Paper PH-01

Jack of all Listings, A New Approach for Review of Clinical Data

Hardik Panchal, Celgene Corporation, NJ

ABSTRACT

For an on-going clinical trial, an important process of data review is conducted typically by members of the study team (physician, statistician, data management, programmer) using different outputs where patient data is presented. The purpose of data review is to monitor study progress, assess quality of data and identify trends. Outputs used for this review include listings of study domains like Demographics, Medical History, Adverse Events, Concomitant Medications, Labs, Drug Exposure, Efficacy. Traditional approach is to produce one listing of all Adverse Events, another listing of all Concomitant Medications and so on.

This paper proposes a data listing programmed in SAS and output in MS Excel where in all patient data from different study domains is not only presented together but is also interleaved chronologically by date. This kind of listing offers a more powerful tool for data review. Some of the advantages compared to the traditional approach are:

- Reviewers don't need to sift through different output files to get an assessment of patients' status.
- Reviewers can get a full picture of what information is collected by study sites at each of the patient visits.
- At reviewers' discretion, rows of data from different domains (say AE and Study Drug Exposure) can be put together using data filter feature of MS Excel to identify relational issues in the data.

This kind of all-in-one listing in MS Excel has the potential to reduce the number of traditional data review and patient profile outputs. This translates to savings in time, effort and money for the sponsor.

ABBREVIATIONS

AE – Adverse Event

ConMed – Concomitant Medications

EDC – Electronic Data Capture

ECOG – Eastern Co-operative Oncology Group

ECG - Electrocardiography

TEAE – Treatment Emergent Adverse Event

BACKGROUND

Clinical data review is an activity for most on-going studies where different functions of the study team like physician, statistician, programmer, data management, etc review the collected patient data. Data is reviewed to identify potential data issues, inconsistencies among data collected in different domains, alignment with study protocol and identification of trends in safety and efficacy. For this review, the study team is provided certain pre-defined outputs (or data reports). The study programmer is usually responsible for programming and generating these outputs. Many of the outputs are data listings - some listings are domain specific presenting all records/observations from a particular domain like AE or ConMed, some other reports facilitate cross-checking between domains, eg. AE records put together with Dosing, or AE records put together with Study Discontinuation. There are other reports which are patient specific, also called patient profiles. This paper aims to consolidate all these data reports in to one single output, while providing a much better tool for data review.

Traditional listings are domain specific. That is, an output listing all Adverse Events, another output listing all Concomitant Medication, another output listing all Laboratory observations and so on. A reviewer interested in learning about a particular patient might have to sift through different outputs by different domains. From data-review perspective, a more powerful listing would be one that lists all required information from different datasets (domains) and interleaves observations by date so that a chronological sequence of events can be presented starting from patients' data collected at screening to on-study procedures through follow-up and/or study exit. Such a sequence would indicate whether the patients' condition as shown by an abnormal Lab result is also reflected elsewhere, say in Vital signs result or Physical examination result or an abnormal lab result resulted in reduction or delay of study drug administration and/or whether an AE was recorded. This requires that all data from one domain (say Adverse Event) be stacked on to another domain (say Concomitant Medication) which in turn is stacked on to data from third domain (say Labs) and so on to create a stack of the entire study database. The common thread weaving through such a stack is the patient's ID and the date of topic. The date of topic varies by domain, for example, in AE, ConMeds and Drug Administration datasets, start and stop are the dates of topic. For Labs, Vital Signs, ECG, Tumor/Efficacy Assessment, date of collection or date of assessment are the dates of topic.

Idea is that a clinical study database is an amalgamation of patient data around events (AE, Hospitalization, Medical History, Study Entry, Follow-up Visit, ..), collections (Labs, Vitals, Scans..) and drug administration (Study drug, ConMed, Transfusions, ..) each of which has a date as part of the record. If one scans through the entire database looking for such dates with associated record descriptions and puts these records in a certain order, then a coherent chronological patient history can be created.

FEATURES/USAGE/ADDRESS THE NEED

Proposed listing Image 1:

Note that all data in this and subsequent images are to illustrate the use and format of the proposed listing and do not reflect any real study, patient or treatment.

