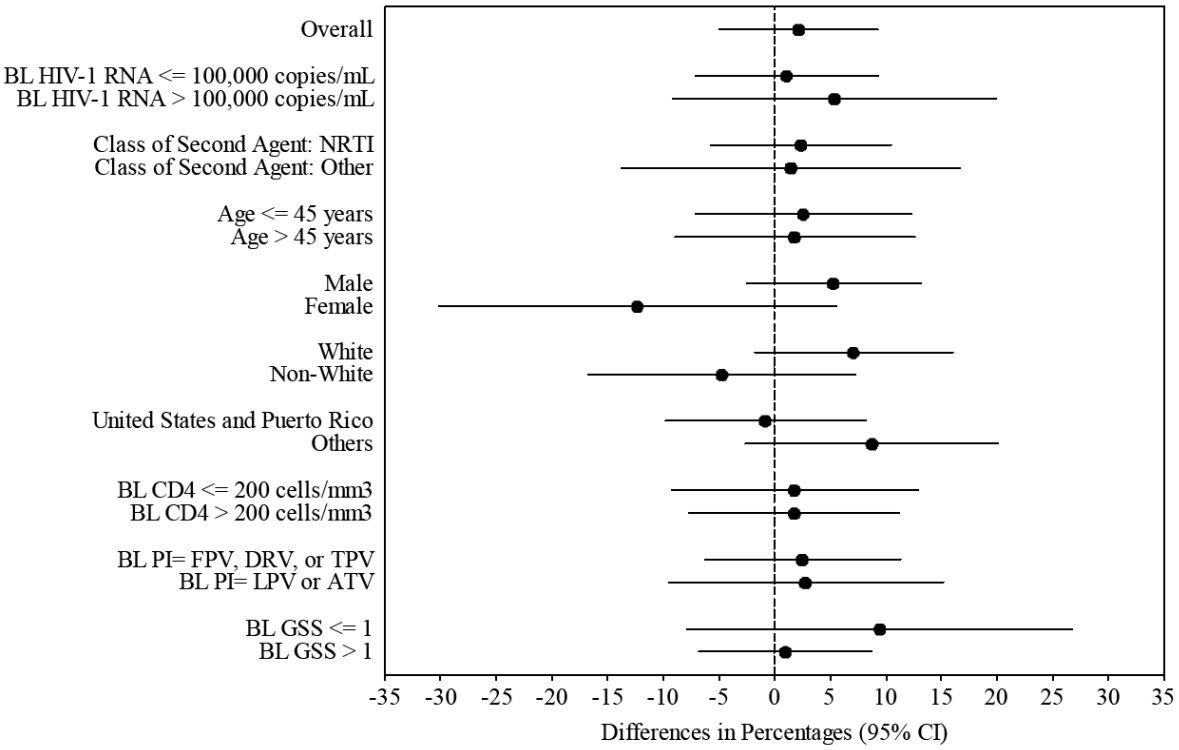


Source: Figure 9.1, study report for week 96.

The sponsor has also conducted a number of analyses to examine the efficacy endpoint using different algorithms and different threshold. They claim that these analyses generally support the noninferiority of EVG relative to RAL.

The sponsor also performed some subgroup analyses. They concluded "The 95% CIs for treatment differences in the percentage of subjects with virologic success for HIV-1 RNA < 50 copies/mL included zero for all subgroups, suggesting no treatment differences across subgroups". In the reviewer's opinion, the conclusion of treatment heterogeneity based on whether the 95% CIs for treatment differences in the percentage of subjects with virologic success for HIV-1 RNA < 50 copies/mL included zero in all subgroups may not be appropriate. Many researchers prefer to assess treatment heterogeneity using testing treatment by subgroup interaction.

Figure 2: sponsor's subgroup analysis of snapshot results.



CD4 cell counts increased following administration of study drug, and the mean increases were similar between the EVG and RAL treatment groups at all time points. Mean (SD) baseline CD4 cell counts were 259 (204.4) cells/mm3 in the EVG group and 264 (207.9) cells/mm3 in the RAL group. At Week 96, the mean (SD) increases from baseline in CD4 cell count were 205 (191.5) cells/mm3 in the EVG group and 198 (162.2) cells/mm3 in the RAL group.

The sponsor concluded

1. The analysis of the primary efficacy endpoint was evaluated in the Week 48 CSR (using the Week 48 dataset) and was evaluated again using the Week 96 dataset. Using the Week 96 dataset, results of the primary analysis of the primary efficacy endpoint were identical to those using the Week 48 dataset. These results confirm the robustness of the Week 48 outcome that EVG once daily is noninferior to RAL twice daily when administered for 48 weeks to HIV-1 infected, antiretroviral treatment-experienced adults in combination with a fully active RTV-boosted PI and an active second agent. Based on a TLOVR analysis, 59.0% of subjects (207 of 351 subjects) in the EVG group and 57.8% of subjects (203 of 351 subjects) in the RAL group achieved and maintained confirmed HIV-1 RNA < 50 copies/mL at Week 48. The stratum-adjusted difference between treatment groups (EVG − RAL) was 1.1%, and the 95% CI was −6.0% to 8.2%. The lower bound of the 2-

sided 95% CI of the stratum-weighted difference (EVG − RAL) in the response rate was −6%, which is greater than the prespecified noninferiority margin of −10%.

2. Using the Week 96 dataset, the percentages of subjects in each treatment group with HIV-1 RNA < 50 copies/mL at Week 48 (snapshot analysis, ITT) were also identical to those using the Week 48 dataset, and confirmed the results of the TLOVR analysis. Elvitegravir and RAL had comparable rates of virologic success. Based on the snapshot analysis, 59.8% of subjects (210 of 351 subjects) in the EVG group and 57.5% of subjects (202 of 351 subjects) in the RAL group had virologic success at Week 48. The stratum-adjusted difference between treatment groups (EVG − RAL) was 2.2%, and the 95% CI was −5.0% to 9.3%. Outcomes for the percentages of subjects with HIV-1 RNA < 50 copies/mL at Week 96 were similar for the EVG and RAL groups, and demonstrate durable efficacy. Based on a TLOVR analysis, 47.6% of subjects (167 of 351 subjects) in the EVG group and 45.0% of subjects (158 of 351 subjects) in the RAL group were classified as responders. The stratum-adjusted difference between groups (EVG − RAL) was 2.6%, and the 95% CI was −4.6% to 9.9%. Based on the snapshot analysis, 52.4% of subjects (184 of 351 subjects) in the EVG group and 53.0% of subjects (186 of 351 subjects) in the RAL group were classified as a virologic success. The stratum-adjusted difference between treatment groups (EVG − RAL) was −0.5%, and the 95% CI was −7.9% to 6.8%.
3. Analyses of the other secondary virologic endpoints support the primary endpoint and provide further robustness to the overall analysis of efficacy.
4. The percentages of subjects with PVF were similar in the EVG and RAL treatment groups. At Week 96, the KM estimates for the percentages of subjects with PVF for HIV-1 RNA cutoff at 50 copies/mL were 45% for the EVG group and 46% for the RAL group. The median time to PVF was 1014 days in the EVG group and 961 days in the RAL group (p = 0.99).
5. CD4 cell counts increased following administration of study drug, and the mean increases were similar between the EVG and RAL treatment groups at all time points. At Week 96, the mean (SD) increases from baseline in CD4 cell count were 205 (191.5) cells/mm3 in the EVG group and 198 (162.2) cells/mm3 in the RAL group.

