Successful FDA Advisory Meetings by using Analysis Datasets

Sandy Chang, Gilead Sciences, Inc., Foster City, California
Steve Wong, Gilead Sciences, Inc., Foster City, California

ABSTRACT

Analysis datasets have become a crucial contributing factor to successful filing of New Drug Applications (NDAs) for Gilead Sciences, Inc. (GSI). In the span of a mere 23 months GSI has received three NDA approvals and we feel that a significant factor of that success can be directly attributed to the use of analysis datasets. The analysis datasets are divided into different categories, each of which serves a particular analysis purpose. The analysis datasets contain both variables that have been defined in the statistical analysis plan (SAP) and key contributed variables (derived variables that support planned and ad hoc analyses). In addition, the analysis datasets are designed for "one PROC away" use as suggested by the Analysis Dataset Modeling Team (ADaM) working group of the Clinical Data Interchange Consortium (CDISC). Finally, these datasets are standardized and consistent across all studies. As such, they save programming time, decrease the learning curve, increase productivity and allow us more resource allocation flexibility. Using our analysis datasets, GSI statistical programmers can generate outputs for ad-hoc requests in one working day or less. During the advisory committee meetings for our recent NDAs, the pre-defined analysis datasets were used to generate real-time responses to questions coming through from Food and Drug Administration (FDA) advisors.

INTRODUCTION

FDA advisory committees provide the FDA with independent opinions and recommendations on the development, safety and effectiveness of NDA applications and on FDA policies. Each committee consists of individuals with recognized expertise and judgment in a specific field. Members have the training and experience necessary to evaluate information objectively and to interpret its significance.

The NDAs include data to show the safety and effectiveness of human drugs. The advisory committee receives summary information about the applications and copies of FDA's review of the NDA documents. Based on this information, the advisory committee may recommend approval or rejection of the NDA. They may also ask for more information before they can make a recommendation. The FDA generally follows an advisory committee's recommendation, but is not legally bound to do so. The committees are advisory – they provide their expertise and recommendations – but FDA makes the final decisions.

Not every NDA submission will have an advisory committee meeting. But, for those NDAs with the advisory committee meetings, the date for the meeting can be the most important date for the sponsor companies, especially for smaller, startup companies. During the advisory committee meetings, members on the advisory committee will express their opinions as well as ask questions of the sponsors. The responses by the sponsors to these questions can potentially make a difference between an approval and a rejection.

Companies typically spend months ahead of the advisory committee meetings to prepare for their primary presentation as well as backup slides in anticipation of the questions. At GSI, we spend at least 6 months ahead of the advisory committee meetings preparing the materials. Even for those submissions without advisory committee meetings, GSI will go through the same preparations to answer FDA questions from the submissions as well as for preparing the Package Inserts (is an extension of professional labeling that may be distributed to patients when drug dispensed. It contains important information about the drug in lay language and may describe benefits, risk, how to recognize risks, dosage and administration). GSI treats all of the FDA's questions with highest priority and responds back to the FDA in the shortest time possible.

In order to provide quick turnaround time and ensure accuracy, GSI relies on the pre-defined analysis datasets. This paper provides an overview of how GSI structures its analysis datasets, and how those analysis datasets and analysis dataset metadata are useful to achieve our goal, which is gaining regulatory approval and successful marketing of new therapies through a reproducible, transparent, efficient and validated approach.

WHAT IS AN ANALYSIS DATASET?

Analysis datasets are the data used for statistical analysis and reporting by the sponsor. The overall principle in designing Statistical Analysis Datasets and related metadata is clear and unambiguous communication of the content, source and quality of the datasets submitted in support of the statistical analysis performed by the sponsor. (ADaM Statistical Analysis Dataset Model: General Considerations, 2004) An analysis dataset serves as a central depository of raw data and analyzable variables derived from one or more of the raw datasets. The derived variables are used as inputs to produce statistical summaries such as tables, figures, and listings supporting a statistical analysis. The Statistical Analysis Plan (SAP) outlines the essential safety and efficacy analysis, as well as defining the primary and secondary endpoints that serve as a blueprint to generate analysis datasets.