	A	B	C	D	E	F	G	H	I	
1	**PT**	**dsnx_**	**vblx_**	**label**	**ndtx_**	**CPEVENT**	**VISIT**	**_1**	**_2**	**_3**
2	1001	DEMO	DOBDT	Date of Birth	23-Oct-36	Visit 0	1	Gender: Maled	Race: White, Non-Hispanic and Non-Latino	Race, other:
3	1001	CNHX	PDIAGDT	Date of Primary Diagnosis	25-Mar-08	Visit 0	1	Stage at Primary Diagnosis: II	Anatomic Site of Primary Diagnosis: OrganX	Other, specify:
4	1001	PTSS	STARTDT	Start Date	28-Mar-08	Visit 0	1	Drug Allergies:	Adverse Event Verbatim: CHRONIC PHARYNGITIS	SOC Text: Infectior
5	1001	PTSS	STARTDT	Start Date	2-Apr-08	Visit 0	1	Drug Allergies:	Adverse Event Verbatim: POST OPERATIVE SCAR ON CHEST WALL, LEFT	SOC Text: Skin and disorders
6	1001	BSA	BSADT	Date of BSA Assessment	17-Dec-08	Visit 0	1	Date of Height, Weight, BSA Assessment: 2008	Height (cm): 165	Height (cm)-EXC:
7	1001	ECOG	ASSESSDT	Date of Assessment	17-Dec-08	Visit 0	1	ECOG Score: 1	Date of Assessment: 17DEC2008	
8	1001	ENRL	INFCONDT	Informed Consent Date	17-Dec-08	Visit 0	1	Patient Initials: XXX	Which treatment regimen assigned?: ArmX	Informed Consent
9	1001	PAPN	ASSESSDT	Date of Assessment	17-Dec-08	Visit 0	1	No Sensory Neuropathy: Check	PAPN Score:	Date of Assessmer
10	1001	PE	PEDT	Date of Physical Exam	17-Dec-08	Visit 0	1	Date of Physical Exam: 17DEC2008		
11	1001	SELF	ASSESSDT	Date of Assessment	17-Dec-08	Visit 0	1	PRO Question Text: I AM BOTHERED BY THE WAY MY HANDS OR NAILS LOOK.	PRO Question Response: Not at All	Date of Assessmer
12	1001	SELF	ASSESSDT	Date of Assessment	17-Dec-08	Visit 0	1	PRO Question Text: I FEEL BLOATED.	PRO Question Response: Not at All	Date of Assessmer
13	1001	VITL	VITALDT	Date of Vitals	17-Dec-08	Visit 0	1	Date of Vitals: 17DEC2008		
14	1001	ECG	ECGDT	Date of ECG	18-Dec-08	Visit 0	1	Date of ECG: 18DEC2008		
15	1001	LESN	SCANDT	Date of Scan, PE, etc.	18-Dec-08	Visit 0	1	Sum of Longest Diameters (cm): 2.4	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG
16	1001	LESN	SCANDT	Date of Scan, PE, etc.	18-Dec-08	Visit 0	1	Sum of Longest Diameters (cm): 2.4	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG
17	1001	ENRL	RANDDT	Date of Randomization	23-Dec-08	Visit 0	1	Patient Initials: XXX	Which treatment regimen assigned?: ArmX	Informed Consent
18	1001	CMED	STARTDT	Start Date	24-Dec-08	MEDS	901	Drug Name: DEXAME	Preferred Drug Name: DEXAME	Continuing:
19	1001	CMED	STARTDT	Start Date	24-Dec-08	MEDS	901	Drug Name: DIPHEN	Preferred Drug Name: DIPHEN	Continuing:
20	1001	CMED	STARTDT	Start Date	24-Dec-08	MEDS	901	Drug Name: ONDANS	Preferred Drug Name: ONDANS	Continuing:
21	1001	CMED	STARTDT	Start Date	24-Dec-08	MEDS	901	Drug Name: RANITI	Preferred Drug Name: RANITI	Continuing:
22	1001	DOSE	DOSEDT	Dose Date	24-Dec-08	Visit 1	2	Dose Date: 24DEC2008	Dosing Drug: Study Drug 2	Dose (mg/m^2):
23	1001	DOSE	DOSEDT	Dose Date	24-Dec-08	Visit 1	2	Dose Date: 24DEC2008	Dosing Drug: Study Drug 1	Dose (mg/m^2): 2(

Column A is the patient number. Columns B to E represent the value and attributes of the Date of topic with column B/C/D/E for source dataset/variable name/variable label/date value respectively. Columns F and G are visit label and visit number associated with the date from respective datasets. Columns H onwards are domain specific {variable label: value} pairs, eg. “Gender: Male”, “ECOG Score: 1”. These are named _1, _2, _3, .. in the column header (first row). Variables _1, _2, _3, .. is a unified naming (or renaming) convention used so that information from different datasets can be stacked with fewer or minimum required variables making it a user friendly presentation/review. Keeping different datasets (AE, ConMed, Labs, ..) stacked with their original variable names would create an enormous number of variables (or columns) in the final output which can be very difficult to review.

Since the listing is sorted by patient ID and date of topic, any patient's initial records will describe data collected at screening/baseline as in the image above. Notice that, among other things, the listing reveals when the patients' first dose of study drug occurred (row 22, Dose Date 24-Dec-08) and how much dose was administered (image is truncated due to space limitation).

Image 2:

	A	B	C	D	E	F	G	H	I	J
1	**PT**	**dsnx_**	**vblx_**	**label**	**ndtx_**	**CPEVENT**	**VISIT**	**_1**	**_2**	**_3**
24	1001	ECOG	ASSESSDT	Date of Assessment	24-Dec-08	Visit 1	2	ECOG Score: 1	Date of Assessment: 24DEC2008	
25	1001	LAB	COLLECDT	Collection Date	24-Dec-08	Visit 1	2	Laboratory Test Name: HGB	Laboratory Test Unit: G/L	Laboratory Test Value: 1
26	1001	LAB	COLLECDT	Collection Date	24-Dec-08	Visit 1	2	Laboratory Test Name: PLATE	Laboratory Test Unit: X10^9/L	Laboratory Test Value: 1
27	1001	LAB	COLLECDT	Collection Date	24-Dec-08	Visit 1	2	Laboratory Test Name: RBC	Laboratory Test Unit: X10^12/L	Laboratory Test Value: 4
28	1001	PE	PEDT	Date of Physical Exam	24-Dec-08	Visit 1	2	Date of Physical Exam: 24DEC2008		
29	1001	SELF	ASSESSDT	Date of Assessment	24-Dec-08	Visit 1	2	PRO Question Text: I AM BOTHERED BY THE WAY MY HANDS OR NAILS LOOK.	PRO Question Response: Not at All	Date of Assessment: 24D
30	1001	SELF	ASSESSDT	Date of Assessment	24-Dec-08	Visit 1	2	PRO Question Text: I FEEL BLOATED.	PRO Question Response: Not at All	Date of Assessment: 24D
31	1001	LAB	COLLECDT	Collection Date	30-Dec-08	Visit 2	3	Laboratory Test Name: HGB	Laboratory Test Unit: G/L	Laboratory Test Value: 1
32	1001	LAB	COLLECDT	Collection Date	30-Dec-08	Visit 2	3	Laboratory Test Name: PLATE	Laboratory Test Unit: X10^9/L	Laboratory Test Value: 1
33	1001	LAB	COLLECDT	Collection Date	30-Dec-08	Visit 2	3	Laboratory Test Name: RBC	Laboratory Test Unit: X10^12/L	Laboratory Test Value: 4
34	1001	LAB	COLLECDT	Collection Date	6-Jan-09	Visit 3	4	Laboratory Test Name: HGB	Laboratory Test Unit: G/L	Laboratory Test Value: 1
35	1001	LAB	COLLECDT	Collection Date	6-Jan-09	Visit 3	4	Laboratory Test Name: PLATE	Laboratory Test Unit: X10^9/L	Laboratory Test Value: 3
36	1001	LAB	COLLECDT	Collection Date	6-Jan-09	Visit 3	4	Laboratory Test Name: RBC	Laboratory Test Unit: X10^12/L	Laboratory Test Value: 4
37	1001	AE	STARTDT	Start Date	10-Jan-09	AES	902	Adverse Event Verbatim: ANEMIA	SOC Text: Blood and lymphatic system disorders	PT Text: Anemia
38	1001	DOSE	DOSEDT	Dose Date	14-Jan-09	Visit 4	5	Dose Date: 14JAN2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):
39	1001	DOSE	DOSEDT	Dose Date	14-Jan-09	Visit 4	5	Dose Date: 14JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 200
40	1001	ECOG	ASSESSDT	Date of Assessment	14-Jan-09	Visit 4	5	ECOG Score: 1	Date of Assessment: 14JAN2009	
41	1001	PAPN	ASSESSDT	Date of Assessment	14-Jan-09	Visit 4	5	No Sensory Neuropathy: Check	PAPN Score:	Date of Assessment: 14J
42	1001	PE	PEDT	Date of Physical Exam	14-Jan-09	Visit 4	5	Date of Physical Exam: 14JAN2009		
43	1001	SELF	ASSESSDT	Date of Assessment	14-Jan-09	Visit 4	5	PRO Question Text: I AM BOTHERED BY THE WAY MY HANDS OR NAILS LOOK.	PRO Question Response: Not at All	Date of Assessment: 14J
44	1001	SELF	ASSESSDT	Date of Assessment	14-Jan-09	Visit 4	5	PRO Question Text: I FEEL BLOATED.	PRO Question Response: Not at All	Date of Assessment: 14J
45	1001	AE	STARTDT	Start Date	28-Jan-09	AES	902	Adverse Event Verbatim: FATIGUE	SOC Text: General disorders and administration site conditions	PT Text: Fatigue
46	1001	LESN	SCANDT	Date of Scan, PE, etc.	28-Jan-09	WEEK6	220	Sum of Longest Diameters (cm): 2.3	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG
47	1001	LESN	SCANDT	Date of Scan, PE, etc.	28-Jan-09	WEEK6	220	Sum of Longest Diameters (cm): 2.3	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG

Image 2 of the listing presents patient data collected across visits 1 through 4 (column F, CPEVENT) and a visit labeled Week 6. Information presented is from different source datasets – ECOG, Lab, Physical Exam (PE), Patient Reported Outcome (SELF), AE, Efficacy (LESN). An AE onset (Anemia) is reported between Visits 3 and 4. Another AE onset (Fatigue) is reported between Visit 4 and Week 6 observations. All information collected about AE like AE grade, outcome, relationship to study drug, action taken, serious is also displayed in the Excel output under subsequent truncated columns beyond '_3' (Excel column J). Visits 2 and 3 did laboratory collection. Hgb, Platelet and RBC results are displayed. Next study drug dosing after Visit 1 occurs at Visit 4 which can be seen in rows 38 and 39. Visit 4 also had ECOG assessment, physical exam and PRO questionnaire filled out.

Site compliance to study protocol can also be assessed through this kind of listing. One can review the various study domains or data fields collected at a certain visit against assessments that were required in the schedule of observations and procedures (also called Table of Events) in the study protocol. Dataset names in column B (dsnx_) can be replaced with dataset labels for reviewers unfamiliar with the dataset naming conventions. The column C (vblx_) variable name can be dropped since date variables are described by their respective labels in column D (label) in above image. A revised listing is shown in Image 3.

Image 3:

	A	B	C	D	E	F	G	H	I
1	**Patient**	**Dataset/Domain**	**Field Label**	**Date**	**Visit Label**	**Visit**	**_1**	**_2**	**_3**
2	1005	Demographics	Date of Birth	4-Apr-53	BASELINE	1	Gender: Male	Race: White, Non-Hispanic and Non-Latino	Race, other:
3	1005	Pre-Treatment Diagnoses Signs & Symptoms	Start Date	7-Nov-08	BASELINE	1	SOC Text: Vascular disorders	PT Text: Arteriosclerosis	Continuing: Check
4	1005	Cancer History	Date of Primary Diagnosis	18-Nov-08	BASELINE	1	Stage at Primary Diagnosis: IV	Anatomic Site of Primary Diagnosis: OrganY	Other, specify:
5	1005	Prior and Concomitant Medications	Start Date	2-Dec-08	MEDS	901	Preferred Drug Name: LIDOCAINE	Continuing:	Indication: ANAESTESIA FOR BRONCHOSCOPY
6	1005	Prior and Concomitant Medications	Stop Date	2-Dec-08	MEDS	901	Preferred Drug Name: LIDOCAINE	Continuing:	Indication: ANAESTESIA FOR BRONCHOSCOPY
7	1005	Pre-Treatment Diagnoses Signs & Symptoms	Start Date	2-Dec-08	BASELINE	1	SOC Text: Investigations	PT Text: Bronchoscopy	Continuing:
8	1005	Subject Enrollment Notification	Informed Consent Date	23-Dec-08	BASELINE	1	Patient Initials: UUU	Which treatment regimen assigned?: ArmX	Informed Consent Date: 23DEC2008
9	1005	ECG	Date of ECG	25-Dec-08	BASELINE	1	Date of ECG: 25DEC2008		
10	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm):	Lesion Type: NTLE	Lesion Number: 1
11	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm): 26.1	Lesion Type: TARG	Lesion Number: 1
12	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm): 26.1	Lesion Type: TARG	Lesion Number: 2
13	1005	Height, Weight, BSA Assessment	Date of BSA Assessment	26-Dec-08	BASELINE	1	Date of Height, Weight, BSA Assessment: 2008	Height (cm): 187	Height (cm)-EXC:
14	1005	ECOG Zubrod Performance Status	Date of Assessment	26-Dec-08	BASELINE	1	ECOG Score: 0	Date of Assessment: 26DEC2008	
15	1005	Physician Assessment Sensory Neuropathy	Date of Assessment	26-Dec-08	BASELINE	1	No Sensory Neuropathy: Check	PAPN Score:	Date of Assessment: 26DEC2008
16	1005	Physical Exam	Date of Physical Exam	26-Dec-08	BASELINE	1	Date of Physical Exam: 26DEC2008		
17	1005	Pat. Self Asmnt. Peripheral Neuropathy	Date of Assessment	26-Dec-08	BASELINE	1	FACT-Taxane Question Text: I AM BOTHERED BY THE WAY MY HANDS OR NAILS LOOK.	FACT-Taxane Question Response: Not at All	Date of Assessment: 26DEC2008