The reviewer's assessment and additional analyses

1. Assessment for Noninferiority margin

   The NI margin is prespecified at 10%. Here we assess whether this margin is reasonable.

   We identified the two registrational placebo-controlled randomized double blinded trials, for Raltegravir, to assess the noninferiority margin. Roughly speaking, the margin should be defined well so that EVG's noninferiority to RAL implies EVG's superiority to Placebo with a high confidence. For this purpose, we generally define the margin as the lower bound of the 95% confidence interval of the rate difference of RAL vs. Placebo.

   The results from these two trials, Benchmrk trials were available from the publication (Cooper et al., 2008; Steigbigel et al., 2008). Overall, the treatment effect of Raltegravir vs. Placebo in risk difference is 29% with a 95% C.I. of (22%, 37%). For subjects with phenotypical sensitivity score (PSS) =1 or 2, the treatment effect of Raltegravir vs. Placebo in risk difference is 32% with a 95% C.I. of (23%, 42%). For subjects with PSS=2, the treatment effect of Raltegravir vs. Placebo in risk difference is 33% with a 95% C.I. of (18%, 47%). For subject with more than one active protease inhibitors in their BR, the treatment effect of Raltegravir vs. Placebo in risk difference is 22% with a 95% C.I. of (12%, 32%). For subject first use of Darunavir in their BR, the treatment effect of Raltegravir vs. Placebo in risk difference is 23% with a 95% C.I. of (9%, 37%). For all these examination, including overall and subgroups, only the 9%, the lower bound of 95% C.I. for the risk difference in subject first use of Darunavir in their BR, is below 10%. However, this lower bond of 9% is likely due to the small sample size (188) noting that the point estimate of the risk difference is 23%. See Table 9.

   On the other hand, for subjects who have Etravirine or Maraviroc in their BR, we cannot establish a meaningful noninferiority margin as subjects in Benchmark trials did not use these drugs (they were not approved at that time). We shall assess this particular issue later in this review.

   We conclude that the noninferiority margin of 10% is reasonable, except possible for subjects who have Etravirine or Maraviroc in their BR.

Table 9: NI margin determination

Week 48	RAL (response/all sample)	Placebo (response/all sample)	reference	Risk diff (95% C.I)
Overall	285/459	78/237	Table 2, Steigbigel et al (2008)	29% (22%, 37%)
Subjects with PSS=1 or 2	182/276	44/131	Figure 1, Cooper et al (2008)	32% (23%, 42%)
Subjects with PSS= 2	99/139	24/62	Figure 1, Cooper et al (2008)	33% (18%, 47%)
Subject with more than one active protease inhibitors in OBT	188/264	64/130	Figure 1, Cooper et al (2008)	22% (12%, 32%)
Subjects first use of Darunavir	91/119	37/69	Figure 3, Cooper et al (2008)	23% (9%, 37%)
Subjects with Etravirine or Maraviroc in BR	NA	NA	NA	No margin

Source: reviewer's analysis.

2. Obtain the primary endpoint using the snapshot algorithm with the standardized analysis window. The pre-specified primary efficacy endpoint for trial 145, proportion of subjects responding by the FDA Time to Loss of Virologic Response (TLOVR) algorithm at Week 48, is appropriate. However, FDA reviewers now prefer a simpler endpoint, the proportion of subjects responding at Week 48, referred as Snapshot approach. We expect that the snapshot algorithm facilitates the communication between FDA and industry.

   Using the analysis window specified in the FDA's snapshot guidance, both the sponsor and the reviewer obtained the same results (Table 10), which are very similar to what the sponsor originally submitted based on their proposed analysis window.

   In addition, the virologic response rates in the RAL arm observed in trial 145 are 58% at week 48 and 54% at week 96. These rates are similar to the corresponding virologic success rates observed in Benchmrk trials.

Table 10: Snapshot results based on the FDA's analysis window.

	Week 48**(Days 294 to 377)**		Week 96 (**Days 630 to 713)**	
	EVG (n=351)	RAL (n=351)	EVG (n=351)	RAL (n=351)
HIV-1 RNA < 50 copies/mL	**211 (60.1%)**	**205 (58.4%)** 60%	**184 (52.4%)**	**188 (53.6%)** 55%
Virologic Failure	**115 (32.8%)**	**110 (31.3%)**	**128 (36.5%)**	**109 (31.1%)**
HIV-1 RNA >= 50 copies/mL	57 (16.2%)	65 (18.5%)	44 (12.5%)	31 (8.8%)
Discontinued Study Drug Due to Lack of Efficacy	9 (2.6%)	6 (1.7%)	14 (4.0%)	19 (5.4%)
Discontinued Study Drug Due to Other Reasons and Last Available HIV-1 RNA >= 50 copies/mL	48 (13.7%)	37 (10.5%)	67 (19.1%)	57 (16.2%)
HIV-1 RNA >= 50 copies/mL at Background Regimen Switch and HIV-1 RNA < 50 copies/mL at Week 48	1 (0.3%)	2 (0.6%)	3 (0.9%)	2 (0.6%)
No Virologic Data in Week 96 Window	**25 (7.1%)**	**36 (10.3%)**	**39 (11.1%)**	**54 (15.4%)**
Discontinued Study Drug Due to AE or Death	8 (2.3%)	18 (5.1%)	9 (2.6%)	24 (6.8%)
Discontinued Study Drug Due to Other Reasons and Last Available HIV-1 RNA < 50 copies/mL	14 (4.0%)	17 (4.8%)	28 (8.0%)	30 (8.5%)
Missing Data during Window but on Study Drug	3 (0.9%)	17 (8.4%)	2 (0.6%)	0 (0%)

Source: reviewer's analysis, consistent with the sponsor's result. The highlighted virologic success rate for RAL are obtained from the ISENTRESS's label (current and previous versions)

3. In this application, there is essentially one single trial to support EVG in the targeted population. We would like to see whether there is sufficient quantity of evidence to support effectiveness of EVG. It has been FDA's position that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness.

