PhUSE US Connect 2018

Paper DV03

Benefits of a Rapidly Growing Library of Clinically Relevant Visualization Templates/Patterns

Eric S. Herbel, Integrated Clinical Systems, Inc., Frenchtown, NJ, USA

ABSTRACT

Over the years many new clinically relevant data visualizations have been defined – many with specific clinical intent, such as identifying patients with potential drug induced liver toxicity, potential renal impairment, increased risk of adverse events or targeted adverse events, oncology study targeted visualizations (Waterfall, Swimmer Plot) and of course different ways to visualize outlier patients. There's a tremendous benefit of continually adding to the library of clinically relevant templates – based on ongoing input from many pharma companies and the US FDA, NCI, etc. The challenge is to not only rapidly provide updates to deploy the new templates, but also to provide an easy to use interface for users to use the new templates/patterns via a drag & drop interface with any of their data structures – SDTM, ADaM as well as EDC raw data format or legacy data structures.

INTRODUCTION

While there's a continued push toward use of standard data structures, CDISC SDTM and ADaM, many organizations continue to review clinical data on ongoing studies in raw EDC format. Alternatively, some organizations spend a large amount of time and resources to transform these to SDTM to be able to use standard data visualizations using general market products. This paper presents several options to address this – focusing on an alternative solution – to build in a high degree of intelligence in flexible templates that can be used directly with EDC raw data format, in addition to standard data structures.

DATA STANDARDIZATION AND STUDY DATA COLLECTION

There has been a steady, increasing trend toward data standardization, especially with the FDA announcement a few years back that 'Studies started after Dec 17, 2016 must use standards in the data catalog (SDTM, ADaM, SEND, define.xml)'. That means – when submitting data to the FDA as part of an application, the data for studies started after Dec 17, 2016 must be submitted in standards based format[1]. At the same time, there's been a significant adoption of EDC or eSource as a data collection method for clinical trials data – using solutions from vendors such as Medidata, Oracle, Omnicomm, Medrio, IBM/Merge, etc. – with more EDC vendors emerging on a regular basis.

Each of these vendors provide a method for exporting the data being collected on a regular basis – typically as SAS datasets or SAS transport files, or ODM XML data stream/web services. Often, this provides a nightly export of data or sometimes more frequent, into receiving databases either in sponsor's databases or hosting solution vendors databases. The data is most often made available directly in this 'raw EDC data format' – not at all in standard CDISC format.

Many vendors also offer the option of doing periodic or ongoing transformation of the EDC raw data format into SDTM format, but this option is usually quite expensive.

DATA VISUALIZATION OF ONGOING CLINICAL STUDIES

Most companies appear to be reviewing clinical data from ongoing studies using the raw EDC format, while a few companies either have external vendors transform the EDC data to standard format for them, or setup internal teams to do that – transforming either to standard CDISC format, or an internal database standard structure. But either of these options typically require considerable personnel and resources to map and process the EDC raw data on a regular basis.

Most companies appear to wait until later in the process to have the study data transformed to CDISC standard format prior to conducting study data analysis and eventual submission to regulatory authorities.

A major reason for some of the companies transforming ongoing study data to a standard format is to allow the use of

general market products to visualize the data in standard format, as it often requires a considerable amount of time and resources to develop these data visualizations– as these general market tools require a considerable amount of time to define the visualizations. So the tendency is to try to reuse the definitions against a standard data format.

An alternative to this would be to define these data visualizations directly against the EDC raw data format, but that typically requires a considerable amount of time and resources to do so.

A third alternative, the subject of this paper, is to use tools that have built in intelligent, clinically relevant templates such that the definition of a wide variety of data/graphic visualizations is extremely quick – as the complexities of the visualization are encapsulated in the template.