Image 4:

	A	B	C	D	E	F	G	H	I	J
1	**PT**	**dsnx_**	**vblx_**	**label**	**ndtx_**	**CPEVENT**	**VISIT**	**_1**	**_2**	**_3**
44	1001	SELF	ASSESSDT	Date of Assessment	14-Jan-09	Visit 4	5	PRO Question Text: I FEEL BLOATED.	PRO Question Response: Not at All	Date of Assessment: 1
45	1001	AE	STARTDT	Start Date	28-Jan-09	AES	902	Adverse Event Verbatim: FATIGUE	SOC Text: General disorders and administration site conditions	PT Text: Fatigue
46	1001	LESN	SCANDT	Date of Scan, PE, etc.	28-Jan-09	WEEK6	220	Sum of Longest Diameters (cm): 2.3	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG
47	1001	LESN	SCANDT	Date of Scan, PE, etc.	28-Jan-09	WEEK6	220	Sum of Longest Diameters (cm): 2.3	Sum of Longest Diameters (cm)-EXC:	Lesion Type: TARG
48	1001	RESP	RESPDT	Response Date	28-Jan-09	WEEK6	220	Target Response Criteria: SD	Overall Response: SD	Tumor Progression:
49	1001	AE	STOPDT	Stop Date	4-Feb-09	AES	902	Adverse Event Verbatim: FATIGUE	SOC Text: General disorders and administration site conditions	PT Text: Fatigue
50	1001	DOSE	DOSEDT	Dose Date	4-Feb-09	Visit 5	8	Dose Date: 04FEB2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):

Notice above that the same AE record is presented twice as per the dates of topic – in this case Start and Stop date of AE. Row 45 shows onset of AE of Fatigue after Visit 4 and row 49 shows that it resolved at Visit 5 (since the Stop date of AE is same as the Dose date of Visit 5). In other words, the listing presents the span of an AE in the study, at the same time, presenting all other events, assessments and drug administrations at/around/between the onset and resolution of AE. The 'span' feature is also relevant to other domains where both - start and stop are collected as in Concomitant/Prior Medications, Study Drug Administration, Hospitalization, etc.

As in the above 4 images, this listing displays records in chronological order all the way up to the last data point collected for the patient. For on-going patients it can tells the reviewer which is the last Visit or Cycle that the patient is on. For patients who have discontinued study drug, it tells the reviewer which is the last visit or cycle or the number of visits or cycles the patient had the study drug before discontinuing. It displays records from Study Drug Discontinuation domain indicating the reason and date of discontinuation and depending on study design, whether the patient is continuing in study follow-up period. Depending on how and when data is specified to be collected in the study protocol, it also displays information about patient status in follow-up. Is the patient on another therapy in follow-up? What kind of therapy (or therapies)? Are attempts of patient phone contact during follow-up successful? How many weeks or months of patient follow-up data is available before study closure? All these and many more questions can be answered from this single output.

Using the data filter feature of MS Excel, the following in-depth data review can be done. Some of the preferred data filters for Safety and Efficacy that reviewers might zoom in –

Image 5 - AE and Study Drug Administration (or Dosing):

	A	B	C	D	E	F	G	H	I	J
1	Patient	Dataset/Domain	Field Lab	Da	Visit Labe	V	_1	_2	_3	_4
34	1007	Study Drug Dosing	Dose Date	29-Dec-08	C1D1	2	Dose Date: 29DEC2008	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 869
35	1007	Study Drug Dosing	Dose Date	29-Dec-08	C1D1	2	Dose Date: 29DEC2008	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
57	1007	Study Drug Dosing	Dose Date	5-Jan-09	C1D8	3	Dose Date: 05JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2):	Total Calculated Dose (mg):
58	1007	Adverse Events	Start Date	10-Jan-09	AES	902	SOC Text: Skin and subcutaneous tissue disorders	PT Text: Alopecia	Grade: 2	Outcome (Check one): Ongoing
59	1007	Study Drug Dosing	Dose Date	13-Jan-09	C1D15	4	Dose Date: 13JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
64	1007	Study Drug Dosing	Dose Date	20-Jan-09	C2D1	5	Dose Date: 20JAN2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 900
65	1007	Study Drug Dosing	Dose Date	20-Jan-09	C2D1	5	Dose Date: 20JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
85	1007	Study Drug Dosing	Dose Date	26-Jan-09	C2D8	6	Dose Date: 26JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
86	1007	Study Drug Dosing	Dose Date	2-Feb-09	C2D15	7	Dose Date: 02FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
87	1007	Adverse Events	Start Date	6-Feb-09	AES	902	SOC Text: Investigations	PT Text: Weight increased	Grade: 1	Outcome (Check one): Resolved
100	1007	Adverse Events	Start Date	9-Feb-09	AES	902	SOC Text: Blood and lymphatic system disorders	PT Text: Neutropenia	Grade: 1	Outcome (Check one): Resolved
101	1007	Adverse Events	Stop Date	12-Feb-09	AES	902	SOC Text: Blood and lymphatic system disorders	PT Text: Neutropenia	Grade: 1	Outcome (Check one): Resolved
104	1007	Study Drug Dosing	Dose Date	25-Feb-09	C3D1	8	Dose Date: 25FEB2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 859
105	1007	Study Drug Dosing	Dose Date	25-Feb-09	C3D1	8	Dose Date: 25FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
125	1007	Study Drug Dosing	Dose Date	3-Mar-09	C3D8	9	Dose Date: 03MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
126	1007	Study Drug Dosing	Dose Date	10-Mar-09	C3D15	10	Dose Date: 10MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
147	1007	Study Drug Dosing	Dose Date	17-Mar-09	C4D1	11	Dose Date: 17MAR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 851
148	1007	Study Drug Dosing	Dose Date	17-Mar-09	C4D1	11	Dose Date: 17MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
152	1007	Study Drug Dosing	Dose Date	24-Mar-09	C4D8	12	Dose Date: 24MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
153	1007	Study Drug Dosing	Dose Date	30-Mar-09	C4D15	13	Dose Date: 30MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
169	1007	Study Drug Dosing	Dose Date	7-Apr-09	C5D1	14	Dose Date: 07APR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 883
170	1007	Study Drug Dosing	Dose Date	7-Apr-09	C5D1	14	Dose Date: 07APR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
191	1007	Adverse Events	Start Date	9-Apr-09	AES	902	SOC Text: Nervous system disorders	PT Text: Hypoaesthesia	Grade: 1	Outcome (Check one): Ongoing at time of Death

