

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning

Xiangchen (Bob) Cui, Vertex Pharmaceuticals, Cambridge, MA

### ABSTRACT

The Center for Drug Evaluation and Research (CDER) issued draft guidance in December 2012, which urges sponsors to submit a clinical dataset that describes and summarizes the characteristics and outcomes of clinical investigation at the level of the individual study site (summary level clinical site data). The agency uses it to facilitate use of a risk-based approach for the timely identification of clinical investigator sites for on-site inspection by CDER during the review of marketing applications. CDER approved Vertex Pharmaceuticals Incorporated two NDAs (hepatitis C and cystic fibrosis) in 2011 and 2012, respectively. This paper provides two different examples of preparing summary level clinical site data for these two drug NDAs. The sharing of hands-on experiences in this paper is intended to assist readers to fully prepare summary level clinical site data for their NDA submissions.

### INTRODUCTION

FDA uses onsite inspections to ensure that clinical investigators, sponsors, and IRBs comply with FDA regulations while developing investigational drugs or biologics. In 1977, FDA established the Bioresearch Monitoring Program to develop cross-center guidelines for inspections of clinical investigators, sponsors, and IRBs. The main objectives of this program are to “protect the rights, safety and welfare of subjects involved in FDA-regulated clinical trials and to verify the accuracy and reliability of clinical trial data submitted to FDA in support of research or marketing applications. The Bioresearch Monitoring Program for drugs is managed by the Office of Scientific Investigations (**OSI**) and for biologics by the Office of Inspections and Surveillance. Medical reviewers, who are responsible for approving or disapproving a product, consult with Bioresearch Monitoring reviewers to choose which clinical trial sites to inspect. Bioresearch Monitoring reviewers analyze various factors, such as the clinical trial protocol risk, high subject enrollment at one site, and the clinical investigator's inspection history to determine which sites to inspect.

The current submission format for study-specific data in sponsors' NDA and BLA packages does not facilitate efficient site selection for FDA because these data are submitted as subject level data.

In December 2012, Center for Drug Evaluation and Research (**CDER**) issued a draft guidance relating to electronic submissions. **Guidance for Industry: Providing Submissions in Electronic Format—Summary Level Clinical Site Data for CDER's Inspection Planning** [1] is one in a series of guidance documents intended to assist sponsors and applicants making certain regulatory submissions to FDA in electronic format. The purpose of this guidance is to assist applicants in the submission of a clinical dataset that describes and summarizes the characteristics and outcomes of clinical investigations at the level of the individual study site (summary level clinical site data). This dataset is intended to facilitate use of a risk-based approach for the timely identification of clinical investigator sites for on-site inspection by **CDER** during the review of marketing applications [1].

**CDER** recommends that applicants submit a summary level clinical site dataset with all NDAs, BLAs, or supplements that contain clinical data, preferably with the NDA/BLA submission. Alternatively, a summary level clinical site dataset can be provided for a pre-NDA or pre-BLA meeting. This dataset should include data for each major (e.g., pivotal) clinical study submitted to support safety and efficacy. Summary level site data are not requested for biopharmaceutical, clinical pharmacology, or animal studies. Reviewers in the **OSI** are available to assist applicants with questions about submitting these summary level clinical data [1].

**FDA** also issued another draft guidance in November 2011, October 2012, and November 2012, respectively, **Specifications for Preparing and Submitting Summary Level Clinical Site Data for CDER's Inspection Planning** [2], which provides current FDA specifications for preparing and submitting a summary-level clinical-site dataset in electronic form for NDAs, BLAs, and NDA or BLA supplemental applications submitted to FDA's **CDER**.

Vertex Pharmaceuticals submitted two NDAs (hepatitis C and cystic fibrosis) in 2010 and 2011, respectively. The Office of Scientific Investigations (**OSI**) requested two types of data to be submitted to the NDA: 1) data to address the clinical data submitted in the NDA that will be used for inspection background materials

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

(Items I and II in Appendix A); and 2) data (**Summary Level Clinical Site Data**) to address the site selection process (Appendix B).

The item 1 in Appendix A specifies the request for general study related information and specific clinical investigator information, which was provided mainly by our Clinical Operations and item 2 is the request for site level data, mainly site-specific individual subject data ("line") listings from the datasets, prepared by Biometrics. Appendix B specifies Summary Level Clinical Site Data for data integrity review and inspection planning in NDA and BLA submissions, prepared by Biometrics.

We used Appendix A for the preparation for these two NDA submissions. There are minor differences between Appendix A and **Specifications for Preparing and Submitting Summary Level Clinical Site Data for CDER's Inspection Planning** [2] where the latter has added more variables. Among them are STUDY (Study Number), STUDYTL (Study Title), Domain (Domain Abbreviation), SPONNO (Sponsor Number), SPONNAME (Sponsor Name), UNDERIND (Under IND), NDA (NDA Number), BLA (BLA Number), SUPPNUM (Supplement Number), FINLMAX (Maximum Financial Disclosure Amount), and MINITIAL (Investigator Middle Initial). The information can be obtained from company Regulatory Affairs, and Clinical Operations. However the other "risk factors" of enrollment, site-specific efficacy, protocol violations, deaths, AEs, and SAEs remain intact which need to be derived by statistical programming from clinical database.

This paper will mainly illustrate the hands-on experience of preparing these derived variables in SLCS dataset. Hence Appendix A will be used in the following, instead of [2].

In fact, the draft guidance [2] has been revised two times since November 2011. Its evolution is underway. There are six collaborative working groups among FDA, industry and academia to improve standards, tools and processes across regulatory review and computational science for Pharmaceutical Industry, whose name is **FDA/PhUSE Computational Sciences (CSS) Working Group**. Among them, **Working Group 2 is working on Standardizing data within the Inspection Site Selection Process** [3]. The specification of summary level clinical site data for CDER's inspection planning is still under development.

This paper will provide hands-on experiences from two NDA submissions with different endpoint types (discrete endpoint and continuous endpoint). While examples are illustrated, some sample of SAS codes are provided as your reference.

