PharmaSUG 2020 - Paper EP-99

Color Data Listings and Color Patient Profiles

Charley Z. Wu, Atara Biotherapeutics

1 ABSTRACT:

During clinical trials, there are frequent datacuts for safety data review, interim data analysis, CSRs. Data reviewers usually need to review the data carefully to ensure data accuracy and integrity. They frequently complain that they have already reviewed the same data many times before, and they don't like to review the same data again. They would rather pay more attention to new and updated data. However, most of data listings/patient profiles/reports cannot tell what are new data and what are old data.

To solve this issue, we developed color data listings and color patient profiles. The idea is that we can set the first datacut as a benchmark, all future data changes are then highlighted with different colors. For example, updated data are colored pink. New data are colored green. Deleted data are colored grey. Unchanged data are not colored. By doing so, reviewers can easily identify any data changes since previous datacut.

Though we just implemented this in regular data listings and patient profiles, we already got very positive feedback from data reviewers. It usually takes them 1-2 weeks to finish the review of all data listings and patient profiles. They can now finish data review in a couple of days. It also makes data review an enjoyable process as colored data changes pop up to reviewer's eyes.

2. INTRODUCTION:

In the life cycle of clinical trials, there are frequent datacuts for safety data review, interim data analysis, conference presentations, CSRs, etc. Medical Monitors, Statisticians, Clinical Data Managers, Pharmacovigilance will need to review the data carefully to ensure data accuracy and integrity. These functions frequently complain that they have already reviewed the same data many times before, and they don't like to review the same data repeatedly. They prefer to see new and updated data only. However, most of data listings/patient profiles/reports cannot tell what are new data and what are old data.

To solve this issue, we developed color data listings and color patient profiles. The idea is that we can set the first datacut as benchmark, all future data changes are then highlighted with different colors. For example, update data are colored pink. New data are colored green. Deleted data are colored grey. Unchanged data are not colored. By doing so, reviewers can easily identify any new, updated, or deleted data since last datacut.

3. FOLDER AND DATA STRUCTURE SETUP

In order to produce color data listings correctly and robustly, proper data and folder structures should be setup first.

- **Benchmark folder**: This is the initial folder that future data will be compared to. This is usually the first datacut of a study which has all SAS datasets. It can be anywhere on the SAS server. For example:

F:\ Project1\Study999_100\Datasets\Snapshots\20180312\

- **Current folder**: This is for current SAS datasets that will be compared with benchmark folder above. For example:

  F:\Project1\Study999_100\Datasets\Current\

- **Output folder**: This is the where color data listings will be created. They can be in PDF, XLS, or RTF formats. For example:

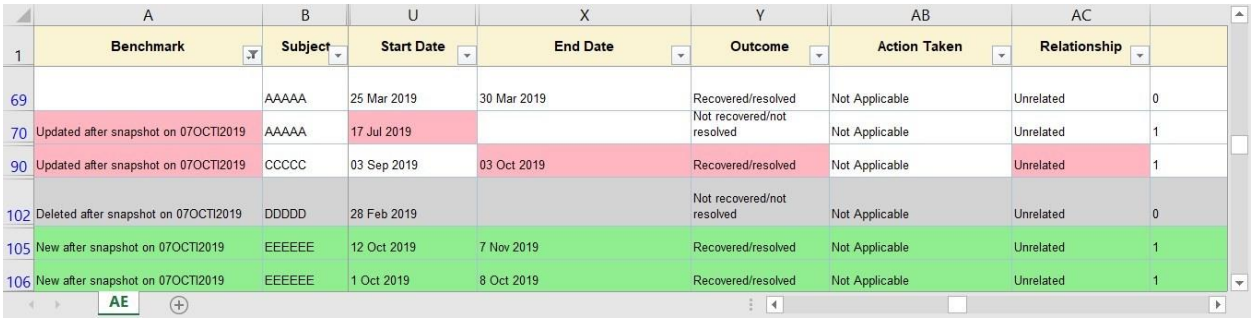
  F:\Project1\Study999_100\Output\Color_Listings\

- **SAS program folder:** This is where SAS programs reside.
- **Temp SAS dataset folder:** In order to produce color data listings, a lot of data processing is required. During the process, many temp datasets are created. We keep two set of datasets produced for color listings.