Data is filtered to keep only Study Drug Dosing and Adverse Events for column B (Dataset/Domain). Above snapshot shows how the onset of AEs occurs with respect to study drug administration. It shows that AEs occurred after drug administration visits of Cycle-1-Day-8, C2D15 and C5D1. In addition to AE information, it can be seen whether any AE resulted in dose adjustment (reduction or delay). Interleaving of AEs with Dosing as in above image might give more perspective than if Dosing and AE records are presented in separate outputs or separate sections within patient profile.

Image 6 – Study Dosing and Response Assessment:

Below image shows Lesion measure (efficacy parameter) and Response Evaluation with respect to dosing visits.

	A	B	C	D	E	F	G	H	I	J
1	Patient	Dataset/Domain	Field Lab	Da	Visit Labe	V	_1	_2	_3	_4
9484	1010	Lesion Identification & Evaluation	Date of Scan, PE, etc.	17-Feb-09	BASELINE	1	Sum of Longest Diameters (cm): 5.8	Lesion Type: TARG	Lesion Number: 1	Site Code: LU
9535	1010	Study Drug Dosing	Dose Date	19-Feb-09	C1D1	2	Dose Date: 19FEB2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 468
9536	1010	Study Drug Dosing	Dose Date	19-Feb-09	C1D1	2	Dose Date: 19FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
9667	1010	Study Drug Dosing	Dose Date	26-Feb-09	C1D8	3	Dose Date: 26FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
9777	1010	Study Drug Dosing	Dose Date	5-Mar-09	C1D15	4	Dose Date: 05MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
9966	1010	Study Drug Dosing	Dose Date	16-Mar-09	C2D1	5	Dose Date: 16MAR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 450
9967	1010	Study Drug Dosing	Dose Date	16-Mar-09	C2D1	5	Dose Date: 16MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
10096	1010	Study Drug Dosing	Dose Date	24-Mar-09	C2D8	6	Dose Date: 24MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2):	Total Calculated Dose (mg):
10148	1010	Study Drug Dosing	Dose Date	31-Mar-09	C2D15	7	Dose Date: 31MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
10173	1010	Lesion Identification & Evaluation	Date of Scan, PE, etc.	31-Mar-09	WEEK6	220	Sum of Longest Diameters (cm): 3.9	Lesion Type: TARG	Lesion Number: 1	Site Code: LU
10198	1010	"Response Evaluation"	Response Date	31-Mar-09	WEEK6	220	Overall Response: PR	Response Assessment was not at a schedul:	Tumor Progression:	AE/Toxicity (record on AE CRF):
10266	1010	Study Drug Dosing	Dose Date	7-Apr-09	C3D1	8	Dose Date: 07APR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 462
10267	1010	Study Drug Dosing	Dose Date	7-Apr-09	C3D1	8	Dose Date: 07APR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
10336	1010	Study Drug Dosing	Dose Date	14-Apr-09	C3D8	9	Dose Date: 14APR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2):	Total Calculated Dose (mg):
10395	1010	Study Drug Dosing	Dose Date	21-Apr-09	C3D15	10	Dose Date: 21APR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 131
10630	1010	Lesion Identification & Evaluation	Date of Scan, PE, etc.	8-May-09	WEEK12	221	Sum of Longest Diameters (cm): 1.8	Lesion Type: TARG	Lesion Number: 1	Site Code: LU
10655	1010	"Response Evaluation"	Response Date	8-May-09	WEEK12	221	Overall Response: PR	Response Assessment was not at a schedul:	Tumor Progression:	AE/Toxicity (record on AE CRF):
10664	1010	Study Drug Dosing	Dose Date	15-May-09	C4D1	11	Dose Date: 15MAY2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 342
10665	1010	Study Drug Dosing	Dose Date	15-May-09	C4D1	11	Dose Date: 15MAY2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 75	Total Calculated Dose (mg): 98
10791	1010	Study Drug Dosing	Dose Date	22-May-09	C4D8	12	Dose Date: 22MAY2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 75	Total Calculated Dose (mg): 98
10894	1010	Study Drug Dosing	Dose Date	29-May-09	C4D15	13	Dose Date: 29MAY2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 75	Total Calculated Dose (mg): 98
11101	1010	Lesion Identification & Evaluation	Date of Scan, PE, etc.	11-Jun-09	WEEK18	222	Sum of Longest Diameters (cm): 1.5	Lesion Type: TARG	Lesion Number: 1	Site Code: LU
	1010	"Response	Response	11-Jun-09	WEEK18	222	Overall Response: PR	Response Assessment was	Tumor Progression:	AE/Toxicity (record on AE CRF):

Data is filtered on Study Drug Dosing, Lesion Identification and Response Evaluation for column B (Dataset/Domain). It can be seen in above image that the baseline lesion measurement of 5.8 cm became 3.9 cm by Cycle-2-Day-15 after about 5 dosing visits, 1.8 cm by C4D1 after another 3 dosing visits and 1.5 cm by C5D1. Overall Response is determined to be Partial Response (PR) for the visits shown.