### TWO NDA SUBMISSIONS FROM VERTEX PHARMACEUTICALS INCORPORATED

Vertex Pharmaceuticals Incorporated submitted two NDAs (hepatitis C and cystic fibrosis) to CDER in 2010 and 2011, respectively, shown in Table 1.

Indication	Primary Efficacy Endpoint
Hepatitis C Virus (HCV)	Proportion of subjects achieving sustained viral response (SVR): undetectable HCV RNA 24 weeks after last planned dose of study treatment
Cystic Fibrosis (CF)	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1)

**Table 1 Two NDA Submissions with Different Primary Efficacy Endpoints**

Our experience shows that the preparation of item 1 and 2 in Appendix A, and summary level clinical site data in Appendix B was the close collaboration among Clinical Operations, Regulatory Affairs, and Biometrics.

### INTRODUCTION TO SUMMARY LEVEL CLINICAL SITE DATASET (SLCS)

Exhibit 1 in Appendix B provides summary level clinical site data elements, and Exhibit 2 provides a sample data submission for the variables identified in Exhibit 1. The SLCS contains the following information:

1. IND Number
2. Trial Number Site ID
3. Treatment Arm
4. Enrollment (Number of Subjects Enrolled and Number of Subjects Screened)
5. Number of Subject Discontinuations
6. Endpoint
7. Endpoint Type

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

8. Site-specific Efficacy
9. Protocol Violations
10. Deaths
11. AEs
12. SAEs
13. Financial Disclosure
14. Name, Address and Contact information of the Primary Investigator

It is worthwhile to point out that the subjects who changed the sites during the studies were counted in the original sites in our preparations.

The screen failure subjects were not in our clinical database. Each study team from Clinical Operations provided the information in excel format, which was converted into a SAS dataset shown in Display 1 and combined with the number of subjects enrolled for the number of subjects screened at each site.

Country	Investigator	Subject	DOB	Gender	Screening DT	Screen Fail DT	Screen Fail Reason	Comm	siteid
United States	Nezam Aldhal	V10105	01MAR1985	Female	23JUL2008	13AUG2008	Cannot be randomized within allowed screening period	N/A	101
United States	Sanjeev Arora	V10402	01MAY1951	Male	18JUN2008	20JUN2008	Other lab value out of range and/or clinically significant lab abnormal	N/A	104
United States	Sanjeev Arora	V10409	01NOV1953	Male	20JUN2008	24JUN2008	Other lab value out of range and/or clinically significant lab abnormal	N/A	104
United States	Michael Bennett	V10502	01AUG1947	Female	12JUN2008	17JUN2008	Other lab value out of range and/or clinically significant lab abnormal	N/A	105
United States	Michael Bennett	V10504	01APR1958	Male	18JUN2008	23JUN2008	Other		105
United States	Michael Bennett	V10506	01MAR1958	Female	26JUN2008	30JUN2008	Not genotype 1	N/A	105
United States	Michael Bennett	V10509	01AUG1950	Male	22JUL2008	08AUG2008	Subject withdrew consent prior to randomization	N/A	105

**Display 1 Snapshot of Screen Failures**

Display 2 shows the example of IND Number, Trial Number Site ID, Treatment Arm, Number of Subjects Enrolled, Number of Subjects Screened, Number of Subject Discontinuations, Endpoint, and Endpoint Type.

IND	TRIAL	SITEID	ARM	ENROLL	SCREEN	DISCON
71832	VX07-950-108	101	T-8/PR - VX950, Placebo, Copegus, Pegasys	2	12	0
71832	VX07-950-108	101	T-12/PR - VX950, Copegus, Pegasys	4	12	0
71832	VX07-950-108	101	Pbo-12/PR - Placebo, Copegus, Pegasys	5	12	2
71832	VX07-950-108	104	T-8/PR - VX950, Placebo, Copegus, Pegasys	2	11	0
71832	VX07-950-108	104	T-12/PR - VX950, Copegus, Pegasys	1	11	0
71832	VX07-950-108	104	Pbo-12/PR - Placebo, Copegus, Pegasys	6	11	0

**Display 1 Snapshot of SLCS Dataset's Partial Variables**

The treatment arm was not clearly specified in CDER's specification [2] or Appendix B regarding which one: planned treatment or actual treatment. We used planned treatment arms in our two preparations of SLCS datasets.

The following SAS codes show the derivation of ENROLL, SCREEN, and DICSON.

```

**1. ENROLL: Number of subjects enrolled;
proc sort data=adsl out=_adsl;by siteid armcd arm;run;
proc freq data=_adsl noprint;
    tables siteid*armcd*arm/out=_enroll(drop=percent);
run;
proc sort data=_enroll(rename=(count=ENROLL));by siteid armcd arm;run;
*** the number of full analysis;
proc freq data=_adsl noprint;
    tables siteid*armcd*arm/out=_fas(drop=percent);
    where fasfl='Y';
run;
proc sort data=_fas(rename=(count=FASFL));by siteid armcd arm;run;

**2. SCREEN: Number of subjects screened;

```

Experiences in Preparing Summary Level Clinical Site Data within NDA’s Submission for FDA’s Inspection Planning, continued

```

data adsl_all;
  set adsl datain.screen;**combine enrolled subjects with screen
failures;
run;
proc freq data=adsl_all noprint;
  tables siteid/out=_screen(drop=percent rename=(count=SCREEN));
run;
proc sort data=_screen;by siteid;run;
data _enroll;
  merge _enroll _screen;
  by siteid;
run;
proc sort data=_enroll;by siteid armcd arm;run;
data _enroll;
  merge _enroll _fas;
  by siteid armcd arm;
run;

**3. DISCONT: Number of Subject Discontinuations;
proc freq data=_adsl noprint;
  tables siteid*armcd*arm/out=_discont(drop=percent);
  where CMPSTYFN=2;
run;
proc sort data=_discont(rename=(count=DISCONT));by siteid armcd
arm;run;

```

**SITE-SPECIFIC EFFICACY**

Item II in Appendix B explains four variables for the site-level efficacy results in SLCS dataset. OSI will use these variables to support site selection and are not intended to support evaluation of efficacy. For each study and investigator site, these variables are shown in Table 2. The primary efficacy endpoint in the study should be used in the calculation of these variables.