   The sponsor originally planned two identical trials with one conducted in US and Puerto Rico and the other in Europe, Australia, Canada, and Mexico. Both Phase 3 trials were identical in design with the same 1:1 randomization ratio and there were independent unique patient identification numbers across both studies. Given the declining numbers of patients with unsuppressed viremia, 7 months after the first screening subject visit, Gilead unified two trials.

   Using the sponsor's original snapshot results at week 48, the (1-0.000625) % confidence interval of the virologic response rate between EVG and RAL is (-

10.4%, 15.0%). Therefore the 99.935% C.I. is still about -10%. It implies that the significance level for meeting the noninferiority criteria is $0.000625=0.025^2$, suggesting the single trial had the significance level equivalent to that two trials would achieve. Using the FDA's snapshot results, the (1-0.000625)% confidence interval of the virologic response rate between EVG and RAL is (-11.0%, 14.4%). According to the results in our assessment for noninferiority margin, -11% is generally acceptable and the superiority of EVG to Placebo can be established with high confidence.

Based on the results, we conclude that the efficacy results based on the single trial is sufficient enough to support the efficacy of EVG.

In addition, EVG is also a component in the 4-drug fixed-dose combination tablet (the QUAD single-tablet regimen [STR], Stribild®) which is comprised of EVG, a pharmacokinetic enhancer cobicistat (COBI), and the current standard-of-care dual NRTI/NtRTI backbone FTC/TDF (Truvada® [TVD]). Stribild® is approved August 2012 indicated for HIV infected, treatment naïve adults and the approval further provides some indirect support of EVG.

4. Modified stratified analysis. The trial 145 was a combination of the old trials 144 and 145, with 144 conducted in the United States (US) and 145 outside of US; therefore the geographic region becomes a natural stratification factor in the combined trial 145. The current stratum-adjusted analysis the sponsor performed only include the original stratification factors: baseline viral load and the second agent in the background drug. In our analysis we include the additional geographic region. Stratified analysis based on stratified by HIV RNA levels, class of the second agent in the BR, and region (US vs. non USA).

   Using the data listing in Table 11, we redo the stratified analysis (CMH analysis for risk difference). In the sponsor's analysis, the percentage of subjects with HIV-1 RNA < 50 copies/mL at Week 48 (snapshot analysis, ITT) was similar in the EVG and RAL treatment groups: 60% of subjects in the EVG group and 58% of subjects in the RAL group were classified as a virologic success. The stratum-adjusted (with three strata) difference between treatment groups (EVG − RAL) was 1.9%, and the 95% CI was (−5.0% to 8.9%). The percentage of subjects with HIV-1 RNA < 50 copies/mL at Week 96 (snapshot analysis, ITT) was similar in the EVG and RAL treatment groups: 52.4% of subjects (184 of 351 subjects) in the EVG group and 53.0% of subjects (186 of 351 subjects) in the RAL group were classified as a virologic success. The stratum-adjusted difference between treatment groups (EVG − RAL) was −1.0%, and the 95% CI was (−8.1% to 6.2%). Although not the same, the results are generally consistent with the stratified analysis based on only two strata: baseline viral load and the second agent in the baseline background therapy.

Table 11: Snapshot results by stratification factors.

Region	Baseline HIV RNA	Second agent	Sample size in EVG	Sample size in RAL	Virologic response in EVG	Virologic response in RAL
Week 48						
Non USA	baseline HIV RNA>100,000 copies/ml	Non NRTI	4	5	0	2
		NRTI	25	27	16	13
	baseline HIV RNA<=100,000 copies/ml	Non NRTI	27	30	24	22
		NRTI	64	61	48	42
USA	baseline HIV RNA>100,000 copies/ml	Non NRTI	10	12	5	5
		NRTI	51	46	18	14
	baseline HIV RNA<=100,000 copies/ml	Non NRTI	27	22	17	14
		NRTI	143	148	83	93
Week 96						
Non USA	baseline HIV RNA>100,000 copies/ml	Non NRTI	4	5	1	1
		NRTI	25	27	12	12
	baseline HIV RNA<=100,000 copies/ml	Non NRTI	27	30	21	22
		NRTI	64	61	44	40
USA	baseline HIV RNA>100,000 copies/ml	Non NRTI	10	12	6	6
		NRTI	51	46	17	15
	baseline HIV RNA<=100,000 copies/ml	Non NRTI	27	22	15	12
		NRTI	143	148	68	80

Source: the review's analysis.

5. Assessment of newly approved drugs that were not used in Benchmrk trials.

   For subjects with Etravirine or Maraviroc in BR, no meaningful noninferiority margin could be defined as subjects in Benchmrk did not take these medications. However, we report that EVG is numerically better than RAL in subjects who had these two drugs in their BR.

   We assess the noninferiority of EVG to RAL using a newly developed method, hybrid design approach, (Soon et al., 2011). This approach represents a conservative approach in this case.

   At week 48, the noninferiority of EVG relative to RAL is significant at a one-sided level of P_2=0.0025 in subjects who did not have Etravirine or Maraviroc in their BR; the superiority of EVG relative to RAL is not significant and the one-

sided p-value from Fisher's exact test is P_1=0.43 in subjects who have Etravirine or Maraviroc in their BR. Combining two p-values through the Fisher's combination rule, we obtained a hybrid p-value of 0.008. See **Table 12**. The small p-value implies that EVG is noninferior to RAL. If we would allow the noninferiority margin to be 5% (half of the pre-defined 10%) for subjects who have Etravirine or Maraviroc in their BR, then the noninferiority of EVG relative to RAL (with a NI margin of 5%) is significant at a one-sided level of P_{10}=0.16 in subjects who have Etravirine or Maraviroc in their BR. Combining two p-values (P_2=0.0025 and P_{10}=0.16), through the Fisher's combination rule, we obtained a hybrid p-value of 0.004. See **Table 12**.

