Catalog

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);"")

Created a table that defines the cause of death as a number. A friend let me know that most studies he works on uses numbers and not actual terms.

1	STROKE
2	SEPSIS
3	SUDDEN CARDIAC DEATH
4	OTHER INFECTION
5	RENAL FAILURE
6	MALIGNANCY
7	HYPERKALEMIA
8	ACUTE MYOCARDIAL INFARCTION
9	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
Used the list to populate the Cause of Death

EOSDT
To be decided

TRTSDT
To be decided

All portfolio files can be found on my GITHUB portal https://github.com/JonathanStiglingh/Portfolio-Project-3 .

Thank you for taking the time to review this portfolio.

Reason for Discontinue

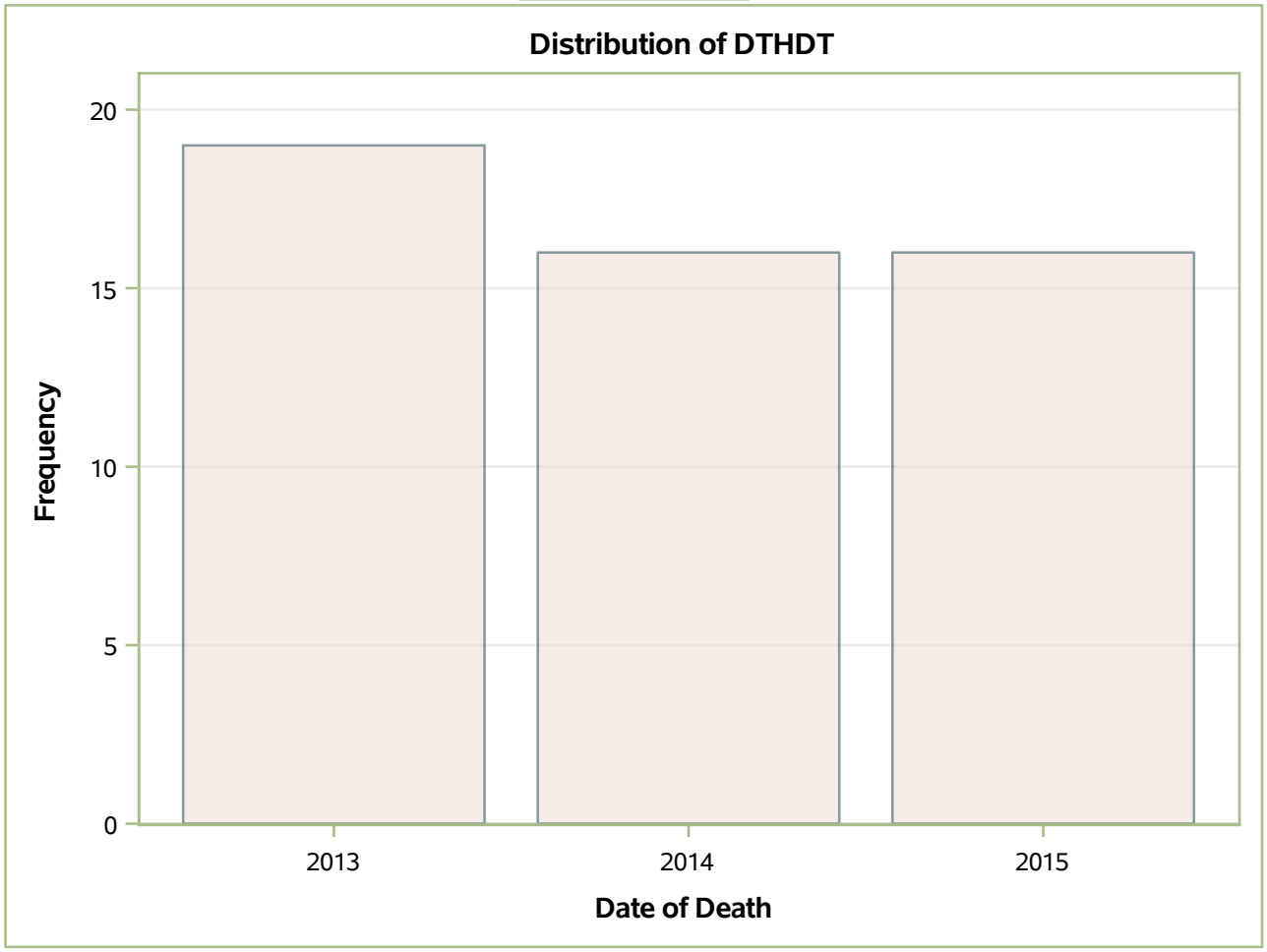
Portfolio Project 3

Reason for Discontinuation from Study	
DCSREAS	Frequency
DEATH	51
COMPLETED	45

Number of Deaths per Year

Portfolio Project 3

Date of Death	
DTHDT	Frequency
2013	19
2014	16
2015	16



Creatinine and GFRE for those with deaths reported

Cause of Death=ACUTE MYOCARDIAL INFARCTION

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.0	1.2
GFREBL	Baseline GFRE (mL/min/1.73m2)	76.9	75.2