Variable	Label	Explanation
TRTEFFR	Treatment Efficacy Result	the efficacy result for each primary endpoint, by treatment arm
TRTEFFV	Treatment Efficacy Result Variance	the variance of the efficacy result (TRTEFFR) for each primary endpoint, by treatment arm
SITEEFFE	Site-specific Efficacy Effect Size	the effect size should be the same representation as reported for the primary efficacy analysis
SITEEFFV	Site-specific Efficacy Effect Size Variance	the variance of the site-specific efficacy effect size (SITEEFFE)

**Table 2 Treatment Efficacy by Treatment Arm and Site, and Site-specific Efficacy Effect Size by Site**

There are two variables to describe the primary endpoint and its type, shown in Table 3.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes or Description	Sample Value
ENDPOINT	Endpoint	Char	String	Plain text label used to describe the primary endpoint as described in the Define file included with each application. (limit 200 characters)	Average increase in blood pressure
ENDPTYPE	Endpoint Type	Char	String	Variable type of the primary endpoint (i.e., continuous, discrete, time to event, or other)	Continuous

**Table 3 Endpoint and Its Type Variables for Site-level Efficacy Variables**

## Experiences in Preparing Summary Level Clinical Site Data within NDA’s Submission for FDA’s Inspection Planning, continued

Table 4 shows our two NDA submissions with binary and continuous endpoints.

Indication	ENDPTYPE (Endpoint Type)	ENDPOINT (Primary Efficacy Endpoint)
Hepatitis C Virus (HCV)	Binary	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment
Cystic Fibrosis (CF)	Continuous	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1) through Week 24

**Table 4 Vertex Two NDA Submissions with Binary and Continuous Endpoints**

In our Hepatitis C Virus (HCV) NDA submission, primary efficacy endpoint shown in Table 4 was equal to proportion of subjects achieved svr24 planned, which was calculated by number of subjects achieved svr24 planned divided by total number of subjects randomized and treated.

For the simple explanation, we can define P1, P2, n1, and n2 as shown below.

P1=proportion of subjects achieved svr24 planned in active drug treatment arm at a given site,

P2=proportion of subjects achieved svr24planned in placebo arm at a given site,

n1=number of subjects randomized and treated in active drug treatment arm at a given site, and n2=number of subjects randomized and treated in placebo arm at a given site

Derivation Rules for Site-specific Efficacy Variables in SLCS are provided by Table 5.

Variable	Derivation Rule
TRTEFFR	Equal to P1 for active drug treatment arm and P2 for placebo arm
TRTEFFV	Equal to (TRTEFFR *(1- TRTEFFR))/number of subjects randomized and treated in a site per treatment arm
SITEEFFE	Equal to P1-P2 Note: This variable is populated only for active drug treatment arms in SLCS.
SITEEFFV	Equal to $p*(1-p)*(1/n1+1/n2)$ where $p=(P1*n1+P2*n2)/(n1+n2)$

**Table 5 Derivation Rules for Site-specific Efficacy Variables in Our Hepatitis C Virus (HCV) NDA submission**

The Display 2 shows the snapshot of site-specific efficacy result by two sites. Note that Arm with the value “Pbo-12/PR-Placebo, Copegus, Pegasys” was the placebo arm. Hence the values for SITEEFFE and SITEEFFV were “missing”. There were two active arms in this study, i.e., T-8/PR and T-12/PR. Each was compared with placebo group.

SITEID	ARM	ENDPOINT	ENDPTYPE	TRTEFFR	TRTEFFV	SITEEFFE	SITEEFFV
110	T-8/PR - VX950, Placebo, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	0.5	0.125	0	0.25
110	T-12/PR - VX950, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	0.67	0.0737	0.17	0.1997
110	Pbo-12/PR - Placebo, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	0.5	0.125		
111	T-8/PR - VX950, Placebo, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	1	0	0.5	0.1667
111	T-12/PR - VX950, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	0.33	0.0737	-0.17	0.1427
111	Pbo-12/PR - Placebo, Copegus, Pegasys	Proportion of subjects achieving sustained viral response (SVR):undetectable HCV RNA 24 weeks after last planned dose of study treatment	Binary	0.5	0.0625		

**Display 2 Snapshot of SLCS Dataset for Site-specific Efficacy Result by Two Sites**

The following SAS codes show the derivation of site-specific efficacy variables in submitting Summary Level Clinical Site Data to FDA for Hepatitis C Virus (HCV) NDA submission.

```

**4. TRTEFFR: Treatment Efficacy Result;
proc freq data=_adsl noprint;
  tables siteid*armcd*arm/out=_svr24pfn(drop=percent);
  where SVR24PFN=1; **Primary Efficacy Endpoint;