At week 96, the noninferiority of EVG relative to RAL is significant at a one-sided level of P_2=0.04 in subjects who did not have Etravirine or Maraviroc in their BR; the superiority of EVG relative to RAL is not significant and the one-sided p-value from Fisher's exact test is P_1=0.3 in subjects who have Etravirine or Maraviroc in their BR. Combining two p-values through the Fisher's combination rule, we obtained a hybrid p-value of 0.07. If we would allow the noninferiority margin of 5% (half of the pre-defined 10%) for subjects who have Etravirine or Maraviroc in BR, then the noninferiority of EVG relative to RAL (with a margin of 5%) is significant at a one-sided level of P_{10}=0.1 in subjects who have Etravirine or Maraviroc in BR. Combining two p-values (P_2=0.04 and P_{10}=0.1) through the Fisher's combination rule, we obtained a hybrid p-value of 0.026. See **Table 12**.

Table 12: Evaluation based on hybrid design (snapshot data at week 48)

Etravirine or Maraviroc in BR	EVG	RAL		Combinational pvalue
	Week 48			
Yes	67% (46/69)	64% (46/72)	P_1=0.43	P=0.008
No	59% (165/282)	57% (159/279)	P_2=0.0025	
Yes	67% (46/69)	64% (46/72)	P_1=0.16	P=0.004
No	59% (165/282)	57% (159/279)	P_2=0.0025	
	Week 96			
Yes	64% (44/69)	58% (42/72)	P_1=0.3	P=0.07
No	50% (140/282)	52% (146/279)	P_2=0.04	
Yes	64% (44/69)	58% (42/72)	P_{10}=0.1	P=0.026
No	50% (140/282)	52% (146/279)	P_2=0.04	

These analyses generally support EVG's noninferiority to RAL despite a total of 141 subjects (20%) in the trial 145 have Etravirine or Maraviroc in their BR while subjects from the reference trial, Benchmrk, trials did not.

6. Heterogeneity of treatment effect of EVG relative to RAL.

Heterogeneity of treatment effect of EVG relative to RAL, expressed in the primary endpoint (virologic response, <50 copies/ml) were observed in gender, in race, and possibly in region at some time points. We communicated with the sponsor seeking for explanations of these heterogeneities. The sponsor confirmed the heterogeneity and stated "clinical relevance of the observed statistical heterogeneities is unclear". They further concluded (b) (4) The sponsor argued that the treatment heterogeneity found in gender and race could be due to multiple testing and that the other important endpoint did not reveal the heterogeneity in gender or in race. We consider the arguments valid.

In below, we first illustrate the observed heterogeneity; we next conduct a thorough assessment of the impact of heterogeneity, followed by our conclusion based on the current data.

In the following Table **13**, we present the subgroup analysis according to race, region (US vs. non US), and gender. For region, we consider two options: 1) Puerto Rico was considered as UA region; 2) Puerto Rico was not considered as UA region. According to the original trial design, the trial 144 is conducted in US including Puerto Rico; we shall use option 1) in the final conclusion.

Table 13: Heterogeneity in primary endpoint (snapshot algorithm on ITT population)

	EVG (n=351)	RAL (n=351)	EVg vs RAl odds ratio (95% C.I.)	p-value of treatment by subgroup interaction in OR: unadjusted (adjusted for stratification factors) logistic regression
Week 48				
Race White Nonwhite	 66%(140/211) 51%(71/140)	 60%(135/226) 56%(70/125)	 1.33(0.90,1.96) 0.81(0.50,1.31)	0.12 (0.08)
Region US Non US	 53%(123/231) 73% (88/120)	 55%(126/228) 64% (79/123)	 0.92(0.64,1.33) 1.53(0.89,2.65)	0.13 (0.14)
Region US* Non US*	 53%(119/224) 72% (92/127)	 57% (122/215) 61% (83/136)	 0.86 (0.59, 1.26) 1.68 (1.00, 2.82)	0.04 (0.05)
Gender Female Male	 47%(28/59) 63%(183/292)	 63%(42/67) 57%(163/284)	 0.54(0.26,1.10) 1.25(0.89,1.74)	0.04 (0.08)
Week 96				
Race White Nonwhite	 61%(128/211) 40%(56/140)	 55%(125/226) 50%(63/125)	 1.25(0.85,1.82) 0.66(0.40,1.07)	0.04 (0.03)
Region US Non US	 46%(106/231) 65% (78/120)	 50%(113/228) 61% (75/123)	 0.86(0.60,1.25) 1.19 (0.71,2.00)	0.33 (0.33)
Region US* Non US*	 46%(102/224) 65% (82/127)	 50% (108/215) 59% (80/136)	 0.83 (0.57, 1.21) 1.28 (0.78, 2.10)	0.18 (0.19)
Gender Female Male	 39%(23/59) 55%(161/292)	 52%(35/67) 153/284(54%)	 0.58(0.29,1.19) 1.05(0.76,1.46)	0.14 (0.23)

Source: reviewer's analysis (US include Puerto Rico and US* does not)

The notable significance of heterogeneity is observed in race (p=0.03) at week 96 and in gender (p=0.04) at week 48. We ignore the significance of heterogeneity in region at week 48 as we consider region US (rather than US*) is more reasonable. These are observed heterogeneities; however we do not have high confidence to conclude that these are definitive clinical heterogeneities. In addition, we do not think it will affect the approvability of the drug even they are true signals for several reasons described below.

- In addition to these three important subgroups analyses, we also conducted at least two other subgroup analyses on the two stratification factors. No notable heterogeneity was found in the two stratification factors. That implies that we have conducted at least 10 subgroup analyses (5 at each of week 48 and week 96). The multiple subgroup analyses increase the chance of false claim if we do not adjust for the multiplicity. In fact, none of the p-values of 0.03 or 0.04 would meet reasonable multiplicity adjustment criteria. Furthermore, heterogeneity for race is observed at week 96 but not at level of 0.1 at week 48; heterogeneity for gender is observed at week 48 but not at level of 0.1 at week 96.
- The main concern of potential heterogeneity here is related to a worry that EVG would not be at least better than placebo in some subgroups such as female subjects or nonwhite subjects. These concerns are alleviated through the following analyses. The response rate of female subjects in EVG group is 48% (28/59) at week 48, which is higher than the 26%, the response rate of female subjects in placebo group in BENCHMRK trials. Although the sample size is very small, the difference between EVG and Placebo rate in female subjects is significant with a two-sided pvalue of 0.097 using the Fisher's exact test and 0.06 using the likelihood ratio chi-square test. This analysis does support that EVG would be better than placebo if they were compared head to head. The response rate of nonwhite subjects in RAL group is 61%, which is higher than the 23%, the response rate of nonwhite subjects in placebo group in BENCHMRK trials. The lower bound of the 95% C.I. of the risk difference between RAL and placebo is 23%. As the lower bound of the 95% C.I. of the risk difference of EVG and RAL is -17%, whose absolute value is less than 23%, we are confident that EVG would be at least better than Placebo in nonwhite groups. Here we use week 48 data in (Cooper et al., 2008) as reference.
- Heterogeneity is not revealed in any of these factors using TLOVR algorithm, the original primary endpoint defined in the protocol. The difference between heterogeneity revealed through two algorithms (Snapshot and TLOVR) was investigated and the difference is primarily due to a total of 73 discordances between two algorithms which accounts for more than 10% of difference. Among them, 12 subjects are classified as virologic failure using the snapshot algorithm due to a single blip at Week 96 and these subjects had sustained virologic response through Week 96 using the TLOVR algorithm; 2) 54