In addition, the analysis datasets set up at GSI also include variables that are not required in the SAP, but are study related and useful for analysis. The main purpose of deriving these extra variables is to ensure quick turnaround for ad-hoc requests not specified in the SAP.

GSI standardizes analysis datasets for use with a one PROC away approach, as suggested by the ADaM team, which we will discuss in detail later in this paper. We use the analysis datasets along with standard SAS® analysis macro(s), developed by our statistical programming group, to produce most of our statistical summaries.

The following flow chart (Figure 1) describes the process:

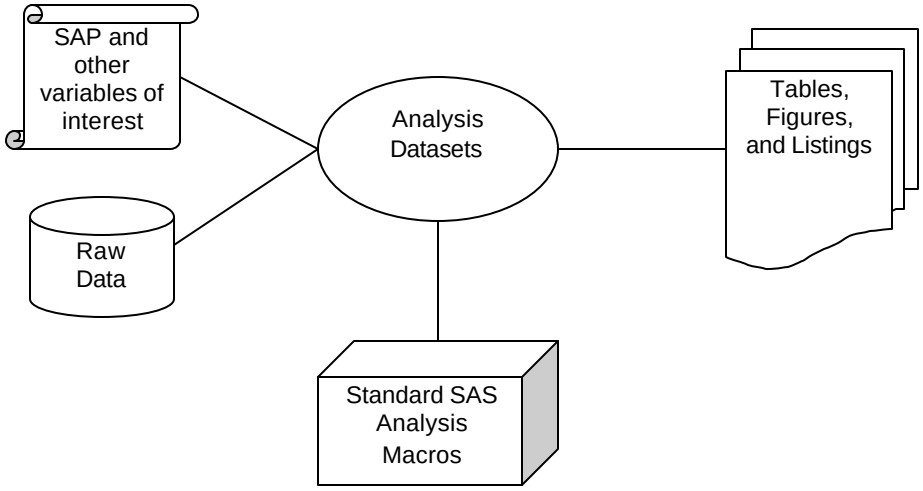


Figure 1: Analysis datasets generation process

TYPES OF ANALYSIS DATASETS

Our analysis datasets are categorized into demographic, laboratory, efficacy, and safety as shown in Figure 2. Each of the analysis datasets consists of derived variables as well as variables obtained directly from the raw data. Most importantly, we structure our analysis datasets to be consistent with ADaM's suggested one PROC away approach.