```

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

```
run;
proc sort data=_svr24pfn(rename=(count=SVR24P_CNT));by siteid armcd
arm;run;
data all1;
  merge _enroll _discont _svr24pfn ;
  by siteid armcd arm;
  if DISCONT=. then DISCONT=0;
  if SVR24P_CNT=. then SVR24P_CNT=0;
  if FASFL>0 then TRTEFFR=round(SVR24P_CNT/FASFL,0.01);
  ** Full Analysis Set Population Flag;
run;
**5. TRTEFFV: Treatment Efficacy Result Variance;
**The variance of a population proportion is defined by the following
formula:  $s = P(1-P) / n$ ;
data all1;
  set all1;
  if FASFL>0 then TRTEFFV=round(TRTEFFR*(1-TRTEFFR)/FASFL,0.0001);
run;
**6. SITEEFFE: Site-Specific Efficacy Effect Size;
proc sort data=all1;by siteid armcd arm;run;
proc sort data=all1 out=pbo_eff(keep=siteid TRTEFFR FASFL);
  by siteid armcd arm;
  where armcd='C';
run;
data all2;
  merge all1(in=a) pbo_eff(in=b rename=(TRTEFFR=TRTEFFR3
FASFL=FASFL3));
  by siteid;
  if armcd='C' then SITEEFFE=.;
  else SITEEFFE=TRTEFFR-TRTEFFR3;
  if a and not b then no_pbo_fg=1;
run;
**7. SITEEFFV: Site-Specific Efficacy Effect Size Variance;
*** The pooled  $p=(p1*n1+p2*n2)/(n1+n2)$ ;
*** The variance= $p*(1-p)*(1/n1+1/n2)$ ;
data all2;
  set all2;
  if armcd^='C' then do;
    p=(TRTEFFR*FASFL+TRTEFFR3*FASFL3)/(FASFL+FASFL3);
    SITEEFFV=round(p*(1-p)*(1/FASFL+1/FASFL3),0.0001);
  end;
run;
```

In our second NDA submission for Cystic Fibrosis (CF), primary efficacy endpoint shown in Table 1 was equal to Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1)

For the simple explanation, we can define M1 and M2 as shown below.

M1= the mean change of the subject-level mean absolute changes in %predicted FEV1 in active drug treatment arm at a given site

M2= the mean change of the subject-level mean absolute changes in %predicted FEV1 in placebo arm at a given site

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

Derivation Rules for Site-specific Efficacy Variables in SLCS are provided by Table 5.

Variable	Derivation Rule
TRTEFFR	Equal to M1 for active drug treatment arm and M2 for placebo arm
TRTEFFV	Equal to the squared standard error of TRTEFFR. Note: For sites having fewer than 2 subjects contributing to TRTEFFR, variance cannot be calculated and TRTEFFV will be set to missing.
SITEEFFE	Equal to M1-M2 Note: This variable is populated only for active drug treatment arms in SLCS.
SITEEFFV	The variance of the site-specific efficacy effect size, assuming unequal variance between treatment arms is calculated as: $S_u^2 = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$ , where Si and ni are standard deviations and sample sizes for each treatment arm. Note: This variable is populated for the active drug treatment arm when at least 2 subjects contribute to TRTEFFR in each treatment arm. Otherwise, SITEEFFEV is set to missing.

**Table 5 Derivation Rules for Site-specific Efficacy Variables in Our Cystic Fibrosis (CF) NDA submission**

The Display 3 shows the snapshot of site-specific efficacy result by two sites. Note that both TRTEFFV and SITEEFFV were set to missing for there was only one subject for VX770 arm at site 004.

SITEID	ARM	ENROLL	ENDPOINT	ENDPTYPE	TRTEFFR	TRTEFFV	SITEEFFE	SITEEFFV
004	VX-770	1	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1) through Week 24	Continuous	12		16.6	
004	PLACEBO	2	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1) through Week 24	Continuous	-4.6	0.04		
045	VX-770	2	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1) through Week 24	Continuous	-2.4	25.69	-2.6	30.11
045	PLACEBO	3	Absolute change from baseline in percent predicted forced expiratory volume in 1 second (%predicted FEV1) through Week 24	Continuous	0.3	4.42		

**Display 3 Snapshot of SLCS Dataset for site-specific efficacy result by two sites**

The following SAS codes show the derivation of site-specific efficacy variables for Cystic Fibrosis (CF) NDA submission.

```

**4. Treatment Efficacy;
proc sort data=adsp out=adspl;by usubjid avisitn;
    where paramcd='PPFEV' and dtype=' ' and anl24fl='Y' and
ABLFL^='Y';
run;
proc univariate data=adspl noprint;
    by usubjid;
    var chg;
    output out=effvar n=n mean=mean;
run;
proc sort data=effvar;by usubjid;run;
proc sort data=adsl;by usubjid;run;
data effvar1;
    merge effvar(in=a)
          adsl(keep=usubjid siteid armcd arm fasfl where=(fasfl='Y'));
    by usubjid;
    if a;

```

Experiences in Preparing Summary Level Clinical Site Data within NDA’s Submission for FDA’s Inspection Planning, continued

```

run;
proc sort data=effvar1;by siteid armcd arm;run;
proc univariate data=effvar1 noprint;
  by siteid armcd arm;
  var mean;
  output out=effvar2 n=n mean=mean STDMEAN=STDMEAN var=variance;
run;
proc sort data=effvar2;by siteid armcd arm;run;
proc transpose data=effvar2 out=_n(drop=_label_ _name_) prefix=n_;
  by siteid;
  var n;
  id armcd;
run;
proc transpose data=effvar2 out=_mean(drop=_label_ _name_)
prefix=mean_;
  by siteid;
  var mean;
  id armcd;
run;
proc transpose data=effvar2 out=_variance(drop=_label_ _name_)
prefix=var_;
  by siteid;
  var variance;
  id armcd;
run;
data effvar3;
merge _n _mean _variance;
by siteid;
SITEEFFE=mean_A-mean_B; *** Site-Specific Efficacy Effect Size;
if n_A>1 and n_B>1 then SITEEFFV=var_A/n_A + var_B/n_B;
***Site-Specific Efficacy Effect Size
Variance;
run;
proc sort data=effvar3;by siteid;run;
data effvar4;
merge effvar2 effvar3;
by siteid;
TRTEFFR=mean; ***Treatment Efficacy Result;
TRTEFFV=STDMEAN*STDMEAN; ***Treatment Efficacy Result Variance;
run;
proc sort data=effvar4 out=_effvar4(keep=siteid armcd arm n TRTEFFR
TRTEFFV SITEEFFE SITEEFFV);by siteid armcd arm;
run;

```

**ENDPOINT TYPE**

To calculate the site-specific efficacy result variable by treatment arm, “TRTEFFR” in Table 2, SLCS requires endpoint type information, in addition to endpoint. Item II in Appendix B explains four endpoint type shown by Table 6.