  - **Compare Folder:** This folder has the final datasets which will produce color data listings. For example:

    F:\ Project1\Study999_100\Datasets\Compare\

    The **data structure** of these datasets is identical to that of source datasets with one additional variable **Benchmark**. This **Benchmark** variable identifies whether a record is new, updated, or deleted. For example:



	A	B	U	X	Y	AB	AC	
1	Benchmark	Subject	Start Date	End Date	Outcome	Action Taken	Relationship	
69		AAAAA	25 Mar 2019	30 Mar 2019	Recovered/resolved	Not Applicable	Unrelated	0
70	Updated after snapshot on 07OCTI2019	AAAAA	17 Jul 2019		Not recovered/not resolved	Not Applicable	Unrelated	1
90	Updated after snapshot on 07OCTI2019	CCCCC	03 Sep 2019	03 Oct 2019	Recovered/resolved	Not Applicable	Unrelated	1
102	Deleted after snapshot on 07OCTI2019	DDDDD	28 Feb 2019		Not recovered/not resolved	Not Applicable	Unrelated	0
105	New after snapshot on 07OCTI2019	EEEEEE	12 Oct 2019	7 Nov 2019	Recovered/resolved	Not Applicable	Unrelated	1
106	New after snapshot on 07OCTI2019	EEEEEE	1 Oct 2019	8 Oct 2019	Recovered/resolved	Not Applicable	Unrelated	1

AE

  - **Diff Folder:** If a data record is new or deleted, it impacts the whole record. So the whole record will colored. This process is relatively easy to achieve. However, if a data record is updated and only a few variables/fields are updated, coloring these variables/fields from the rest of non-changed fields is quite challenging.

    In order to achieve this, we need to have another folder. This folder will identify all individual fields that are updated. For example:

    F:\ Project1\Study999_100\Datasets \Diff\

    The **data structure** of these datasets is as follows. It is an output from PROC COMPARE. It is used to identify base (Benchmark Dataset), compare (Current Dataset), and differences

(DIFF). This dataset is used to color specific fields that have been updated. The example below shows the same AE (Rhinorrhea) is coded to a new HLGT.

VIEWTABLE: Crodatau.Ae_updated_paper

	Type of Observation	Observation Number	Subject name or identifier	eCRF page name	Adverse Event	Adverse EventHLGT	Adverse EventHLGT_CODE
1	BASE	9	AAAAA	Adverse Events	Rhinorrhea		
2	COMPARE	9	AAAAA	Adverse Events	Rhinorrhea	Respiratory tract signs and symptoms	10079101
3	DIF	9	AAAAA			XXXXXXXXXX XXXXX XXXXX XXX XX	XXXXXXXX

4. PROCESS

- Define Benchmark Folder: Can be any datacut folder
- Define Current Folder: Can be any folder which will be compared with Benchmark Folder
- Compare datasets in the current folder with that in the benchmark folder. This is achieved by a macro:

  %find_ds_difference_rave (dsin=AE, order_by= %str(Subject, RecordId), idstm=%str(Subject RecordId));

  This macro will
  - Identify any new records
  - Identify any deleted records
  - Identify any updated records. Then use PROC COMPARE to identify what variables are updated
  - Combine new/updated/deleted/un-changed records together and output the datasets to the COMPARE folder
  - Output PROC COMPARE output to DIFF folder

- Generate colored data listings/Patient profiles. This is achieved by a macro:

  %dataset_difference_color (dsin=AE, order_by= %str(Subject, RecordId), idstm=%str(Subject RecordId));

  This macro will do:
  - If a record is new, color the whole record **green** with PROC REPORT style formatting
  - If a record is deleted, color the whole record **grey** with PROC REPORT style formatting

- If a record is updated, go through each variable and color the updated variable **pink** with PROC REPORT style formatting. This step will need to use the datasets in the DIFF folder. This is the most challenging step.
- If a record is not changed, output it without any formatting.