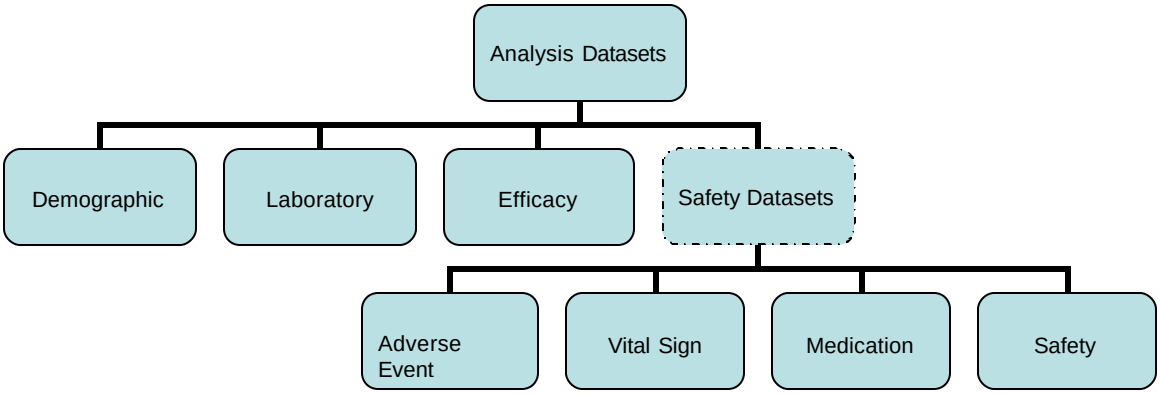


Figure 2: Analysis Dataset Categories

DEMOGRAPHIC analysis dataset contains all subjects' demographic data (i.e., Age, Race, and Gender), disposition data (i.e., Date patient withdrew from the study), treatment groups and key dates such as date of first dose, date of last collected Case Report Form (CRF) and duration on treatment. The dataset has the format of one observation per subject.

LABORATORY analysis dataset contains all subjects' laboratory data, in the format of one observation per subject per test code per visit per accession number. Here, we derive the study visits according to the study window defined in the SAP, as well as re-grade the laboratory toxicity per protocol. For a crossover study, both the visit related to the initial period and as it is related to the beginning of the new study period will be derived. If the laboratory data are collected from multiple local lab centers, this analysis dataset will also centralize the laboratory data and standardize measurement units by using conversion factors.

EFFICACY analysis dataset contains derived primary and secondary endpoint variables as defined in the SAP, as well as other efficacy parameters of interest, such as censor variables pertaining to the time to an efficacy event. This dataset has the format of one record per subject per analysis period.

SAFETY can be categorized into four analysis datasets:
VITAL SIGN analysis dataset captures all subjects' vital signs collected during the trial. This dataset has the format of one observation per subject per vital sign per visit, similar to the structure for the laboratory analysis dataset.

ADVERSE EVENT analysis dataset contains all adverse events (AEs) reported including serious adverse events (SAEs) for all subjects. A treatment emergent flag, as well as a flag to indicate if an event is reported within 30 days after the subject permanently discontinued from the study, will be calculated. This dataset has a format of one record per subject per adverse event per start date. Partial dates and missing AEs start and/or stop dates will be imputed using logic defined in the SAP.

MEDICATION analysis dataset contains the subjects' medication records including concomitant medications and other medications taken either prior to the beginning of study or during the study. This dataset has a format of one record per subject per medication taken per start date. Incomplete and missing medication start or stop dates will be imputed using instructions defined in the SAP.

SAFETY analysis dataset contains other safety variables, whether they are defined in the SAP or not. The Safety analysis dataset, similar to Efficacy analysis dataset in structure, consists of data with one record per subject per analysis period to capture safety parameters for all subjects.

It is crucial to generate analysis datasets in a specific order, as some variables derived from one particular analysis dataset may be used as the inputs to generate other variables in other analysis datasets. For example, the time to event variables in the efficacy and safety analysis datasets are calculated based on the date of the first dose derived in the demographic analysis dataset.

Figure 3 illustrates the sequence of analysis dataset definition:

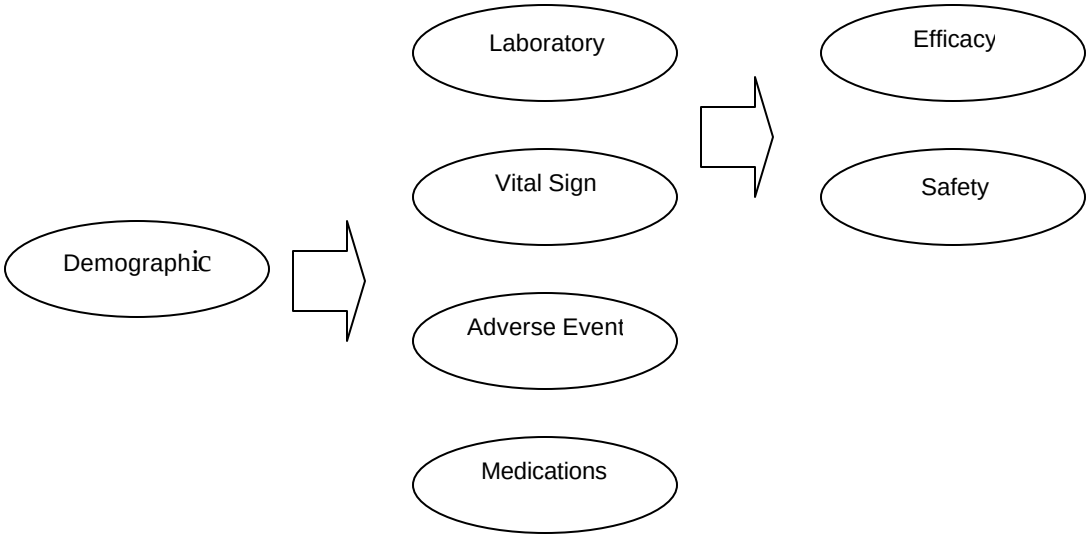


Figure 3: Analysis datasets are generated in sequence

WHY ANALYSIS DATASETS?

Analysis datasets not only contain information needed for performing statistical analysis, but also store imputed and derived variables that help FDA reviewers better understand the GSI data. A particular clinical trial may collect a large amount of individual subject data, but only a portion of the collected data might be needed or useful to determine the efficacy or safety of the drug. Analysis datasets thus help to filter unnecessary information, identify data errors that can not be identified by Data Quality Checks, clean data discrepancies, and prepare transparent and analyzable data.

In particular, analysis datasets from GSI are structured in a one PROC away format so that once the analysis datasets are set up, they can be used in conjunction with our analysis macros and internal utility tools to produce statistical summary outputs in a timely matter. Since we utilize our pre-written, validated analysis macros, the validation tasks are much easier.

The following SAS program and the corresponding Adverse Event Summary output demonstrate how easily the output can be produced using analysis dataset structured in one PROC away approach, along with analysis macros.

Table 1: Adverse Event Summary Table.

Gilead Sciences, Inc.
Study 12345

TABLE
NUMBER AND PERCENT OF PATIENTS REPORTING ADVERSE EVENTS
ANALYSIS POPULATION: SAFETY

	Treatment A	Placebo
NUMBER OF PATIENTS	100	100
NUMBER OF PATIENTS WITH ANY EVENTS	1 (1%)	60 (60%)
NUMBER OF PATIENTS WITH EVENT, PER BODY SYSTEM AND PREFERRED TERM		
BODY AS A WHOLE	1 (1%)	60 (60%)
ASTHENIA	1 (1%)	60 (60%)
FLU SYNDROME	1 (1%)	60 (60%)
ABDOMINAL PAIN	1 (1%)	60 (60%)
HEADACHE	1 (1%)	60 (60%)
BACK PAIN	1 (1%)	60 (60%)
PAIN	1 (1%)	60 (60%)

Source: s12345/final_analysis/prog/t_ae.sas v8.2 01Jan1997:00:01

Program 1: SAS program to produce Adverse Event Summary Table

```
/************************************************************************
***  SAVED AS:        t_ae.sas                                           *
***                                                                      *
***  CODED ON:        01Jan1997 by Sandy Chang                           *
***                                                                      *
***  EXECUTED UNDER:  Unix SAS Version 8.2 at Gilead Sciences            *
***                                                                      *
***  DESCRIPTION:     Number of Percentage of Patients Reporting AE      *
***                                                                      *
***  INPUT  DATA SET: intdata.aae                                        *
***                                                                      *
***  OUTPUT DATA SET:                                                    *
***                                                                      *
***  OUTPUT REPORT:   t_ae.out                                           *
***                                                                      *
***  LAST REVISED:                                                       *
***                                                                      *
************************************************************************/

*** Fetch in input data;
%fetch (     library = intdata,
             data    = aae,
             out     = aae,
             sortby  = ptidc
       );


*** macro initialization call;
%tabdesc (   analfile = final,
             colvar   = trtgrp,
             ptvar    = ptidc,
             ocprint  = N,
             col2=50
         );

*** get counts for patients with events;
%tabcount (  sortord  = 1,
             var      = bodysys,
             stat     = n pct n_any pct_any,
             sortby   = a,
             showall  = N,
             anylabel = PATIENTS WITH EVENT
          );

*** get counts for body system and prefer term.
%tabcount (  sortord  = 1,
             sortby   = f,
             showall  = N,
             var      = bodysys prefer_
          );

*** report writing macro call;
%tabdrpt;
```