Endpoint Type	Explanation
Discrete Endpoints	endpoints consisting of efficacy observations that can take on a discrete number of values (e.g., binary, categorical). Summarize discrete endpoints by an event frequency (i.e., number of events), proportion of events, or similar method at the site for the given treatment.
Continuous Endpoints	endpoints consisting of efficacy observations that can take on an infinite number of values. Summarize continuous endpoints by the mean of the observations at the site

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

	for the given treatment.
Time-to-Event Endpoints	endpoints where the time to occurrence of an event is the primary efficacy measurement. Summarize time-to-event endpoints by two data elements: the number of events that occurred (TRTEFFR) and the number of censored observations (CENSOR).
Other	if the primary efficacy endpoint cannot be summarized in terms of the previous guidelines, a single or multiple values with precisely defined variable interpretations should be submitted as part of the dataset

**Table 6 Four Endpoint Types**

Display 2 and Display 3 show the examples of endpoint types with binary and continuous endpoints, respectively.

### SAFETY PERSPECTIVE VARIABLES IN SUMMARY LEVEL CLINICAL SITE DATA

There are three safety perspective variables in SLCS dataset, shown in Table 7.

Variable	Variable Label	Comments
NSAE	Number of Non-Serious Adverse Events	Total number of treatment emergent non-serious adverse events at a given site per treatment arm
SAE	Number of Serious Adverse Events	Total number of treatment emergent serious adverse events <b>excluding deaths</b> at a given site per treatment arm
DEATH	Number of Deaths	Total number of deaths at a given site per treatment arm

**Table 7 Definition of Safety Variables in SLCS dataset**

Note: NSAE and SAE should include multiple events per subject, which is contrary to AE table reporting programming, and these events are regardless the relationship with treatments. SAE excludes the deaths.

Snapshot of these variables in an example in HCV NDA submission shown in Display 3.

SITEID	ARM	ENROLL	NSAE	SAE	DEATH
135	T-8/PR - VX950, Placebo, Copegus, Pegasys	5	62		
135	T-12/PR - VX950, Copegus, Pegasys	4	54		
135	Pbo-12/PR - Placebo, Copegus, Pegasys	6	69	3	
139	T-8/PR - VX950, Placebo, Copegus, Pegasys	2	22		
139	T-12/PR - VX950, Copegus, Pegasys	2	32		
139	Pbo-12/PR - Placebo, Copegus, Pegasys	2	32		
140	T-8/PR - VX950, Placebo, Copegus, Pegasys	3	33		
140	T-12/PR - VX950, Copegus, Pegasys	3	30		1
140	Pbo-12/PR - Placebo, Copegus, Pegasys	3	30		

**Display 3 Snapshot of SLCS Dataset for Safety Perspective Variables**

The following SAS codes show the derivation of NSAE, SAE, and DEATH variables for Hepatitis C Virus (HCV) NDA submission.

```

**8. NSAE: Number of Non-Serious Adverse Events;
**9. SAE: Number of Serious Adverse Events;
proc sort data=adae out=_adae;
  by siteid armcd arm aeser;
  where aeany='Y' and trtemfl='Y' and aedthfl^='Y';
run;
proc freq data=_adae noprint;
  tables siteid*armcd*arm*aeser/out=_aecnt(drop=percent);
run;
proc sort data=_aecnt;by siteid armcd arm aeser;run;
proc transpose data=_aecnt out=_aecnt2(drop=_name_ _label_);

```

Experiences in Preparing Summary Level Clinical Site Data within NDA’s Submission for FDA’s Inspection Planning, continued

```

by siteid armcd arm;
id aeser;
var count;
run;
proc sort data=_aecnt2;by siteid armcd arm;run;
proc sort data=all2;by siteid armcd arm;run;
data all3;
merge all2 _aecnt2(rename=(N=NSAE Y=SAE));
by siteid armcd arm;
run;
**10. DEATH: Number of Deaths;
proc sort data = _adsl(where = (STREAS = "Death")) out = _adsld;
by siteid armcd arm usubjid;
run;
proc freq data=_adsld noprint;
tables siteid*armcd*arm/out=_death(drop=percent);
run;
proc sort data=_death;by siteid armcd arm;run;
proc sort data=all3;by siteid armcd arm;run;
data safety;
merge all3 _death(rename=(count=DEATH));
by siteid armcd arm;
run;

```

**PROTOCOL VIOLATIONS**

The specification for number of protocol violations is described in Table 8. It is unclear from FDA specification [2] if the violations exclusively include programmatically derived ones per study Statistical Analysis Plan (SAP), or also include protocol violations from study logs from each study.

Variable	Variable Label	Comments
PROTVIOL	Number of Protocol Violations	Number of deviations from the protocol noted by the sponsor for a given site. This value should include multiple violations per subject. Any duplicate violation for individual subjects is not excluded from the count. For sites with no subjects enrolled, PROTVIOL is set to missing.

**Table 8 Definition of the Variable for Number of Protocol Violations in SLCS dataset**

The Display 4 shows the snapshot of number of protocol violations in SLCS from HCV study.

The ADaM dataset ADPP was derived only from the study SAP.