5. DISCUSSIONS

Data review is always a tedious process for Medical Monitors, Biostat, and CDM, especially when one is responsible for multiple studies. Yet, data review is a critical part in clinical trials. Data review meetings are routinely scheduled in the life cycle of clinical trials. Data listings, patient profiles, CSRs are usually cumulative. However, data reviewers do not like to review the same data they had reviewed before. Rather, they prefer to review data incrementally. That is to review any new/updated/deleted data. Certainly, they also want to see the whole data whenever needed.

Color data listing discussed here is a solution to the challenge. We create data listings, patient profiles with complete datasets as usual. But we go one more step to color any new/updated/deleted data in the listings. So it is very easy for data reviewers to know what data need more attention.

Though we usually use the first datacut as benchmark, it can be changed to any subsequent datacut as needed. Furthermore, we can use this process to identify any data differences between data transfers, patient profiles.

So far, we only implemented this on data listings and patient profiles. But if people are curious to know any data differences between different versions of CSRs, DSURs, this process can be used to color-code these reports as well.

After we implemented this process, we got overwhelming positive feedback from data reviewers. It usually takes them a couple of weeks to finish whole data reports review. Now they can finish the data review in a couple of days.

6. Sample Output

6.1. Sample Color Data Listings

	A	B	G	J	Q	U	X	Y	AB	AC
1	Benchmark	Subject	eCRF Page	Adverse Event	Adverse EventPT	Start Date	End Date	Outcome	Action Taken	Relationship
71		CCCCC	Adverse Events	Mechanical Fall	Fall	16 Feb 2019	16 Feb 2019	Recovered/resolved	Not Applicable	Unrelated
90	Updated after snapshot on 07OCTI2019	CCCCC	Adverse Events	Contact dermatitis right wrist	Dermatitis contact	03 Sep 2019	03 Oct 2019	Recovered/resolved	Not Applicable	Unrelated
91		CCCCC	Adverse Events	Urinary frequency	Pollakiuria	24 Jun 2019		Recovering/resolving	Dose Not Changed	Possibly Related
102	Deleted after snapshot on 07OCTI2019	DDDDD	Adverse Events	Possible pelvic tumour (most likely benign). No further information available		28 Feb 2019		Not recovered/not resolved	Not Applicable	Unrelated
103		EEEEEE	Adverse Events	Mechanical fall	Fall	27 Jul 2019	27 Jul 2019	Recovered/resolved	Dose Not Changed	Unrelated
104		EEEEEE	Adverse Events	Mechanical fall	Fall	26 Aug 2019	26 Aug 2019	Recovered/resolved	Not Applicable	Unrelated
105	New after snapshot on 07OCTI2019	EEEEEE	Adverse Events	Diabetes	Diabetes mellitus	12 Oct 2019	7 Nov 2019	Recovered/resolved	Not Applicable	Unrelated
106	New after snapshot on 07OCTI2019	EEEEEE	Adverse Events	Redness in foot secondary to insect bite	Arthropod bite	1 Oct 2019	8 Oct 2019	Recovered/resolved	Not Applicable	Unrelated
107		MMMMM	Adverse Events	Nausea	Nausea	28 Nov 2018	28 Nov 2018	Recovered/resolved	Not Applicable	Unrelated
114		NNNNNN	Adverse Events	Post-infusion weakness	Asthenia	01 May 2019	02 May 2019	Recovered/resolved	Dose Not Changed	Related
115	Updated after snapshot on 07OCTI2019	FFFFFF	Adverse Events	Worsening weakness	Asthenia	23 May 2019	20 Oct 2019	Not recovered/not resolved	Not Applicable	Unrelated
116	Updated after snapshot on 07OCTI2019	FFFFFF	Adverse Events	MS relapse	Multiple sclerosis relapse	15 Jul 2019	25 Sep 2019	Recovered/resolved	Not Applicable	Unrelated
117	New after snapshot on 07OCTI2019	FFFFF	Adverse Events	Worsening of existing neurological symptoms		16 Dec 2019		Not recovered/not resolved	Not Applicable	Unrelated
118		QQQQQ	Adverse Events	Common cold	Nasopharyngitis	24 Aug 2019	09 Sep 2019	Recovered/resolved	Not Applicable	Unrelated
119		QQQQQ	Adverse Events	Chest pain (musculoskeletal)	Musculoskeletal chest pain	30 Aug 2019	09 Sep 2019	Recovered/resolved	Not Applicable	Unrelated
120	New after snapshot on 07OCTI2019	GGGGG	Adverse Events	Acute Bronchitis	Bronchitis	28 Oct 2019	01 Nov 2019	Recovered/resolved	Not Applicable	Unrelated
121		GGGGG	Adverse Events	Sweating	Hyperhidrosis	15 Jan 2019		Not recovered/not resolved	Dose Not Changed	Unrelated
122		GGGGG	Adverse Events	Sunburn	Sunburn	26 Jan 2019	02 Feb 2019	Recovered/resolved	Not Applicable	Unrelated