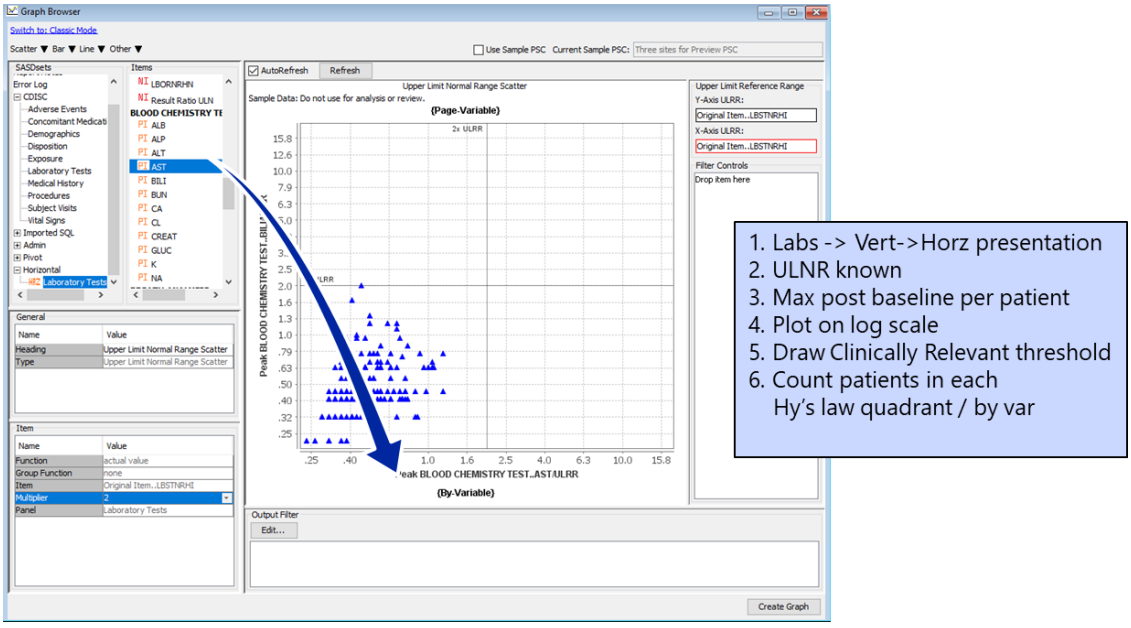
RAPID DEVELOPMENT OF QUICK USE CLINICALLY RELEVANT TEMPLATES

One such tool is the commercial product JReview®, which has been developed over the past 24 years with large amount of input from pharma and biotech companies, CROs, the National Cancer Institute and the US FDA. We see continually evolving ways of visualizing clinical data, and so release updated versions of the product on a frequent basis – at least one major release per year – to incorporate the new ideas as new intelligent, clinically relevant templates.