SITEID	ARM	ENROLL	PROTVIOL
108	T-8/PR - VX950, Placebo, Copegus, Pegasys	4	0
108	T-12/PR - VX950, Copegus, Pegasys	3	1
108	Pbo-12/PR - Placebo, Copegus, Pegasys	2	0
109	T-8/PR - VX950, Placebo, Copegus, Pegasys	4	1
109	T-12/PR - VX950, Copegus, Pegasys	1	1
109	Pbo-12/PR - Placebo, Copegus, Pegasys	4	0
110	T-8/PR - VX950, Placebo, Copegus, Pegasys	2	0
110	T-12/PR - VX950, Copegus, Pegasys	3	0
110	Pbo-12/PR - Placebo, Copegus, Pegasys	2	0

**Display 4 Snapshot of SLCS Dataset for Number of Protocol Violations Variable**

The following SAS codes show the derivation of PROTVIOL variable for Hepatitis C Virus (HCV) NDA submission.

```

**11. PROTVIOL Number of Protocol Violations;
proc sort data=adpp;by siteid armcd arm;run;

```

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

```
proc freq data=adpp noprint;
  tables siteid*armcd*arm/out=_adpp(drop=percent);
run;
proc sort data=_adpp;by siteid armcd arm;run;
proc sort data=all4;by siteid armcd arm;run;
data safety2;
  merge safety _adpp(rename=(count=PROTVIOL));
  by siteid armcd arm;
  if NSAE=. then NSAE=0;
  if SAE=. then SAE=0;
  if DEATH=. then DEATH=0;
  if PROTVIOL=. then PROTVIOL=0;
run;
```

However the variable PROTVIOL in SLCS from our second CF study submission was derived from SDTM DV domain, which included protocol violations from study logs provided by the clinical study team.

The Display 5 shows the snapshot of number of protocol violations in SLCS from CF study. The numbers of protocol violations in Display 5 were bigger than ones in Display 4, for protocol violations from logs were included in that calculation.

SITEID	ARM	ENROLL	PROTVIOL
002	VX-770	1	5
002	PLACEBO	3	16
007	VX-770	1	8
007	PLACEBO	1	20
015	VX-770	1	3
015	PLACEBO	2	10

### Display 5 Snapshot of SLCS Dataset for Number of Protocol Violations Variable

The following SAS codes show the derivation of PROTVIOL variable for Cystic Fibrosis (CF) NDA submission.

```
**11. PROTVIOL Number of Protocol Violations;
proc sort data=sdtm.dv out=dv;by usubjid;run;
data dv;
  merge dv(in=a) adsl(keep=usubjid siteid armcd arm);
  by usubjid;
  if a;
run;
proc freq data=dv noprint;
  tables siteid*armcd*arm/out=_dv(drop=percent
rename=(count=PROTVIOL));
run;
proc sort data=_dv;by siteid armcd arm;run;
```

## NAME, ADDRESS AND CONTACT INFORMATION OF THE PRIMARY INVESTIGATOR

Clinical Operations provided the information in excel format, which was converted into a SAS dataset, and was included in SLCS dataset.

## CONCLUSION

This paper provides hands-on experiences from preparing summary level clinical site datasets from two NDA submissions. Some sample of SAS codes are provided as your reference. We hope that they can make your life a little easier when you are working on preparing summary level clinical site datasets for NDA submission.

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

### REFERENCES

- [1] Guidance for Industry: Providing Submissions in Electronic Format — Summary Level Clinical Site Data for CDER's Inspection Planning  
<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/UCM332468.pdf>
- [2] Specifications for Preparing and Submitting Summary Level Clinical Site Data for CDER's Inspection Planning  
<http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ucm332466.pdf>
- [3] FDA/PhUSE Computational Sciences (CSS) Working Group 2: Standardizing data within the Inspection Site Selection Process  
[http://www.phusewiki.org/wiki/index.php?title=Standardizing\\_data\\_within\\_the\\_Inspection\\_Site\\_Selection\\_Process](http://www.phusewiki.org/wiki/index.php?title=Standardizing_data_within_the_Inspection_Site_Selection_Process)

### ACKNOWLEDGMENTS

Appreciation goes to Qunming Dong, Leif Bengtsson, and Tathabbai Pakalapati for their valuable review and comments.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Xiangchen (Bob) Cui, Ph.D.  
Enterprise: Vertex Pharmaceuticals, Inc.  
Address: 88 Sidney Street  
City, State ZIP: Cambridge MA, 02139  
Work Phone: 617-341-6069  
Fax: 617-460-8060  
E-mail: [xiangchen\\_cui@vrtx.com](mailto:xiangchen_cui@vrtx.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

## **Appendix A:**

### **I. Request for general study related information and specific Clinical Investigator information**

A. Please include the following information in a tabular format in the original NDA for each of the completed Phase 3 clinical trials:

1. Site number
2. Principal investigator
3. Location: City State, Country, to include contact information (phone, fax, email)

B. Please include the following information in a tabular format by site in the original NDA for each of the completed Phase 3 clinical trials:

1. Number of subjects screened for each site by site
2. Number of subjects randomized for each site by site
3. Number of subjects treated who prematurely discontinued for each site by site

C. Please include the following information in a tabular format in the NDA for each of the completed Phase 3 clinical trials:

1. Name, address and contact information of all CROs used in the conduct of the clinical trials
2. The location (actual physical site where documents are maintained and would be available for inspection) for all source data generated by the CROs with respect to their roles and responsibilities in conduct of respective studies
3. The location (actual physical site where documents are maintained and would be available for inspection) of sponsor/monitor files (e.g. monitoring master files, drug accountability files, SAE files, etc.)

### **II. Request for Site Level Data**

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

1. For each site in the pivotal clinical trials: Name of primary investigator, accurate address and phone number, e-mail contact
2. For each pivotal trial: Sample blank CRF and case report data tabulations for the site with coding key
3. For each pivotal trial: Site-specific individual subject data ("line") listings from the datasets:
  - a. Line listings for each site listing the subject/number screened and reason for subjects who did not meet eligibility requirements
  - b. Line listings by site and subject, of treatment assignment (randomization)
  - c. Line listings by site and subject, of drop-outs and discontinued subjects with date and reason
  - d. Line listings by site of evaluable subjects/ non-evaluable subjects and reason not evaluable
  - e. Line listings by site and subject, of AEs, SAEs, deaths and dates
  - f. Line listings by site and subject, of protocol violations and/or deviations reported in the NDA, description of the deviation/violation
  - g. Line listings by site and subject, of the primary and secondary endpoint efficacy parameters or events.
  - h. Line listings by site and by subject, concomitant medications (as appropriate to the pivotal clinical trials)
  - i. Line listings by site and by subject, of laboratory tests performed for safety monitoring

### **III. Request for Individual Patient Data Listings format:**

DSI is piloting a risk based model for site selection. Electronic submission of site level datasets will facilitate the timely selection of appropriate clinical sites for FDA inspection as part of the application and/or supplement review process. Please refer to the attached document, "Summary Level Clinical Site Data for Data Integrity Review and Inspection Planning in NDA and BLA Submissions" for further information. We request that you provide datasets, as outlined, for each pivotal study submitted in your application.