Some of the other review capabilities –

- Need to know how many patients, what patient IDs and what kind of Adverse Events resulted in Discontinuation of Study drug?

Solution: Filter on AE dataset (column B or 'dsnx_' above in image) and among those observations filter on AE Action taken variable (will be one of the columns from _1, _2, _3, …. (typically max number of columns overall for the entire listing can be somewhere between _20 to _35, depending on how many variables (or data fields) of interest need to be included).

- Need to put AE and Dosing records together to review what AEs caused modification or adjustment of study drug administration (dose reduced or interrupted)?

Solution: Filter on column B or 'dsnx_' keeping AE and DOSE records/observations (Image 5 above). Similarly, depending on need or style of review, (AE and Medical History) or (AE and Concomitant Medication) or (ECG and Medical History) records can be put together to identify relationships or inconsistencies.

- Need to know the incidence of an AE that is of special interest for your study?

Solution: Go to the column which lists AE Preferred term and filter that column for the particular AE that is of interest (say Dysgeusia) for your study to get a list of patient IDs to whom that AE occurred. To see the profile of any of the patients from this list, remove the filter on AE Preferred term and apply filter for the patient ID (first column).

- Need to know how many patients had Visit 7 dosing or how many patients had efficacy assessment at Visit 9?

Solution: Filter CPEVENT (column F) for Visit 7 and filter 'dsnx_' (column B) for 'DOSE'. Similarly, apply different filter for efficacy and visit 9.

The bottom-line is that this listing is like a kaleidoscope. A Rubik's cube is another analogy. Point is that nearly all answers are within this single-output listing; the user or reviewer just needs to manipulate it to see what they need to see. Each member of a study team may use it in their own way. Flexibility of MS Excel combined with programmatic technique of appending all data together in conveniently placed rows and columns provides immense possibilities of data review.

Traditional listing outputs for data review have the limitation that they are static (pre-defined standard or study-specific outputs are tailored to answer a fixed set of questions) and the reviewer (say study scientist or statistician) often comes back to the programmer for additional outputs. Much time and effort is spent in collecting specifications and programming/QC of the new outputs because the original set of data reports didn't satisfy the reviewer's needs. The proposed output actually pushes the work from the programmer to the reviewer by giving them the flexibility to make their own reports. This way, more time and effort of the study team is spent on the critical activity of reviewing the study data than in specifications/development/QC of new reports. More review provides more opportunity to identify data issues early leading to cleaner data and avoiding unseemly outcomes such as unlocking databases because an issue was found *after* DB lock.

Like the raw/reported data are presented in the listing images shown earlier, derived patient-specific data fields that are of interest can also be merged. Information about patients meeting efficacy criteria, the last known status of patient in the study, whether the patient had any on-study serious AEs can be derived and included in the listing. Image below shows two such columns that are shaded.

Image 7:

	A	B	C	D	E	F	G	H
1	Patient	Best Overall Response	Treatment Related TEAE of Grade>=3	Dataset/Domain	Field Label	Date	Visit Label	Visit
2								
3								
4								
685	1015	(PR) Partial Response	Yes	Demographics	Date of Birth	4-Dec-57	BASELINE	1
686	1015	(PR) Partial Response	Yes	Prior and Concomitant Medications	Start Date	18-Jun-08	MEDS	901
687	1015	(PR) Partial Response	Yes	Prior and Concomitant Medications	Start Date	18-Jun-08	MEDS	901
688	1015	(PR) Partial Response	Yes	Prior and Concomitant Medications	Start Date	18-Jun-08	MEDS	901
689	1015	(PR) Partial Response	Yes	Prior and Concomitant Medications	Start Date	18-Jun-08	MEDS	901
690	1015	(PR) Partial Response	Yes	Prior and Concomitant Medications	Start Date	18-Jun-08	MEDS	901

Another advantage of this listing is that it helps considerably in QC of statistical programming deliverables like Analysis Datasets, Tables, Listings, Graphs. Before the final output is exported to MS Excel, it is in SAS dataset format. This dataset can be saved and used by the study programmer to answer frequent adhoc requests from study team. Quick data-checks done using output SAS dataset. For instance:

- (1) To derive date of last known alive or last point of contact with the patient with regard to study the following code could be used.

```
PROC SQL;
      create table LCONTACT as
/* PT is patient ID. Max(ndtx_) fetches the latest date from all dates of topic as the last contact date */
      select pt, max(ndtx_) as lcontact format=date9.
      from ALLCDTS /*ALLCDTS is final output datatset of proposed listing.*/
      where ndtx_ is not missing
          & pt in (select pt
                   from DERIVED.D_POP
                   where .. /* applicable where condition to look up patients from a select population */)
          & .. /* additional where clauses as applicable to keep or exclude certain dates of topic, for example
in this case, date of death */
      group by pt
      order by pt;
quit.
```

- (2) To list patient observations that are reported/collected out of chronological order such as Visit/Cycle 5 date for a patient is in May 2014 but Visit/Cycle 6 date for the same patient is in March 2014.

