Catalog

Jonathan Stiglingh

Given Dataset:

Row	USUBJID	EOSSTT	DCSREAS	CREATBL	GFREBL	DTHDT	DTHCAUS	EOSDT	TRTSDT
1	XYZ-001-001	DISCONTINUED	DEATH	1.1	95,00	2013/11/02	SEPSIS	2013/11/02	2013/08/10
2	XYZ-001-002	DISCONTINUED	DEATH	1.2	58.5	2015/03/20	SUDDEN CARDIAC DEATH	2015/03/20	2014/12/29
3	XYZ-001-003	COMPLETED	COMPLETED	0.8	105,00			2010/10/12	2010/07/11
4	XYZ-001-004	DISCONTINUED	DEATH	0.9	78.9	2015/04/03	RENAL FAILURE	2015/04/03	2015/01/31
5	XYZ-001-005	COMPLETED	COMPLETED	1.4	68.3			2015/01/23	2014/10/26
6	XYZ-001-006	COMPLETED	COMPLETED	1.3	93.5			2015/01/25	2014/10/02

USUBJID
Completed the listing to follow the order as given above: XYZ-001-007 etc.

EOSSTT
Generated random list using the three given below. Randomly removed and changed spelling of words to simulate errors to be found.
=UPPER(CHOOSE(RANDBETWEEN(1;3);M1;M2;M3))

  Discontinued
  Completed
  Unknown

DCSREAS
Complete death status by using the Completed and Discontinued selected option from EOSSTT to generate the status of subject. Randomly removed and changed spelling of words to simulate errors to be found.

=IF(C8="COMPLETED";"Completed";IF(C8="DISCONTINUED";"DEATH";""))

CREATBL
Generating random creatinine levels using the Formula below.

=RAND()+RANDBETWEEN(0;1) then =LEFT(E8;3) to only use the first 3 digits including the ,

GFREBL
Generating random GFRE (kidney function) values using the Formula below.

=RAND()*10+RANDBETWEEN(50;90) then =LEFT(F8;4)

DTHDT
Random Death Date generated using the DCSREAS column and the Formula below using dates 2013/01/01 and 2015/12/31.
=IF(D76="Death";RANDBETWEEN(M1;M2);"")

DTHCAUS
=IF(D8="DEATH";CHOOSE(RANDBETWEEN(1;9);M1;M2;M3;M4;M5;M6;M7;M8;M9);"")

STROKE

SEPSIS
SUDDEN CARDIAC DEATH
OTHER INFECTION
RENAL FAILURE
MALIGNANCY
HYPERKALEMIA
ACUTE MYOCARDIAL INFARCTION
ATHEROSCLEROTIC HEART DISEASE

EOSDT
Generating end of study date using death date and random date between 2010/10/12 and 2015/12/31.

=IF(D8="DEATH";G8;RANDBETWEEN(I4;M2))

TRTSDT
Generating fake start date using the below formula.

=I8-RANDBETWEEN(50;115)

Cleaning Activities:

USUBJID
Nothing required.

EOSSTT:
You can see by the following that there are a few changes that need to be made to the data.

* All words must be capitalized.
* Words must be completed that are abbreviated.
* Missing data to be updated according to completion instructions if provided. In this case

The FREQ Procedure

EOSSTT				
EOSSTT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Frequency Missing = 5				
COMPLETED	38	40.00	38	40.00
Completed	1	1.05	39	41.05
DISCONTINUED	45	47.37	84	88.42
Disc	3	3.16	87	91.58
Unk	7	7.37	94	98.95
Unknown	1	1.05	95	100.00

Cleaning the missing data

Row	USUBJID	EOSSTT	DCSREAS	CREATBL	GFREBL	DTHDT	DTHCAUS
17	XYZ-001-017		COMPLETED	.4	87.8	.	
33	XYZ-001-033		DEATH	1.4	69.2	07/20/2014	MALIGNANCY
71	XYZ-001-071		COMPLETED	1.3	63.1	.	
83	XYZ-001-083			1	82.4	01/30/2014	
92	XYZ-001-092		COMPLETED	1.6	56.1	.	

You can se that the logic behind this column is if the subject completed the study it would state complete. If the subject passed away or has a death date the it would indicate Discontinued.

In the above the following is relevant:

Subject XYZ-001-017, 071 & 92 – Subject completed the study so Completed is required.
Subject XYZ-001-033   – Subject has Death (DCSREAS) and a Death Date(DTHDT). Discontinued Applied.
Subject XYZ-001-083   – Subject has Death Date DTHDT only. Discontinued Applied and Death applied to DCSREAS as well to save time.

Same logic applied to rest of the dataset.

Instruction are that all Unknown are to be removed.

Finally the variable frequency looks are follows:
The FREQ Procedure

EOSSTT				
EOSSTT	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
COMPLETED	45	46.88	45	46.88
DISCONTINUED	51	53.13	96	100.00

DCSREAS
After cleaning the missing values there is nothing that needs to be updated here.