In addition, One PROC away does not mean that a formatted table can be generated in a single statistical procedure. Rather, it implies that each statistic in the table can be replicated by running a standard statistical analysis routine such as a SAS® PROC or S-PLUS function using the analysis dataset as input. This flexibility helps the FDA reviewers replicate and explore these data with little or no data manipulation, thus allowing reviewers to focus on the study results, not on programming.

Analysis datasets present clear, structured and analyzable data. Rather than recomputed derived variables from the raw data each time they are used, analysis variables are defined only once in one particular analysis dataset. They can be used to generate variables in other analysis datasets, as well as to generate any statistical tables and figures, and subject listings. For example, the date of study day 1, defined in the Demographic analysis dataset, can then be used to calculated the study duration in Demographic, to derive study visits in the Laboratory, or to identify the time of the first event for any Efficacy endpoint.

In addition, using our analysis datasets, GSI statistical programmers can generate validated analysis output for ad-hoc requests for those variables that have been defined in the analysis datasets in one working day or less. With the standardization of the analysis datasets, our programmers know where to find the data to minimize the time for searching for a particular variable.

Prior to the FDA advisor committee meetings, hundreds of ad-hoc requests were anticipated and all responses were generated in the a very short time. During the advisory committee meetings for our recent NDAs, the pre-defined analysis datasets were used to generate real-time responses to questions coming through from FDA advisors. We had less than 45 minutes to provide instant responses for GSI's representatives at the meetings and we were able to come through with the answers.

Finally, organized, structured, and well documented analysis datasets help FDA reviewers understand our data better, generate fewer questions, and shorten the review time. In particular, by using the analysis datasets we discussed above, GSI has received three NDA approvals in the span of 23 months.

STANDARDIZATION

GSI standardizes its analysis datasets in terms of file structure and contents. The standardization is applied to all studies regardless of the therapeutic area. The naming conventions for Efficacy and Safety analysis datasets might differ from one therapeutic area to another. However, the basic structure and the overall content of the analysis datasets are similar, if not identical, across all studies and projects *within* a therapeutic area. In particular, the formats of the analysis datasets are strictly enforced at GSI to ensure standardization across all studies, and no customization of the overall design or general structure is allowed or needed for any individual study.

It is difficult and challenging to standardize the details of analysis datasets across studies. The study designs and scope of the studies vary. The standardization outlines here do not mandate that all derived variables and the "contents" of each of the analysis datasets be identical across all studies. Rather they intend to standardize the file "structure", such as one observation per subject for demographic analysis dataset. This also means that the file names should be consistent across studies. The exact variables and the definition of those variables of each analysis dataset may be study specific, but it is required that the relevant information related to a common analysis purpose be included together. For example, demographic related data should be included in demographic analysis dataset, and efficacy related data should be included in efficacy analysis dataset, and so on. Finally, it is important that the attributes of key common variables and data code lists be consistent across studies as much as possible.

From time to time, we evaluate our standard analysis datasets and seek improvements to make them easier to program, simpler to understand, and to follow the FDA reviewers and ADaM latest suggestions. The current CDISC standard for raw data, Study Data Tabulations Model V3.1 and the upcoming ADaM Analysis Dataset and Documentation models will be evaluated in the context of upcoming FDA guidances.