subjects (22 in EVG and 32 in RAL) were classified as having rebound before Week 96 using the TLOVR algorithm, but resuppressed at Week 96 and therefore were classified as responder using the snapshot algorithm. There are 2 in EVG and 3 in RAL had switched background drug and were considered failure.

All 54 subjects achieved confirmed viral suppression (HIV-1 RNA < 50 copies/mL), followed by confirmed viral rebound (≥ 50 copies/mL) prior to the upper bound of Week 96 analysis window. However, all these 54 subjects had last HIV-1 RNA levels < 50 copies/mL in the Week 96 analysis windows.

- 43 of the 54 subjects had confirmed HIV-1 RNA rebound between ≥ 50 and < 400 copies/mL after initially achieving confirmed HIV-1 RNA < 50 copies/mL;
- 11 subjects who had HIV-1 RNA levels rebound to ≥ 400 copies/mL after achieving confirmed HIV-1 RNA levels of < 50 copies/mL, resistance testing was performed. Eight of the 11 subjects did not develop new resistance mutations and the remaining three subjects developed new resistance mutations; none developed integrase inhibitor resistance-associated mutations.

The 43 subjects will not be classified as failure if 400 copies/ml was the threshold. In addition, the discordance rate of 10% is similar to the discordance rate of 9.4% in Benchmrk 018 and 8.6% in Benchmrk 019,

Note that, the above discussion of heterogeneity is mostly about approvability of EVG, where we focus on whether EVG should at least better than placebo. The discussion is however not about comparing EVG with RAL.

The following table, Table 14, provides the direct comparison of EVG and RAL.

Table 14: subgroup analyses by gender and race

	EVG (n=351)	RAL (n=351)	EVG vs RAl risk difference (95% C.I.)
Week 48			
Gender			
Female	47%	63%	-15%(-32%, 2.0%)
Male	63%	57%	5.3%(-2.7%,13%)
Week 96			
Race			
White	61%	55%	5.4%(-3.9%,14.6%)
Nonwhite	40%	50%	-10%(-22.3%,1.5%)

Source: reviewer's analysis (US include Puerto Rico and US* does not), (Cooper et al., 2008).

7. It appears that the rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy in 0145 is remarkably higher than the same rate in Benchmark trials.

   Gilead acknowledges that the overall rates of study drug discontinuation at Week 96 in Trial GS-US-183-0145 (EVG 41.2% vs. RAL 41.9%) is balanced between the two groups (EVG vs. RAL) but higher than those in the BENCHMRK studies (RAL 15.0% vs. Placebo 17.3%). Specifically, the rates of study drug discontinuation due to reasons other than adverse event, death, lack of efficacy, or pregnancy were generally higher in both treatment groups of Trial GS-US-183-0145 than in those of BENCHMRK trials

   Gilead considers that these observations may be associated with change in the medical landscape in antiretroviral treatment since the initiation of the BENCHMRK clinical studies. At the time of initiation of GS-US-183-0145 (June 2008), a few novel and potent antiretroviral agents became available that had not been at the time of initiation of BENCHMRK (February 2006): in the US Darunavir was approved in June 2006, RAL in October 2007, and Etravirine in January 2008. These new agents made possible for clinicians to construct effective regimens that could reestablish virologic suppression (HIV-1 RNA < 50 copies/mL) in many patients, even the most treatment-experienced patients who are failing current therapy. As a result, although the trial populations of both studies were treatment-experienced patients, those in the BENCHMRK studies were more likely to be failing therapy at baseline due to lack of available active agents, compared with subjects in Trial GS-US-183-0145. In BENCHMRK, the absence or paucity of available therapeutic options outside the trial likely led to greater motivation for each subject to adhere to study drug and procedures. In contrast, subjects in Trial GS-US-183-0145 may not have had as much motivation

to remain in the trial due to availability of potent and active agents outside the trial.

Gilead conducted additional analyses of Trial GS-US-183-0145 to evaluate the effect of potential antiretroviral treatment options at baseline on the rates of study drug discontinuation due to reasons other than adverse event, death, lack of efficacy, or pregnancy. At Week 96 data cut, subjects with more treatment options at baseline (i.e. 0 to 1 class resistance) had higher rates of study drug discontinuation due to administrative reasons; non-compliance, lost to follow-up, withdrawal of consent, protocol violation and investigator's discretion (EVG 37.5% [48/128] vs RAL 42.7% [59/138]) than those without any treatment options (2 or 3 class resistance) in both groups (EVG 28.3% [63/223] vs RAL 17.8% [38/213]) per subject disposition. This new analysis further support our interpretation that the higher study drug discontinuation rate observed in Trial GS-US-183-0145, compared with the BENCHMRK studies, is a result of a complex interplay between the behavior of study subjects, their disease characteristics (i.e. antiretroviral resistance) and the medical landscape of antiretroviral treatment before and during the study.

The reviewer did not find any data that contradict to the sponsor's explanation. In addition, the virologic response rates in the RAL arm observed in trial 145 are 58% at week 48 and 54% at week 96. These rates are similar to the corresponding virologic success rates observed in Benchmrk trials. That means, we could interpret with a certain level of comfort, that most of the subjects who discounted the trial while their last viral loads >50 copies/ml in the trial will not have virologic success even if they were not discontinued.

3.3 Evaluation of Safety

Please refer to the clinical review By Dr. Fleischer, Russell D.