### **Appendix B:**

#### **Summary Level Clinical Site Data for Data Integrity Review and Inspection Planning in NDA and BLA Submissions**

##### **I. INTRODUCTION**

The purpose of this electronic submission of a single new clinical site dataset is to facilitate the timely evaluation of data integrity and selection of appropriate clinical sites for FDA inspection as part of the application and/or supplement review process.

## **II. DESCRIPTION OF THE SUMMARY LEVEL CLINICAL SITE DATASET**

The summary level clinical site data are intended (1) to clearly identify individual clinical investigator sites within an application or supplement, (2) to specifically reference the studies to which those clinical sites are associated, and (3) to present the characteristics and outcomes of the study at the site level.

For each study used to support efficacy, data should be submitted by clinical site and treatment arm for the population used in the primary analysis to support efficacy. As a result, a single clinical site may contain multiple records depending on the number of studies and treatment arms supported by that clinical site.

The site-level efficacy results will be used to support site selection and are not intended to support evaluation of efficacy. To this end, for each study used to support efficacy, the summary level clinical site dataset submission should include site-specific efficacy results by treatment arm and the submission of site-specific effect sizes.

The following paragraphs provide additional details on the format and structure of the efficacy related data elements.

### **Site-Specific Efficacy Results**

For each study and investigator site, the variables associated with efficacy and their variable names are:

- Treatment Efficacy Result (TRTEFFR) – the efficacy result for each primary endpoint, by treatment arm (see below for a description of endpoint types and a discussion on how to report this result)
- Treatment Efficacy Result Variance (TRTEFFV) – the variance of the efficacy result (treatEffR) for each primary endpoint, by treatment arm
- Site-specific Efficacy Effect Size (SITEEFFE) – the effect size should be the same representation as reported for the primary efficacy analysis
- Site-specific Efficacy Effect Size Variance (SITEEFFV) – the variance of the site-specific efficacy effect size (SITEEFFE)
- Endpoint (endpoint) – a plain text label that describes the primary endpoint as described in the Define file data dictionary included with each application.
- Treatment Arm (ARM) – a plain text label for the treatment arm that is used in the Clinical Study Report

In addition, for studies whose primary endpoint is a time-to-event endpoint, include the following data element:

- Censored Observations (CENSOR) – the number of censored observations for the given site and treatment.

## Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

If a study does not contain a time-to-event endpoint, record this data element as a missing value.

To accommodate the variety of endpoint types that can be used in analyses please reference the below endpoint type definitions when tabulating the site-specific efficacy result variable by treatment arm, "TRTEFFR".

- Discrete Endpoints – endpoints consisting of efficacy observations that can take on a discrete number of values (e.g., binary, categorical). Summarize discrete endpoints by an event frequency (i.e., number of events), proportion of events, or similar method at the site for the given treatment.
- Continuous Endpoints – endpoints consisting of efficacy observations that can take on an infinite number of values. Summarize continuous endpoints by the mean of the observations at the site for the given treatment.
- Time-to-Event Endpoints – endpoints where the time to occurrence of an event is the primary efficacy measurement. Summarize time-to-event endpoints by two data elements: the number of events that occurred (TRTEFFR) and the number of censored observations (CENSOR).
- Other – if the primary efficacy endpoint cannot be summarized in terms of the previous guidelines, a single or multiple values with precisely defined variable interpretations should be submitted as part of the dataset.

In all cases, the endpoint description provided in the "endpoint" plain text label should be expressed clearly to interpret the value provided in the (TRTEFFR) variable.

The site efficacy effect size (SITEEFFE) should be summarized in terms of the primary efficacy analysis (e.g., difference of means, odds ratio) and should be defined identically for all records in the dataset regardless of treatment.

The Define file for the dataset is presented in Exhibit 1.

### III. CREATING AND SUBMITTING THE DATA FILE (SUBMISSION TEMPLATE AND STRUCTURE)

A sample data submission for the variables identified in Exhibit 1 is provided in Exhibit 2. The summary level clinical site data can be submitted in SAS transport file format (\*.xpt). The file may be submitted electronically through the FDA Electronic Submission Gateway (ESG) referencing the active IND number or via secure CD addressed to the Division of Scientific Investigations point of contact.

**Exhibit 1: Summary Level Clinical Site Data Elements**

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes or Description	Sample Value
IND	IND Number	Num/Char	6 digit identifier	FDA identification number for investigational new drug	010010
TRIAL	Trial Number	Char	String	Study or Trial identification number	ABC-123
SITEID	Site ID	Num/Char	String	Investigator site identification number	50
ARM	Treatment Arm	Num/Char	String	Plain text label for the treatment arm as referenced in the clinical study report (limit 200 characters)	Active (e.g. 25mg), Comparator drug product name (e.g. Drug x), or Placebo
ENROLL	Number of Subjects Enrolled	Num	Integer	Total number of subjects enrolled at a given site	20
SCREEN	Number of Subjects Screened	Num	Integer	Total number of subjects screened at a given site	100
DISCONT	Number of Subject Discontinuations	Num	Integer	Number of subjects discontinuing from the study after being enrolled at a site	5
ENDPOINT	Endpoint	Char	String	Plain text label used to describe the primary endpoint as described in the Define file included with each application. (limit 200 characters)	Average increase in blood pressure
ENDPTYPE	Endpoint Type	Char	String	Variable type of the primary endpoint (i.e., continuous, discrete, time to event, or other)	Continuous
TRTEFFR	Treatment Efficacy Result	Num	Floating Point	The efficacy result for each primary endpoint, by treatment arm	0, 0.25, 1, 100
TRTEFFV	Treatment Efficacy Result Variance	Num	Floating Point	The variance of the efficacy result (TRTEFFR) for each primary endpoint, by treatment arm	0, 0.25, 1, 100
SITEEFFE	Site-Specific Efficacy Effect Size	Num	Floating Point	The effect size should be the same representation as reported for the primary efficacy analysis	0, 0.25, 1, 100
SITEEFFV	Site-Specific Efficacy Effect Size Variance	Num	Floating Point	The variance of the site-specific efficacy effect size (SITEEFFE)	0.065