The FREQ Procedure

DCSREAS				
DCSREAS	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
COMPLETED	45	46.88	45	46.88
DEATH	51	53.13	96	100.00

CREATBL
You will notice that there at are two instances where the CREATBL is 0 – as per instruction these values have been confirmed by sponsor to be specific values. Updating as per subject ID.

The FREQ Procedure

CREATBL				
CREATBL	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
.1	6	6.25	6	6.25
.2	5	5.21	11	11.46
.3	1	1.04	12	12.50
.4	7	7.29	19	19.79
.5	6	6.25	25	26.04
.6	5	5.21	30	31.25
.7	4	4.17	34	35.42
.8	10	10.42	44	45.83
.9	4	4.17	48	50.00
0	2	2.08	50	52.08
0.8	1	1.04	51	53.13
0.9	1	1.04	52	54.17
1	2	2.08	54	56.25

CREATBL				
CREATBL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.1	4	4.17	58	60.42
1.2	2	2.08	60	62.50
1.3	11	11.46	71	73.96
1.4	6	6.25	77	80.21
1.5	7	7.29	84	87.50
1.6	2	2.08	86	89.58
1.7	3	3.13	89	92.71
1.8	3	3.13	92	95.83
1.9	4	4.17	96	100.00

GFREBL
Correct the value per subject to 91.1 and 63.8.

Row ▲	USUBJID	EOSSTT	DCSREAS	CREATBL	GFREBL
1	80 XYZ-001-080	DISCONTINUED	DEATH	1.50	911.00
2	97 XYZ-001-097	COMPLETED	COMPLETED	1.30	638.00

DTHDT
To be decided

DTHCAUS
To be decided

EOSDT
To be decided

TRTSDT
To be decided

Jonathan Stiglingh

```sas
 1
 2
 3
 4  libname Death '/folders/myfolders/Portfolio3';
 5
 6  %let path=/folders/myfolders/Portfolio3/;
 7  filename N01 "&path\Dataset.xlsx";
 8
 9  /************* Importing the Dataset *************/
10  PROC IMPORT DATAFILE=N01
11      REPLACE
12      DBMS=XLSX
13      OUT=DEATH.Dataset;
14      SHEET="Data";
15      GETNAMES=YES;
16  RUN;
17
18
19
20  /************* Cleaning the Dataset *************/
21
22  data Death.Clean;
23      set DEATH.Dataset (drop=K L);
24
25  /************* EOSSTT variable *************/
26      EOSSTT=UPCASE(EOSSTT);
27      if EOSSTT='DISC' then EOSSTT='DISCONTINUED';
28      else if EOSSTT = 'UNK' then EOSSTT='UNKNOWN';
29
30      if DCSREAS = 'COMPLETED' then EOSSTT='COMPLETED';
31      else if DCSREAS = 'DEATH' then EOSSTT='DISCONTINUED';
32      else if DTHDT ^=. then EOSSTT='DISCONTINUED';
33
34
35  /************* DCSREAS variable *************/
36      if DTHDT ^=. then DCSREAS='DEATH';
37
38  /************* CREATBL variable *************/
39  /**** updating as per single observation ******/
40      if USUBJID='XYZ-001-008' then CREATBL=1.1;
41      if USUBJID='XYZ-001-028' then CREATBL=1.3;
42
43
44  /************* GFREBL variable *************/
45      if USUBJID='XYZ-001-080' then GFREBL=91.1;
46      if USUBJID='XYZ-001-097' then GFREBL=63.8;
47
48  /************* DTHDT variable *************/
49  /* No Change */
50
51  /************* DTHCAUS variable *************/
52  /* No Change */
53
54  /************* EOSDT variable *************/
55  /* No Change */
56
```

Portfolio Project 3

Jonathan Stiglingh

```sas
57  /************** TRTSDT variable **************/
58  /* No Change */
59
60  Label    USUBJID = "Unique Subject Identifier"
61           EOSSTT  = "End of Study Status"
62           DCSREAS = "Reason for Discontinuation from Study"
63           CREATBL = "Baseline Creatinine (mg/dL)"
64           GFREBL  = "Baseline GFRE (mL/min/1.73m2)"
65           DTHDT   = "Date of Death"
66           DTHCAUS = "Cause of Death"
67           EOSDT   = "End of Study Date"
68           TRTSDT  = "Date of First Exposure to Treatment";
69
70  run;
71
72  /************** Remove all unknown observations **************/
73  data Death.Clean;
74      modify death.clean;
75      if find(EOSSTT,'UNKNOWN') then remove;
76  run;
77
78
79
80
81  /************** Create the Listing **************/
82  %let outpath=/folders/myfolders/Portfolio3/;
83
84  Proc Sort data=Death.Clean;
85      by USUBJID;
86  run;
87
88  proc export data=demo.needs_review
89      outfile="&outpath\Death_Report.xls"
90      dbms=xls
91      replace;
92  run;
93
94
95
96  /************** TESTING DATASET **************/
97
98  /* data temp; */
99  /*  set Death.Clean; */
100 /*      where GFREBL>110; */
101 /* run; */
102 /*   */
103 /* PROC SORT DATA=death.clean;BY GFREBL; RUN; */
```

Portfolio Project 3