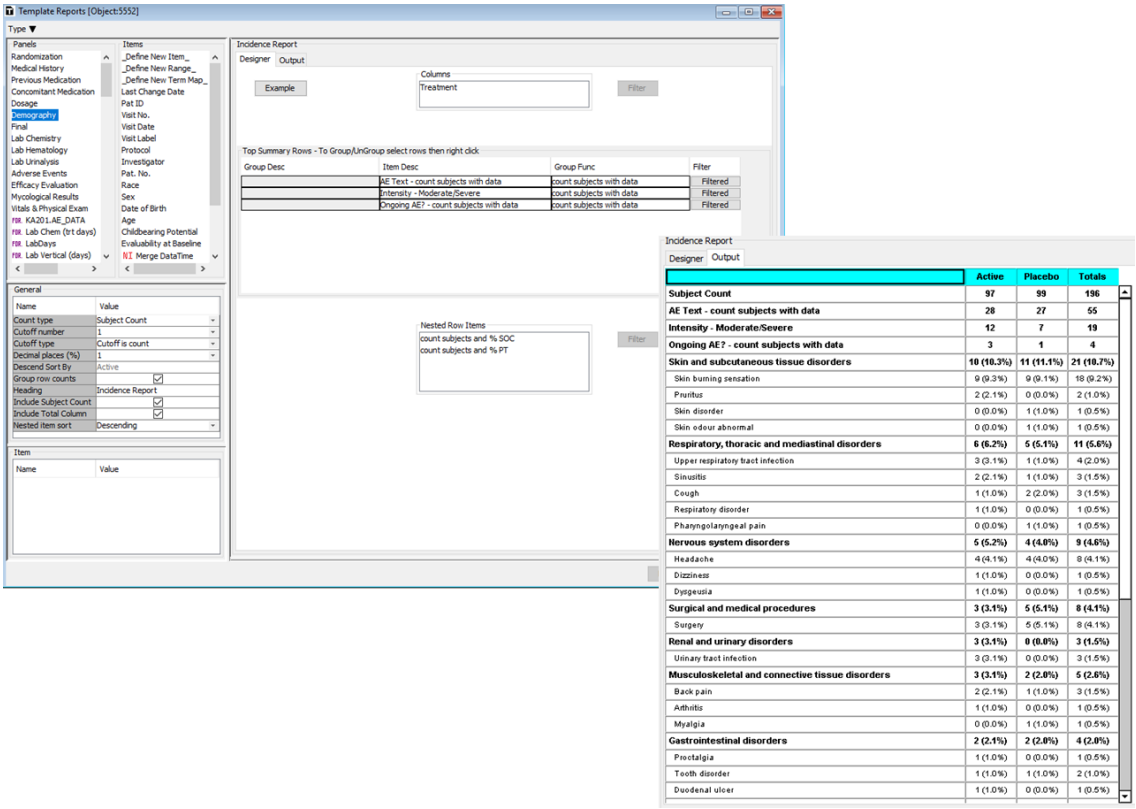
QUICK USE PATTERNS/TEMPLATES

The idea of quick, easy to use clinically relevant intelligent templates – is to provide a user interface where the user can choose from the wide variety of clinically relevant templates, then simply drag and drop the data items that they'd like to include in the template to the appropriate locations on the graph template display. The complexities of the intelligent template is automatically applied for the user – they just specify which variables they'd like to include. So – these variables could equally come from SDTM standard datasets or EDC raw data format – since it's so quick and easy to define.

For example, as the illustration below demonstrates, the user has chosen the 'Upper Limit Normal Range Scatter' – a generic way of describing a Hy's Law chart, but appropriate for use with other types of labs in addition to liver enzymes. In this illustration, as soon as they've drag and dropped the Total Bilirubin (BILI) variable and the AST variable to the y and x axes respectively, the user sees a preview of the results. The interesting thing is that the AST and BILI variables are 'virtual columns' – described below – from vertically shaped lab data. As soon as both axis variables have been selected, the preview displayed has already applied the algorithms/logic/process encapsulated in the template – based on this very commonly used clinically relevant graph. The process steps are described in the steps included in the illustration.



There are many types of clinical relevant graph templates included – but other templates that are helpful include table templates, report templates, patient profile templates and an AE Risk Assessment template – all of which also have embedded intelligence and processes. A good example is the AE Incidence Table or Demographic Summary Table. In the case of the AE Incidence Table, after the user selects that template, they simply drag and drop the variables to be included in the body of the incidence table, such as SOC and PT. After the user selects a few other items – such as a 'column variable', and optionally top summary row variables, a preview of the results is displayed in the 'Output' preview tab. Note that the AE SOC and PTs are automatically presented in descending order of number of patient reports for each SOC, then within the SOC, the PTs included are likewise presented in descending order of number of patient reports for each PT. Further – as is commonly requested, the user can specify a 'cutoff percent' in the 'general' template properties area – such as 1% - which then only displays a PT if at least one cell is at least 1% of the population, and SOCs are only presented if they have at least one PT that meet the cutoff. These and many other options are selected in the 'general template properties' area. An illustration of this is included below.