Experiences in Preparing Summary Level Clinical Site Data within NDA’s Submission for FDA’s Inspection Planning, continued

CENSOR	Censored Observations	Num	Integer	The number of censored observations for the given site and treatment	5
NSAE	Number of Non-Serious Adverse Events	Num	Integer	Total number of non-serious adverse events at a given site. This value should include multiple events per subject.	10
SAE	Number of Serious Adverse Events	Num	Integer	Total number of serious adverse events excluding deaths at a given site. This value should include multiple events per subject.	5
DEATH	Number of Deaths	Num	Integer	Total number of deaths at a given site	1
PROTVIOL	Number of Protocol Violations	Num	Integer	Number of deviations from the protocol noted by the sponsor for a given site. This value should include multiple violations per subject.	20
FINLDISC	Financial Disclosure Amount	Num	Integer	Total financial disclosure amount (\$USD) by the site investigator	50000.00
LASTNAME	Investigator Last Name	Char	String	Last name of the investigator as it appears on the FDA 1572	Doe
FRSTNAME	Investigator First Name	Char	String	First name of the investigator as it appears on the FDA 1572	John
PHONE	Investigator Phone Number	Char	String	Phone number of the primary investigator	555-555-5555, 44-555-555-5555
FAX	Investigator Fax Number	Char	String	Fax number of the primary investigator	555-555-5555, 44-555-555-5555
EMAIL	Investigator Email Address	Char	String	Email address of the primary investigator	john.doe@mail.com
COUNTRY	Country	Char	ISO 3166-1-alpha-2	Country in which the site is located	US
STATE	State	Char	String	Unabbreviated state or province in which the site is located	Maryland
CITY	City	Char	String	Unabbreviated city, county, or village in which the site is located	Silver Spring
POSTAL	Postal Code	Char	String	Postal code for the site	20850
STREET	Street Address	Char	String	Street address and office number at which the site is located	1 Main St, Suite 100

The following is a fictional example of a data set for a placebo-controlled trial. Four international sites enrolled a total of 205 subjects who were randomized in a 1:1 ratio to active or placebo. The primary endpoint was the percent of responders. The site-specific efficacy effect size (SITEEFFE) is the difference between the active and the placebo treatment efficacy result. Note that since there were two treatment arms, each site contains 2 rows in the following example data set and a total of 8 rows for the entire data set.

**Exhibit 2: General Structure of Data Submission Template**

IND	TRIAL	SITEID	ARM	ENROLL	SCREEN	DISCONT	ENDPOINT	ENDTYPE	TRTEFFR
000001	Study 1	001	Active	26	61	3	Percent Responders	Binary	0.48
000001	Study 1	001	Placebo	25	61	4	Percent Responders	Binary	0.14
000001	Study 1	002	Active	23	54	2	Percent Responders	Binary	0.48
000001	Study 1	002	Placebo	25	54	4	Percent Responders	Binary	0.14
000001	Study 1	003	Active	27	62	3	Percent Responders	Binary	0.54
000001	Study 1	003	Placebo	26	62	5	Percent Responders	Binary	0.19

Experiences in Preparing Summary Level Clinical Site Data within NDA's Submission for FDA's Inspection Planning, continued

000001	Study 1	004	Active	26	29	2	Percent Responders	Binary	0.46
000001	Study 1	004	Placebo	27	29	1	Percent Responders	Binary	0.12

TRTEFF V	SITEEFF E	SITEEFF V	CENSO R	NSA E	SA E	DEATH	PROTVI OL	FINLDIS C	LASTNAME	FRSTNAME	PHONE
0.0096	0.34	0.0198	NA	0	2	0	1	0.00	Doe	John	555-123-4567
0.0049	NA	NA	NA	2	2	0	1	0.00	Doe	John	555-123-4567
0.0108	0.33	0.0204	NA	3	2	1	0	45000.00	Washington	George	020-3456-7891
0.0049	NA	NA	NA	0	2	0	3	45000.00	Washington	George	020-3456-7891
0.0092	0.35	0.0210	NA	2	2	0	1	0.00	Jefferson	Thomas	01-89-12-34-56
0.0059	NA	NA	NA	3	6	0	0	0.00	Jefferson	Thomas	01-89-12-34-56
0.0095	0.34	0.0161	NA	4	1	0	0	0.00	Lincoln	Abraham	555-987-6543
0.0038	NA	NA	NA	1	2	0	1	0.00	Lincoln	Abraham	555-987-6543

FAX	EMAIL	COUNTRY	STATE	CITY	POSTAL	STREET
555-123-4560	John@mail.com	RU	Moscow	Moscow	103009	Kremlin Road 1
555-123-4560	John@mail.com	RU	Moscow	Moscow	103009	Kremlin Road 1
020-3456-7890	george@mail.com	GB	Westminster	London	SW1A 2	10 Downing St
020-3456-7890	george@mail.com	GB	Westminster	London	SW1A 2	10 Downing St
01-89-12-34-51	tom@mail.com	FR	N/A	Paris	75002	1, Rue Road
01-89-12-34-51	tom@mail.com	FR	N/A	Paris	75002	1, Rue Road
555-987-6540	abe@mail.com	US	Maryland	Rockville	20852	1 Rockville Pk.
555-987-6540	abe@mail.com	US	Maryland	Rockville	20852	1 Rockville Pk.