3.4 Benefit-Risk Assessment (Optional)

These two drugs belong to the same drug classes. They appear to have comparable efficacies. Although some heterogeneities were observed in gender, race, and possible in region, we do not have high confidence to declare these are real heterogeneities. We hope that some further studies could be conducted to re-assess these heterogeneities in e.g. gender and race.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The virologic response at week 96 of EVG vs. RAL is similar in subjects with older (>=45 years old) and younger age (<45), see Table 15. The virologic response at week 48 of EVG vs. RAL is also similar in subjects with older and younger age. Heterogeneity in gender, race, and region has been discussed earlier and we will present some additional information about them in Section 4.2 together with other factors.

Table 15: Age analysis of snapshot results at week 96

		EVG	RAL	P value (Fisher's exact test)
Age (years)	>=45	54.42%(80/147)	55.62%(94/169)	0.909773
	<45	50.98%(104/204)	51.65%(94/182)	0.918992

Source: reviewer's analysis.

4.2 Other Special/Subgroup Populations

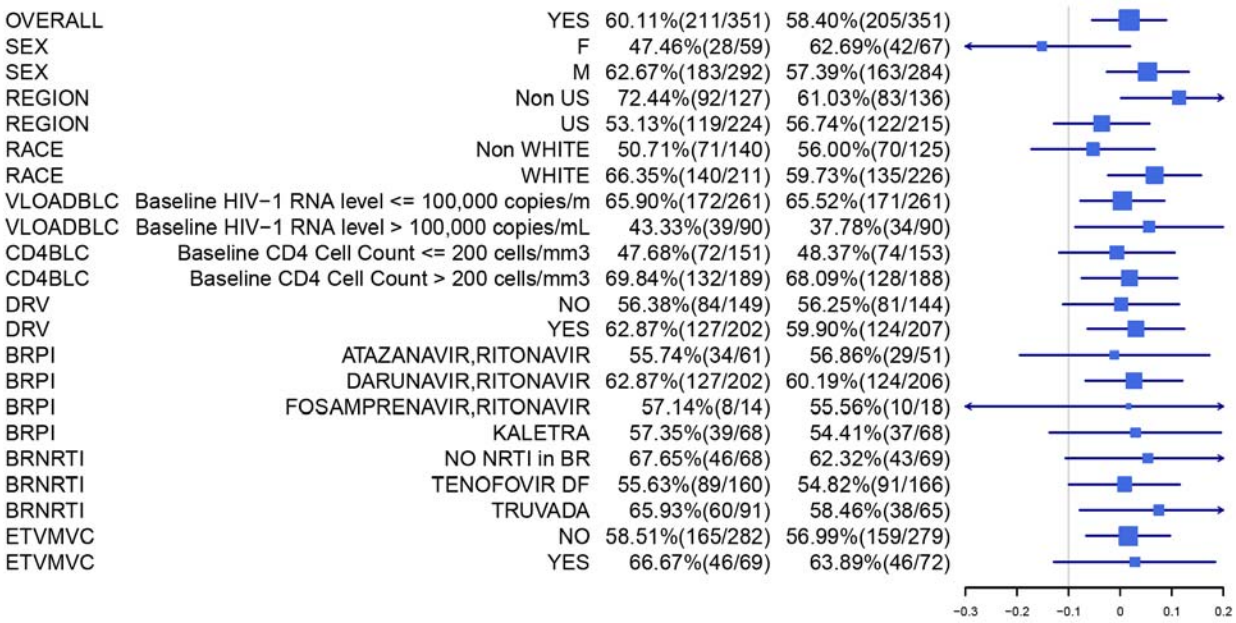
Other subgroup analyses have been performed. No significant treatment heterogeneity is discovered except for gender, race, and geographic region (USA vs. others including Puerto Rico). The three factors were discussed in details earlier.

Table 16: subgroup analysis of snapshot results (week 48)

Demographic/baseline variables	levels	EVG	RAL	Pvalue (Fisher's exact test)
overall		60.11%(211/351)	58.40%(205/351)	0.700961
Gender	F	47.46%(28/59)	62.69%(42/67)	0.106642
	M	62.67%(183/292)	57.39%(163/284)	0.202869
USA*	No	72.44%(92/127)	61.03%(83/136)	0.066853
	Yes	53.13%(119/224)	56.74%(122/215)	0.501923
USA	No	73.33%(88/120)	64.23%(79/123)	0.130912
	Yes	53.25%(123/231)	55.26%(126/228)	0.708046
White	Non WHITE	50.71%(71/140)	56.00%(70/125)	0.459455
	WHITE	66.35%(140/211)	59.73%(135/226)	0.165886
Baseline HIV-1 RNA level	<= 100,000 copies/ml	65.90%(172/261)	65.52%(171/261)	1.000000
	> 100,000 copies/mL	43.33%(39/90)	37.78%(34/90)	0.543852
Baseline CD4 Cell Count	<= 200 cells/mm3	47.68%(72/151)	48.37%(74/153)	0.909180
	> 200 cells/mm3	69.84%(132/189)	68.09%(128/188)	0.739221
Darunavir	not in the BR	56.38%(84/149)	56.25%(81/144)	1.000000
	In the BR	62.87%(127/202)	59.90%(124/207)	0.544155
PI in the BR	ATAZANAVIR,RITONAVIR	55.74%(34/61)	56.86%(29/51)	1.000000
	DARUNAVIR,RITONAVIR	62.87%(127/202)	60.19%(124/206)	0.611510
	FOSAMPRENAVIR,RITONAVIR	57.14%(8/14)	55.56%(10/18)	1.000000
	KALETRA	57.35%(39/68)	54.41%(37/68)	0.862979
Second agent in the BR	not NRTI	67.65%(46/68)	62.32%(43/69)	0.592069
	NRTI	58.30%(165/283)	57.45%(162/282)	0.864872
Etravirine or Maraviroc in BR	No	58.51%(165/282)	56.99%(159/279)	0.732965
	YES	66.67%(46/69)	63.89%(46/72)	0.859714

Source: Reviewer's analysis. For the region variable, variable USA* does not include Puerto Rico as part of USA and variable USA does. ETVMVC=1 means subjects have Etravirine or Maraviroc in BR,

Figure 3: subgroup analysis of snapshot results (week 48)