Variable ‘cyn’ used below in query could be the Visit or Cycle number. In order for ‘cyn’ to be comparable among subsequent and previous visit observations of a patient, it needs to have consistent unit, be it Visit # or Cycle # or Week #.

```
PROC SQL;
      create table OVERLAP as
      select distinct a.pt, a.DSNX_, a.VBLX_, a.nDTX_, a.VISname, a.LABEL,
                           b.DSNX_ as dsnx_b, b.VBLX_ as vblx_b, b.nDTX_ as ndtx_b,
                          b.VISname as visnm_b, b.LABEL as label_b, a.ndtx_-b.ndtx_+1 as overlap_days
      from /*SQL clone feature, merge dataset with itself*/ ALLCDTS a, ALLCDTS b
      where a.pt=b.pt
          & .<a.cyn<b.cyn
          & a.ndtx_>b.ndtx_+10>.;
quit;
```

DEVELOPMENT

A big part of the development program is a play on metadata of study database which is done using standard SAS views – SASHELP.VCOLUMN and SASHELP.VTABLE. It is a one time development process that can be conducted at study start-up. During the on-going study period, the program can be re-run whenever new datasets are extracted from EDC. Such re-runs can produce a fresh output listing that can be shared with reviewers. Following steps describe high level process/approach:-

1. Select a pair of dataset and date variable at a time, eg. AE dataset and AE Start date is one pair; AE dataset and AE Stop date is another pair, Laboratory dataset and Lab collection date is another pair and so on.
2. Create a temporary dataset for each pair for the list of patients that need to be reviewed, it could be one patient, few patients of interest, 5 randomly selected patients from the study or all enrolled patients. Keep domain specific information in these temporary datasets, eg. AE relationship, AE outcome, .. for a pair that is on AE dataset; Lab value, lab units, lab ranges,… for a pair that is on Laboratory dataset… and so on for as many pairs that exist in the database. Based on study teams feedback regarding relevance (or what is critical) to data review, the number of pairs (dataset, date variable) may be reduced.
3. Stack all these temporary datasets. Note that few variables are common in each of temporary datasets like patient ID, visit label, visit number and date of topic. For domain specific variables, in order to be stackable, they are renamed to ‘_1’, ‘_2’, ‘_3’, ….
4. Sort by patient ID and date of topic. Export final dataset to MS Excel.

For item 1 above, collect all date related variables from the entire study database. This can be done programmatically by searching in SASHELP.VCOLUMN (standard SAS view) in the data library of the study/protocol. Fetch all variables that have (or contain) ‘Date’ in label or in format or in variable name itself. Typically in a clinical database, date related variables end with ‘DT’ like ‘BRTHDT’, ‘VISDT’, ‘ASSESDT’, ‘DEATHDT’ so that can also be a search criteria. Any such list has to be manually reviewed to exclude certain variables that do not describe events or findings about the patient (clinical database frequently includes some EDC system related variables that do not essentially describe an event or a collection from patients). Further it is possible that certain patient related dates are not

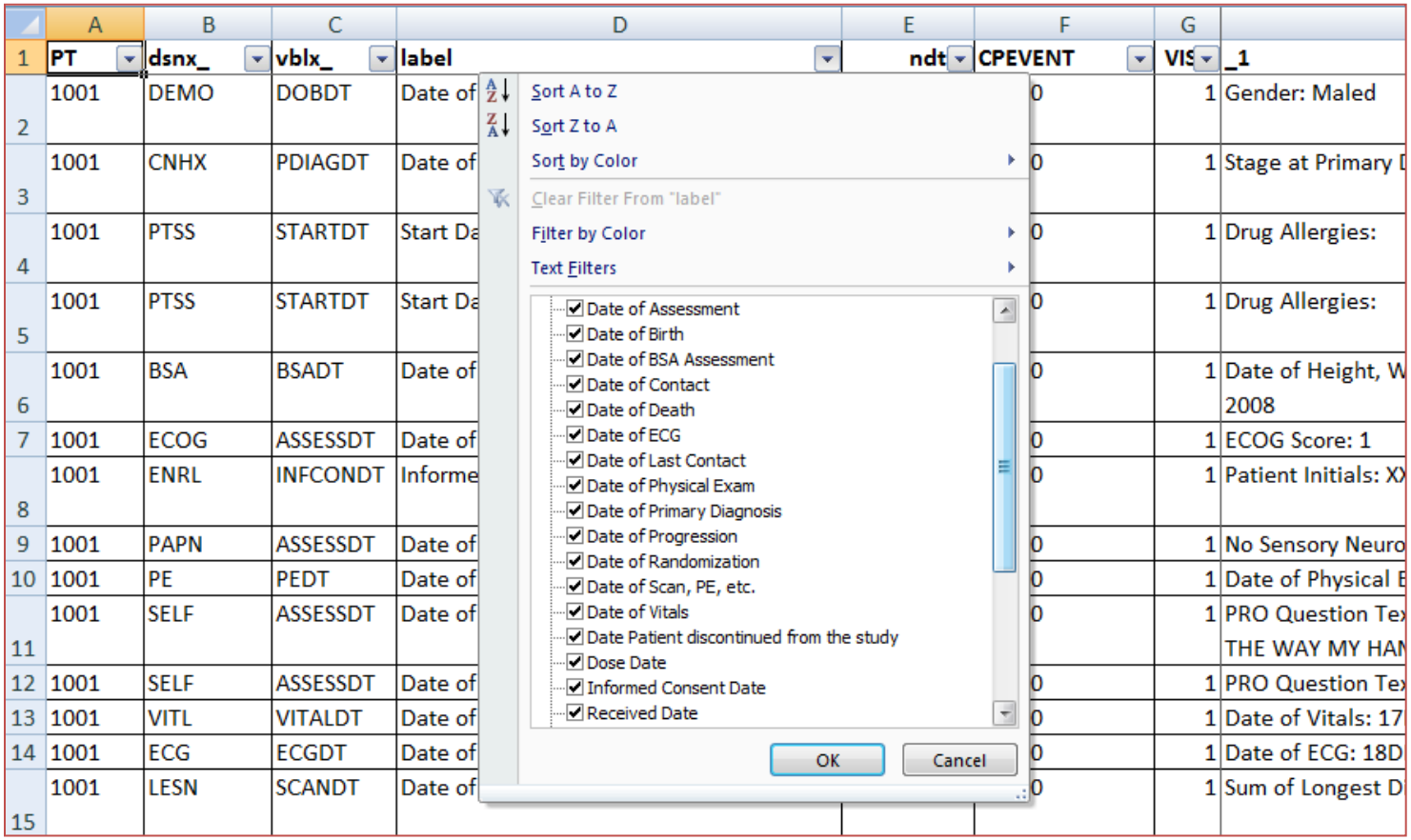
captured by any of above search criteria because the variable is named differently or out of convention and/or the label, format doesn't have 'Date' in it. Such dates should be caught by manual review of the database, and date variable(s) and respective dataset names be added to the above list.