AE

6.2. Sample Color Patient Profile

Concomitant Medications

Benchmark	Seq	Days After First Dose	Medication	ATC	Indication	Route	Start Date	End Date	Dose	Unit
	1	-277	Ampyra	NERVOUS SYSTEM	Multiple Sclerosis, Gait Disturbance	Oral	16 AUG 2018	Yes	10	mg
	2	-1622	cholecalciferol, vitamin D3, (VITAMIN D3)	ALIMENTARY TRACT AND METABOLISM	vitamin D deficiency	Oral	10 DEC 2014	Yes	2000	IU
	3	-2296	baclofen	MUSCULO-SKELETAL SYSTEM	Multiple Sclerosis, spasticity	Oral	04 FEB 2013	Yes	10	mg
	4	-1069	CALCIUM CITRATE ORAL	ALIMENTARY TRACT AND METABOLISM	supplement	Oral	15 JUN 2016	Yes	1	tab
Updated after snapshot on 07OCTI2019	5	-2257	OCUVITE eye + multi (Lutein, Zeaxanthin, daily multivitamin)	SENSORY ORGANS	Visual Loss	Oral	15 MAR 2013	Yes	1	tab
	7	-2041	fluoxetine	NERVOUS SYSTEM	Depression	Oral	17 OCT 2013	24 Feb 2014	10	mg
	8	-1911	fluoxetine	NERVOUS SYSTEM	depression	Oral	24 FEB 2014	27 Aug 2014	20	mg
	9	-1727	fluoxetine	NERVOUS SYSTEM	depression	Oral	27 AUG 2014	12 Jan 2015	10	mg
	10	-1589	fluoxetine	NERVOUS SYSTEM	depression	Oral	12 JAN 2015	08 Jul 2015	20	mg
	11	-1412	fluoxetine	NERVOUS SYSTEM	depression	Oral	08 JUL 2015	09 Jun 2016	10	mg
	12	-2386	fluoxetine	NERVOUS SYSTEM	depression	Oral	06 NOV 2012	17 Oct 2013	20	mg
	13	-2282	cyanocobalamin (vitamin b-12)	BLOOD AND BLOOD FORMING ORGANS	supplement	Oral	18 FEB 2013	Yes	500	mcg
	14	-44	collagen	MUSCULO-SKELETAL SYSTEM	supplement	Oral	06 APR 2019	Yes	1	scoop
	15	-277	dalfampridine	NERVOUS SYSTEM	Multiple Sclerosis	Oral	16 AUG 2018	Yes	10	mg
	16	-2299	Ascorbic Acid	ALIMENTARY TRACT AND METABOLISM	Supplement	Oral	01 FEB 2013	07 Feb 2013	500	mg
Updated after snapshot on 07OCTI2019	17	-236	levocetirizine	RESPIRATORY SYSTEM	rhinitis	Oral	26 SEP 2018	Yes	5	mg
New after snapshot on 07OCTI2019	18	-431	acetaminophen	NERVOUS SYSTEM	Prophylactic antipyretic	Oral	15 MAR 2018	15 Mar 2018	1000	mg

7. CONTACT INFORMATION

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

Author Name: Charley Z. Wu
E-mail: charleyzwu@yahoo.com