Cause of Death=ATHEROSCLEROTIC HEART DISEASE

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.2	1.2
GFREBL	Baseline GFRE (mL/min/1.73m2)	72.9	77.4

Cause of Death=CORONARY ARTERY DISEASE

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.1	1.1
GFREBL	Baseline GFRE (mL/min/1.73m2)	94.1	94.1

Cause of Death=CORONARY THROMBOSIS

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	0.1	0.1
GFREBL	Baseline GFRE (mL/min/1.73m2)	85.0	85.0

Cause of Death=HYPERKALEMIA

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.1	1.1
GFREBL	Baseline GFRE (mL/min/1.73m2)	82.3	80.1

Cause of Death=MALIGNANCY

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.2	1.2
GFREBL	Baseline GFRE (mL/min/1.73m2)	72.0	70.6

Cause of Death=OTHER INFECTION

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.0	0.8
GFREBL	Baseline GFRE (mL/min/1.73m2)	60.8	61.1

Cause of Death=RENAL FAILURE

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	1.1	1.1
GFREBL	Baseline GFRE (mL/min/1.73m2)	79.2	84.1

Cause of Death=SEPSIS

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	0.9	1.1
GFREBL	Baseline GFRE (mL/min/1.73m2)	78.4	77.5

Cause of Death=STROKE

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	0.6	0.5
GFREBL	Baseline GFRE (mL/min/1.73m2)	79.6	82.6

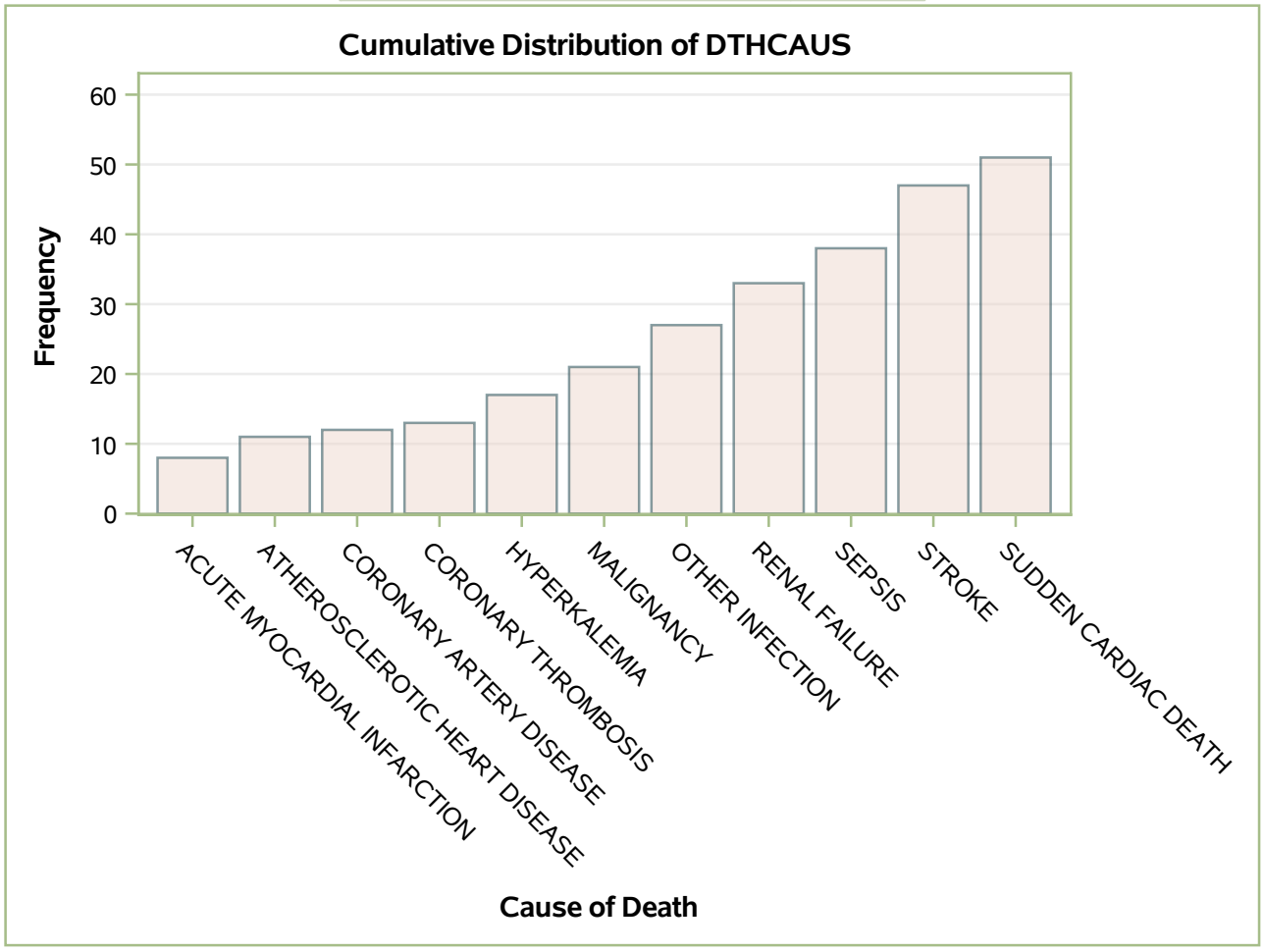
Creatinine and GFRE for those with deaths reported