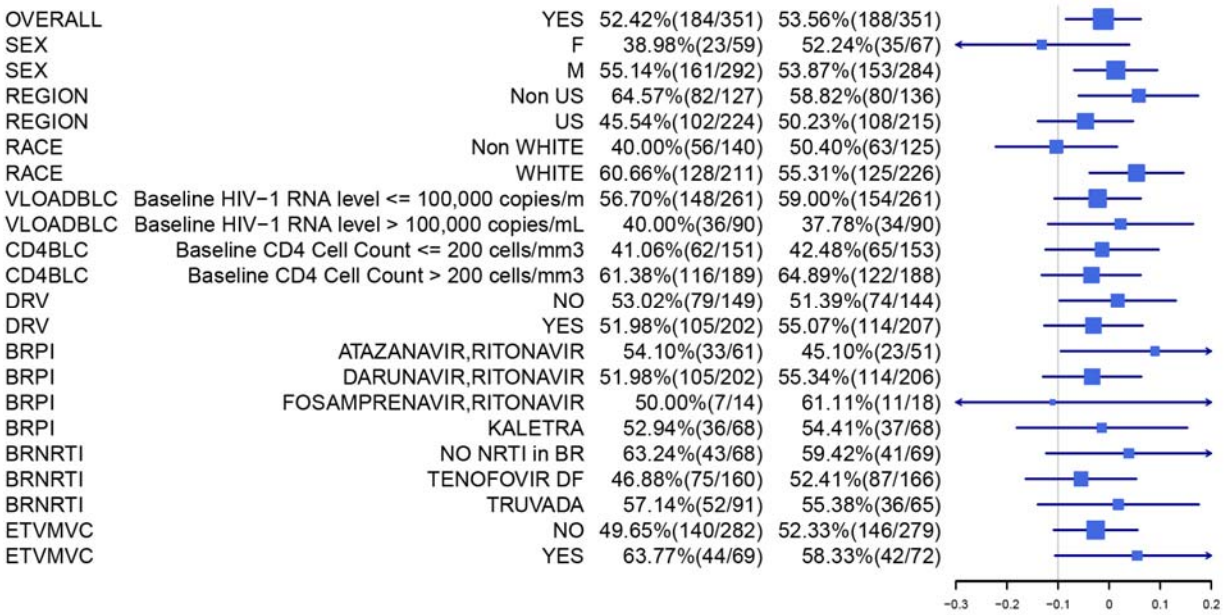
Source: Reviewer's analysis. ETVMVC=YES means Etravirine or Maraviroc is in BR; BRPI and BRNRTI means the first and second agent in BR. DRV=YES means Darunavir in BR. Region is consistent with USA* in **Table 16**.

Table 17: subgroup analysis of snapshot results (week 96)

Demographic/baseline variables	level	EVG	RAL	Pvalue (Fisher's exact test)
overall		52.42%(184/351)	53.56%(188/351)	0.820552
Gender	F	38.98%(23/59)	52.24%(35/67)	0.154712
	M	55.14%(161/292)	53.87%(153/284)	0.801891
USA*	No	64.57%(82/127)	58.82%(80/136)	0.375406
	Yes	45.54%(102/224)	50.23%(108/215)	0.340323
USA	No	65.00%(78/120)	60.98%(75/123)	0.595299
	Yes	45.89%(106/231)	49.56%(113/228)	0.455366
White	Non WHITE	40.00%(56/140)	50.40%(63/125)	0.107817
	WHITE	60.66%(128/211)	55.31%(125/226)	0.286461
Baseline HIV-1 RNA level	<= 100,000 copies/ml	56.70%(148/261)	59.00%(154/261)	0.65767
	> 100,000 copies/mL	40.00%(36/90)	37.78%(34/90)	0.878550
Baseline CD4 Cell Count	<= 200 cells/mm3	41.06%(62/151)	42.48%(65/153)	0.816985
	> 200 cells/mm3	61.38%(116/189)	64.89%(122/188)	0.522235
Darunavir	not in the BR	53.02%(79/149)	51.39%(74/144)	0.815538
	In the BR	51.98%(105/202)	55.07%(114/207)	0.552855
PI in the BR	ATAZANAVIR,RITONAVIR	54.10%(33/61)	45.10%(23/51)	0.448047
	DARUNAVIR,RITONAVIR	51.98%(105/202)	55.34%(114/206)	0.551453
	FOSAMPRENAVIR,RITONAVIR	50.00%(7/14)	61.11%(11/18)	0.721266
	KALETRA	52.94%(36/68)	54.41%(37/68)	1.000000
Second agent in the BR	Not NRTI	63.24%(43/68)	59.42%(41/69)	0.726487
	NRTI	49.82%(141/283)	52.13%(147/282)	0.613942
Etravirine or Maraviroc in BR	No	49.65%(140/282)	52.33%(146/279)	0.554736
	YES	63.77%(44/69)	58.33%(42/72)	0.604766

Source: reviewer's analysis. Variables are the same as those in **Table 16**.

Figure 4: subgroup analysis of snapshot results (week 96)



Source: Reviewer's analysis. ETVMVC=YES means Etravirine or Maraviroc is in BR; BRPI and BRNRTI means the first and second agent in BR. DRV=YES means Darunavir in BR. Region is consistent with USA* in Table 17.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

In this review, we identified some issues. We conducted analyses to address all of them and requested explanations from the sponsor to address some of them. In below, we briefly describe the issues and their resolutions.

1. We found some randomization errors which were confirmed by the sponsor. A total of 30 subjects were incorrectly stratified during randomization due to the data entry errors that occurred when sites randomized subjects to the study and did not report the errors. The sponsor claimed that "The stratification errors either not reported by study sites or reported by study sites have no impact on statistical analysis for efficacy because Gilead used baseline HIV-1 RNA data and the second agent (NRTI or other) at baseline to re-classify subjects' strata in efficacy analysis." Considering errors are inevitable, the review agrees with the sponsor that the errors identified shall not impact the results.
2. In this submission, there is essentially one single trial to support the drug registration of EVG in the targeted population. The analysis the reviewer conducted suggests that the evidence from the trial 145 is as strong as two trials

would provide. We conclude that the evidence based on the single trial is sufficient enough to support EVG.