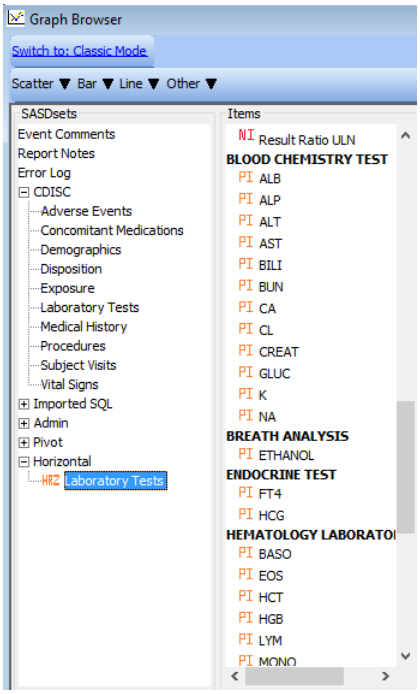


	Active	Placebo	Totals
Subject Count	**97**	**99**	**196**
AE Text - count subjects with data	**28**	**27**	**55**
Intensity - Moderate/Severe	**12**	**7**	**19**
Ongoing AE? - count subjects with data	**3**	**1**	**4**
Skin and subcutaneous tissue disorders	**10 (10.3%)**	**11 (11.1%)**	**21 (10.7%)**
Skin burning sensation	9 (9.3%)	9 (9.1%)	18 (9.2%)
Pruritus	2 (2.1%)	0 (0.0%)	2 (1.0%)
Skin disorder	0 (0.0%)	1 (1.0%)	1 (0.5%)
Skin odour abnormal	0 (0.0%)	1 (1.0%)	1 (0.5%)
Respiratory, thoracic and mediastinal disorders	**6 (6.2%)**	**5 (5.1%)**	**11 (5.6%)**
Upper respiratory tract infection	3 (3.1%)	1 (1.0%)	4 (2.0%)
Sinusitis	2 (2.1%)	1 (1.0%)	3 (1.5%)
Cough	1 (1.0%)	2 (2.0%)	3 (1.5%)
Respiratory disorder	1 (1.0%)	0 (0.0%)	1 (0.5%)
Phanyngolaryngeal pain	0 (0.0%)	1 (1.0%)	1 (0.5%)
Nervous system disorders	**5 (5.2%)**	**4 (4.0%)**	**9 (4.6%)**
Headache	4 (4.1%)	4 (4.0%)	8 (4.1%)
Dizziness	1 (1.0%)	0 (0.0%)	1 (0.5%)
Dysgeusia	1 (1.0%)	0 (0.0%)	1 (0.5%)
Surgical and medical procedures	**3 (3.1%)**	**5 (5.1%)**	**8 (4.1%)**
Surgery	3 (3.1%)	5 (5.1%)	8 (4.1%)
Renal and urinary disorders	**3 (3.1%)**	**0 (0.0%)**	**3 (1.5%)**
Urinary tract infection	3 (3.1%)	0 (0.0%)	3 (1.5%)
Musculoskeletal and connective tissue disorders	**3 (3.1%)**	**2 (2.0%)**	**5 (2.6%)**
Back pain	2 (2.1%)	1 (1.0%)	3 (1.5%)
Arthritis	1 (1.0%)	0 (0.0%)	1 (0.5%)
Myalgia	0 (0.0%)	1 (1.0%)	1 (0.5%)
Gastrointestinal disorders	**2 (2.1%)**	**2 (2.0%)**	**4 (2.0%)**
Proctalgia	1 (1.0%)	0 (0.0%)	1 (0.5%)
Tooth disorder	1 (1.0%)	1 (1.0%)	2 (1.0%)
Duodenal ulcer	1 (1.0%)	0 (0.0%)	1 (0.5%)

WORKING WITH COMMON DATA SHAPES

We've also seen a continued increase in the types of data that are being reported in a vertical shape, i.e., one row per patient/visit/variable, value and other flags associated with the item. The typical reason for reporting data in a vertical shape is for flexibility – to allow for inclusion of variables not originally planned for collection. In addition, this data shape or data structure allows for inclusion of several ancillary values or flags associated with the main test result. The original example of this was lab data – where central labs often report data to sponsors in a vertical shape. Each row contains the patient identification information, visit, collection date, lab analyte, and several ways of reporting the results – in original units, standard units, plus reference ranges in original and standard units, unit of collection, abnormal flag, etc. All this data is great – but tends to be difficult to use for quick reporting or graphing. A solution to the issue was to define the parameters of interest for a specified vertical dataset, for example, for labs – to include the basic items needed for reporting or graphing – such as lab test name, lab test result, low and high reference range, collection date and lab category if available. Based on this information, the system analyzes the data contained in the current lab dataset/table. When the user accesses the 'horizontal view' of the lab dataset/table – they not only see the original items from the dataset, but they also see the list of lab tests (virtual columns) organized by lab category (if available). Then the user can drag and drop the specific (virtual column) lab tests of interest to lab templates, reports, etc., and the system takes care of data transposition on the fly. The additional

information contained in the vertical to horizontal definition provides information to several of the graph templates – that need to know the location of the upper limit of the reference range for a selected lab test. An example of the display of the contents of a horizontal view of a lab dataset is included below.



RAPID ADDITION TO CLINICALLY RELEVANT TEMPLATES

The concept of relying on intelligent, clinically relevant templates to encapsulate the complexities of these graphic and tabular visualizations relies on one very important thing – that the template library included in the product continually grows – expanding based on good ideas from a wide variety of sources.

REFERENCES

1. https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

CONTACT INFORMATION

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

Eric S. Herbel
Integrated Clinical Systems, Inc.
900 State Highway 12
Frenchtown, NJ 08825 USA
Work Phone: +1 908 996 3312
Fax: +1 908 996 7832
Email: eherbel@i-review.com
Web: www.i-review.com

Brand and product names are trademarks of their respective companies.