BENEFITS OF STANDARIZATION

The standardization and consistency of analysis datasets across studies allows users to search information more quickly since all related data is grouped together. With standardization, GSI programmers need only to acquire the basic understanding of the analysis dataset structure and the scope of the analysis datasets. After they are trained on the standard analysis dataset structures, they can then apply the same understanding in programming analysis datasets for any study. This significantly reduces learning curve that study specific analysis dataset customization would otherwise create. As a result, the standardization saves programming time, increases productivity and allows us more resource allocation flexibility.

METADATA

Prior to the advisory committee meetings, companies spend months to prepare for the primary presentations and the backup slides. A typical submission may easily create thousands of backup slides, and each of the backup slides may require one or more analysis outputs to support it. The usage and standardization of analysis datasets achieves the efficiency and accuracy as mentioned above. Equally important, GSI found that analysis dataset metadata, the detailed documentation of the analysis datasets, promotes a clear understanding of the previously created variables, especially for legacy studies. Analysis level metadata allows FDA reviewers to link from a statistical output to the metadata describing the analysis.

At GSI, analysis dataset metadata are documented according to the FDA IT2 and IT3 (??Define IT2, IT3??) Guidelines on Electronic Submission (Item 11 CRF Tabulations). In the Item 11 Documentation, the key component is the document DEFINE.pdf which serves as the roadmap for the FDA reviewers to trace back the variables to their source. CDISC and the FDA are currently defining a new machine-readable metadata, DEFINE.xml that is likely to augment the current guidance in the futute. The content of this metadata builds on the current guidelines and incorporates additional attributes as described by Christiansen and Kubick.

Figure 4 and Figure 5 illustrate examples of GSI Item 11 documentation for analysis dataset metadata.

Figure 4: GSI separates the analysis datasets from the raw CRF datasets in define.pdf.