3. Two "newly" approved drugs, Etravirine or Maraviroc, were used in trial 145 as a part of BR but they were not sued in Benchmrk trials, the reference trials for RAL. Using a newly developed method, hybrid design, the reviewer is able to conclude that the evidence is sufficient to support the noninferiority of EVG relative to RAL, despite about 20% of subjects in trial 145 had the two newly approved drugs in their background therapy.
4. Heterogeneity of treatment effect in gender, race and possible region was observed. We investigate whether they are "real" signal or not. Most importantly, we examine whether they cast doubt on the approvability of the drug or whether they suggest some restrictions of the drug in some subgroups. With the sponsor's arguments (that heterogeneity found in gender and race could be due to multiple testing and the other endpoint did not reveal heterogeneity) and two supporting analyses that the reviewer conducted, the reviewer concludes that the potential treatment heterogeneity dose not affect the approvability as the data support EVG's superiority to placebo, in male and female subjects, in white or non white subjects, see section 3.2.4, reviewer's analysis.
5. Analysis of large percentage of discontinuations. It appears that the rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy in 0145 is remarkably higher than the same rate in Benchmark trials. Gilead acknowledges that the overall rate of study drug discontinuation at Week 96 in trial 145 is balanced between the two groups. Gilead considers that these observations may be associated with change in the medical landscape in antiretroviral treatment since the initiation of the BENCHMRK clinical studies. At the time of initiation of GS-US-183-0145 (June 2008), a few novel and potent antiretroviral agents became available that had not been at the time of initiation of BENCHMRK (February 2006): in the US Darunavir was approved in June 2006, RAL in October 2007, and Etravirine in January 2008. These new agents made possible for clinicians to construct effective regimens that could reestablish virologic suppression (HIV-1 RNA < 50 copies/mL) in many patients, even the most treatment-experienced patients who are failing current therapy. As a result, although the study populations of both studies were treatment-experienced patients, those in the BENCHMRK studies were more likely to be failing therapy at baseline due to lack of available active agents, compared with subjects in Study GS-US-183-0145. In BENCHMRK, the absence or paucity of available therapeutic options outside the study likely led to greater motivation for each subject to adhere to study drug and procedures. In contrast, subjects in Study GS-US-183-0145 may not have had as much motivation to remain in the study due to availability of potent and active agents outside the study. The reviewer did not find any data that contradict to the sponsor's explanation.

5.2 Collective Evidence

The sponsor intended to use a much smaller Phase 2 Trial GS-US-183-0105 and the single arm roll-over Trial GS-US-183-0130 as supportive studies. However, these two trials are not independent with almost all subjects in trial 130 is from trial 105. In

addition, the doses of EVG used in trial 105 are different from the registration doses. Finally, the subjects in the EVG arms vs. subjects in RAL are not blinded. Consequently, the reviewer only relied on trial 145 to assess the evidence of efficacy.

5.3 Conclusions and Recommendations

The sponsor concluded that RAL is noninferior to Isentress, which was approved by the United States (US) Food and Drug Administration (FDA) on 12 October 2007.

The reviewer concurs with the sponsor's above conclusion in general. The concurrence comes after the reviewer carefully assessed many outstanding statistical issues, such as one single trial was submitted to support the drug registration of EVG; newly approved HIV drugs were used in trial 145 while they were not used in historical trials of RAL; heterogeneity is observed in gender and race without clear explanation; a remarkably higher rate of discontinuation due to reasons other than adverse event, death, lack of efficacy, pregnancy higher than in reference trials for RAL.

While the reviewer concludes that the potential treatment heterogeneity dose not affect the approvability, we note that, nevertheless, among male subjects 63% (55%) and 57% (54%) had HIV-1 RNA <50 copies/mL at week 48 (week 96) in the EVG and RAL treatment arms, respectively. Among female subjects, 47% (39%) and 63%(52%) had HIV-1 RNA <50 copies/mL at week 48 (week 96) in the EVG and RAL treatment arms, respectively. It could be a public interest to assess whether the observed poor performance of EVG relative to RAl (-15% at week 48 and -13%) at week 96 in women, is a real signal or not in future trials.

5.4 Labeling Recommendations (as applicable)

The sponsor proposed the following language in their label (b) (4)

The language, as currently stated in the label, may not be appropriate (b) (4)

This particular interpretation has a great uncertainty and might be wrong.


(b) (4)

APPENDICES (Add When Needed)

Reference

Cooper, D. A., Steigbigel, R. T., Gatell, J. M.*, et al.* (2008). Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. *N Engl J Med* **359**, 355-365.
Soon, G. G., Nie, L., Hammerstrom, T., Zeng, W., and Chu, H. (2011). Meeting the demand for more sophisticated study designs. A proposal for a new type of clinical trial: the hybrid design. *BMJ Open* **1**, e000156.
Steigbigel, R. T., Cooper, D. A., Kumar, P. N.*, et al.* (2008). Raltegravir with optimized background therapy for resistant HIV-1 infection. *N Engl J Med* **359**, 339-354.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

LEI NIE
03/22/2013

GUOXING SOON
03/22/2013

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 203093 **Applicant:** Gilead **Stamp Date:** 06/27/2012

Drug Name: Elvitegravir **NDA/BLA Type:** original, standard

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated.	X			
4	Data sets in EDR are accessible and conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? ____X____

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.	X			
Appropriate references for novel statistical methodology (if present) are included.	X			
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	X			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

1. Please explain how your analysis deals with the HIV viral load data which were measured by two (or actually three) assays. We note that, at some time points, HIV viral load was measured by only one of following three HIV RNA assays: 1) Amplicor/standard; 2) Amplicor/ultra sensitive; and 3) Taqman.
2. Please provide information to identify which subjects are originally from Study 144.
3. The randomization code does not indicate whether subjects were randomized using the randomization code for Study 0144 or 0145. Please provide the original randomization (Rand0144 and Rand0145) for Studies 0144 and 0145 and provide source information demonstrating how they were merged into the new Study 145.
4. Please clarify the reasons why the information form L:\m5\datasets\gs-us-183-0145\tabulations\sdtm\96-wk\ds.xpt does not match your Figure 8-1, in the study 145 report. For example, the data showed there are 725 subjects were randomized, but your figure says 724. The data showed the death is 11 but your figure showed 10. There are many other items where the data do not match the figures.
5. Please explain how you define the week 48 and week 96 window for the snapshot analysis. Our preliminary assessment of your snapshot classification reveals some inconsistency between your classifications and ours, which could be related to a different window definition. For example, we consider subject 2475-3203 to be a failure rather than "not missing data but on study." This subject stopped treatment on 2011-2-24, when the viral load was 33,700 on day 700.

Brief summary of controlled clinical trials

The following table contains information on the relevant trials contained in the submission.

Study number	Design	Treatment arms/Sample size	Primary endpoint/Analysis	Sponsor's findings
145	NI trial of EVG vs RAL	EVG (351) vs RAL (351)	Snapshot algorithm of response at week 48	EVG is noninferior to RAL

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

LEI NIE
09/06/2012

GUOXING SOON
09/06/2012