Example of a search query to fetch all date related variables:

```
PROC SQL;
       create table DTVRS as
       select memname, name, type, format, label
       from SASHELP.VCOLUMN
       where libname='DB'
           & (name like '%DT%'              |
              name like '%DATE%'            |
              upcase(label) contains 'DATE' |
              upcase(format) contains 'DATE')
           & memname in (select memname  /*This subquery is to look only in those datasets with patient ID as one of the
variables.*/
                         from SASHELP.VCOLUMN
                         where upcase(name)='PT'
                             & libname='DB')
/* as result of manual review, exclude certain variables that are deemed not helpful for review */
           & memname ^ in ('PKCL'
                           'PKIN')
           & name ^like '%_YYYY'
           & name ^like '%_DD'
           & name ^like '%_INT'
           & name ^like '%_MM'
/*These variables have 'DT' or 'DATE' in their names or formats but since they do not describe any clinical or study
related event, they are excluded.*/
           & upcase(name) ^in ('FMFRSTDT'
                               'FMLASTDT'
                               'CMDTUNK'
                               'CMDTUNKN'
                               'CREATED'
                               'UPDATED')
       order by name, memname;
quit;
```

Below image shows an example set of date variables.

Image 8:



	A	B	C	D	E	F	G	
1	PT	dsnx_	vblx_	label	ndt	CPEVENT	VIS	_1
2	1001	DEMO	DOBDT	Date of		0	1	Gender: Maled
3	1001	CNHX	PDIAGDT	Date of		0	1	Stage at Primary D
4	1001	PTSS	STARTDT	Start Da		0	1	Drug Allergies:
5	1001	PTSS	STARTDT	Start Da		0	1	Drug Allergies:
6	1001	BSA	BSADT	Date of		0	1	Date of Height, W 2008
7	1001	ECOG	ASSESSDT	Date of		0	1	ECOG Score: 1
8	1001	ENRL	INFCONDT	Informe		0	1	Patient Initials: XX
9	1001	PAPN	ASSESSDT	Date of		0	1	No Sensory Neuro
10	1001	PE	PEDT	Date of		0	1	Date of Physical E
11	1001	SELF	ASSESSDT	Date of		0	1	PRO Question Tex THE WAY MY HAN
12	1001	SELF	ASSESSDT	Date of		0	1	PRO Question Tex
13	1001	VITL	VITALDT	Date of		0	1	Date of Vitals: 17
14	1001	ECG	ECGDT	Date of		0	1	Date of ECG: 18D
15	1001	LESN	SCANDT	Date of		0	1	Sum of Longest D

For items 2 & 3:

In below macro call, `&ds` is for dataset name from the study data library, `&var` is the variable name for date of topic, `&vtyp` to distinguish between charater or numeric variable type. Note that `dsnx_` is shown in above Displays as column B and `vblx_` under column C. `&&&ds._rcols` resolves to _1, _2, .. which is a list of variables of interest from the respective domain, shown above in Displays as columns _1, _2, …These variables have been pre-derived in {variable label: value} format in `&DS.3_` datasets.

```
%macro getcdt(/*dataset*/ ds, /*variable*/ var, /*type – char or num*/ vtyp);
      PROC SQL noprint;
            create table TMP_&DS._&VAR as
            select distinct "&ds" as dsnx_ length=15,
                           "&var" as vblx_ length=15, pt, &var as &vtyp %if &vtyp=cdtx_ %then length=15;,
                            cpevent,  visit, &&&ds._rcols
            from &DS.3_
            where &var is not missing
                & pt in (select pt
                         from LIST_OF_PATIENTS);
      quit;

      data ALLCDT; /*Append/Stack all temporary datasets into one.*/
            %if &vtyp=cdtx_ %then length cdtx_ $15;;
            set ALLCDT
                TMP_&DS._&VAR;
      run;
%mend getcdt;

DATA ALLCDT; /*Initialize a null dataset – no rows, no columns*/
      if 0;
run;

DATA _null_;
      set DTVRS; /*Dataset DTVRS is collection of date variables derived in step 1.*/
      length cstrg $250;
           if type='char' then cstrg='%getcdt('||compress(memname)||', '||compress(name)||', '||'cdtx_'||');';
      else if type='num'  then cstrg='%getcdt('||compress(memname)||', '||compress(name)||', '||'ndtx_'||');';
      call execute (cstrg);
run;
```

For item 4:

Final SAS dataset after sorting by patient ID and date of topic is then output to MS Excel using ODS.

```
ods listing close;
ods msoffice2k file="..path..\allcdts.xls" style=styles.minimal;
PROC PRINT data=allcdts noobs width=min;
run;
ods msoffice2k close;
ods listing;
```

PROSPOSED ENHANCEMENTS

Important events like AEs of Special Interest, Deaths or certain efficacy criteria (eg. Responder?) can be programmatically color coded (called Traffic Lighting) in MS Excel to draw reviewer's attention.

TRAINING

Reviewers (Physician, Scientist, Statistician, Project Leadership, ...) are accustomed to using traditional listing outputs like patient profiles and other standard format listings. A training to understand and use the single-output listing can help overcome the resistance of reviewers to move away from traditional outputs. Potentially significant efficiencies (efficiencies that translate to saving of time, effort and money) can be realized if the single-output listing proposed in this paper can meet data review needs that require as many as 25 to 50 outputs.

ACKNOWLEDGMENTS

Thanks to Jian Hua Huang and Shankar Srinivasan for reviewing the paper and providing additional insight.

SOFTWARE ACKNOWLEDGEMENTS

SAS 9.2, Microsoft Office 2007

CONTACT

For your comments and questions, contact the author at: hardikjp@inbox.com