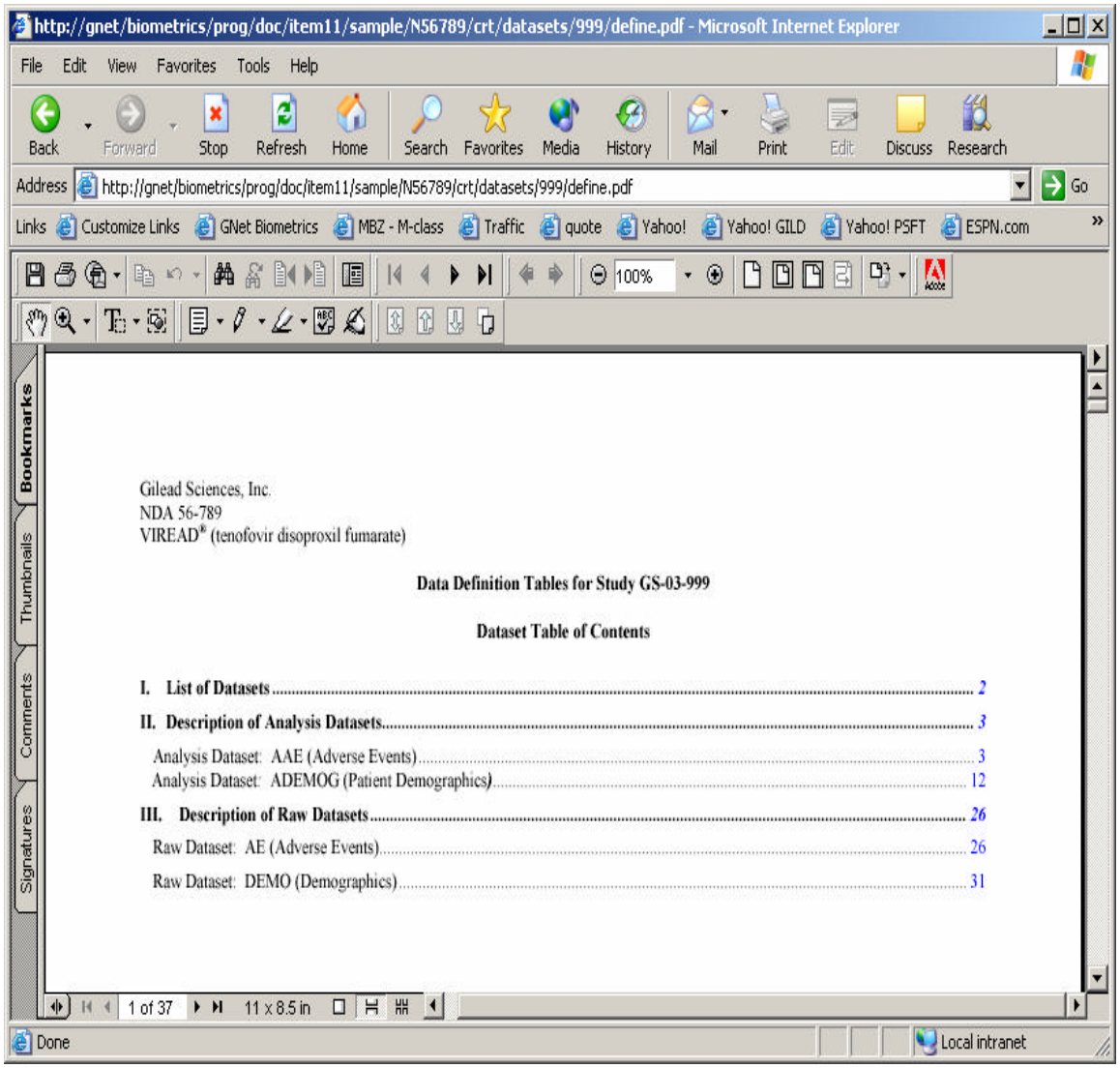
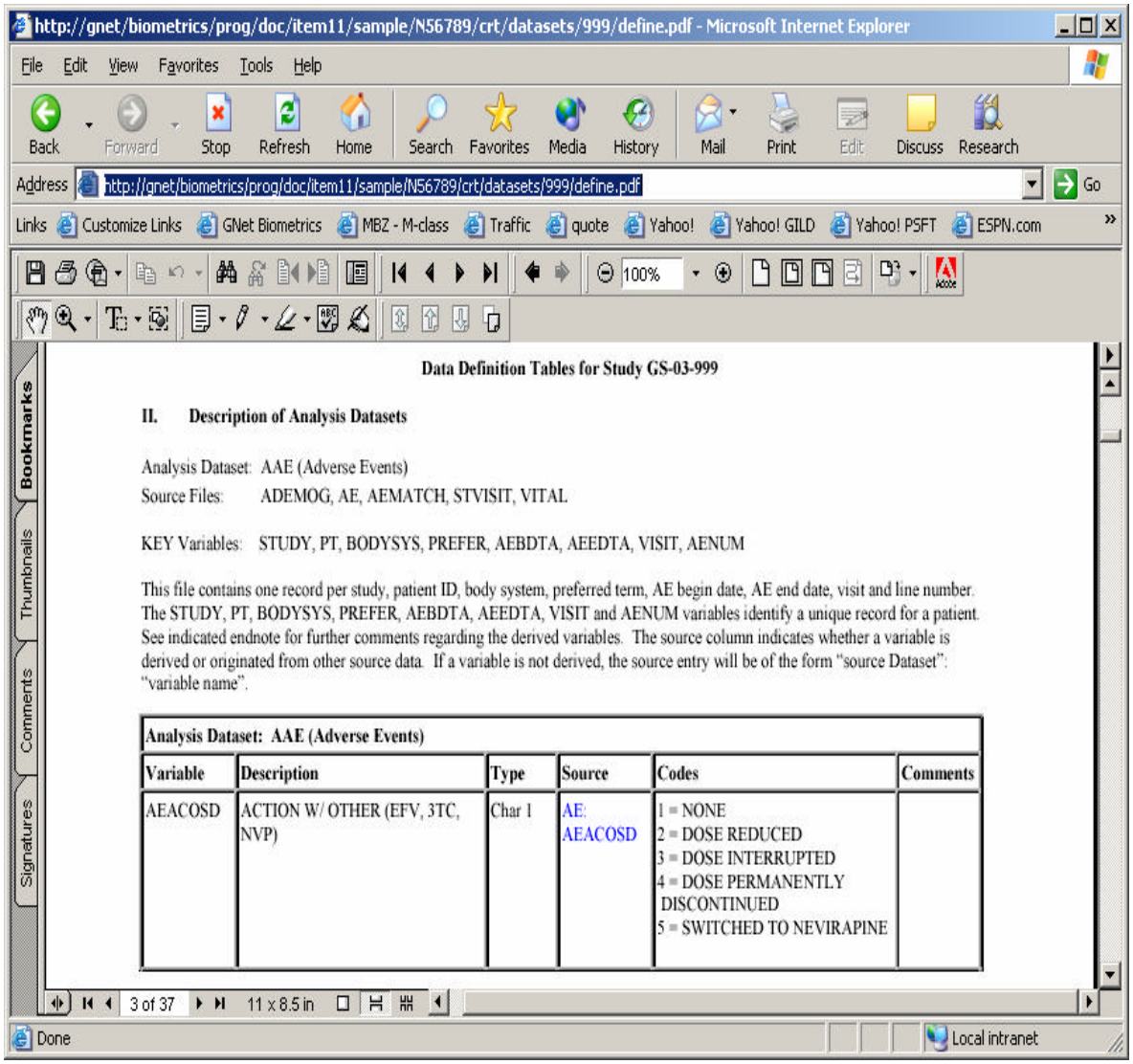


Figure 5: GSI provides a description of each analysis dataset along with key variables and source files.



CONCLUSION

Analysis datasets and their analysis dataset metadata promote the usage of standard structure and analysis documentation. They provide clear, unambiguous communication of the analysis performed by the sponsor. With standard structures, the analysis datasets provide the basis for an efficient analysis model. They help ensure quick programmer turnaround time, maintain consistency across studies, and ease of training and sharing of resources during the critical time of FDA advisory committee meetings.

Regardless whether a submission goes through an advisory committee meeting, the use of analysis datasets has paid GSI handsomely with both immediate returns and long-term benefits.

REFERENCES

CDISC ADaM Statistical Models http://www.cdisc.org/models/adam/index.html

CDISC Study Data Tabulation Models V 3.1 http://www.cdisc.org/models/sds/v3.1/index.html

CDISC website http://www.cdisc.org/standards/index.html

CDISC Submission Metadata Model, D Christiansen and W Kubick, 2001.
http://www.cdisc.org/pdf/SubmissionMetadataModelV2.pdf

FDA Human Drug Advisory Committees http://www.fda.gov/cder/audiences/acspage/#Introduction

Providing Regulatory Submissions in Electronic Format – General Considerations
http://www.fda.gov/cder/guidance/2867fnl.pdf Draft Update http://www.fda.gov/cder/guidance/6028dft.pdf

Providing Regulatory Submissions in Electronic Format – NDAs
http://www.fda.gov/cder/guidance/2353fnl.pdf
Electronic Common Technical Document (eCTD) http://www.fda.gov/cder/regulatory/ersr/#ectd

ACKNOWLEDGMENTS

The authors thank Glenn Itano of Gilead Sciences and Dave Christiansen of Christiansen Consulting for their suggestions regarding this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. The authors may be contacted at:

Sandy Chang
Gilead Sciences, Inc.
333 Lakeside Drive
Foster City, CA 94404
650-522-5285
schang@gilead.com

Steve Wong
Gilead Sciences, Inc.
333 Lakeside Drive
Foster City, CA 94404
650-522-5406
swong@gilead.com