Cause of Death=SUDDEN CARDIAC DEATH

Variable	Label	Mean	Median
CREATBL	Baseline Creatinine (mg/dL)	0.8	0.9
GFREBL	Baseline GFRE (mL/min/1.73m2)	77.3	76.7

Cummulative Cause of Death

Cause of Death	
DTHCAUS	Frequency
ACUTE MYOCARDIAL INFARCTION	8
ATHEROSCLEROTIC HEART DISEASE	3
CORONARY ARTERY DISEASE	1
CORONARY THROMBOSIS	1
HYPERKALEMIA	4
MALIGNANCY	4
OTHER INFECTION	6
RENAL FAILURE	6
SEPSIS	5
STROKE	9
SUDDEN CARDIAC DEATH	4



```
1
2  libname Death '/folders/myfolders/Portfolio3';
3
4  %let path=/folders/myfolders/Portfolio3/;
5  filename N01 "&path\Dataset.xlsx";
6
7  /************** Importing the Dataset **************/
8  PROC IMPORT DATAFILE=N01
9      REPLACE
10     DBMS=XLSX
11     OUT=DEATH.Dataset;
12     SHEET="Data";
13     GETNAMES=YES;
14 RUN;
15
16
17 /************** Cleaning the Dataset **************/
18
19 data Death.Clean;
20     set DEATH.Dataset (drop=Row K L);
21
22 /************** EOSSTT variable **************/
23     EOSSTT=UPCASE(EOSSTT);
24     if EOSSTT='DISC' then EOSSTT='DISCONTINUED';
25     else if EOSSTT = 'UNK' then EOSSTT='UNKNOWN';
26
27     if DCSREAS = 'COMPLETED' then EOSSTT='COMPLETED';
28     else if DCSREAS = 'DEATH' then EOSSTT='DISCONTINUED';
29     else if DTHDT ^=. then EOSSTT='DISCONTINUED';
30
31
32 /************** DCSREAS variable **************/
33     if DTHDT ^=. then DCSREAS='DEATH';
34
35 /************** CREATBL variable **************/
36 /**** updating as per single observation ******/
37     if USUBJID='XYZ-001-008' then CREATBL=1.1;
38     if USUBJID='XYZ-001-028' then CREATBL=1.3;
39
40 /************** GFREBL variable **************/
41     if USUBJID='XYZ-001-080' then GFREBL=91.1;
42     if USUBJID='XYZ-001-097' then GFREBL=63.8;
43
44 /************** DTHDT variable **************/
45 /* No Change */
46
47 /************** DTHCAUS variable **************/
48     DTHCAUS=strip(Upcase(DTHCAUS));
49
50 /************** EOSDT variable **************/
51 /* No Change */
52
53 /************** TRTSDT variable **************/
54 /* No Change */
55 run;
56
57 /************** Remove all unknown observations **************/
58 data Death.Clean;
59     modify Death.Clean;
60     if find(EOSSTT,'UNKNOWN') then remove;
61 run;
62
63 Data Coding;
64     set Death.Clean;
65     length DTHCAUS1 $50;
66 /************** DTHCAUS variable **************/
67 /******** CODING Cause of death term **********/
68
69         if DTHCAUS=1 then DTHCAUS1='ACUTE MYOCARDIAL INFARCTION';
70         if DTHCAUS=2 then DTHCAUS1='ATHEROSCLEROTIC HEART DISEASE';
71         if DTHCAUS=3 then DTHCAUS1='CORONARY ARTERY DISEASE';
72         if DTHCAUS=4 then DTHCAUS1='CORONARY THROMBOSIS';
73         if DTHCAUS=5 then DTHCAUS1='HYPERKALEMIA';
74         if DTHCAUS=6 then DTHCAUS1='MALIGNANCY';
75         if DTHCAUS=7 then DTHCAUS1='OTHER INFECTION';
76         if DTHCAUS=8 then DTHCAUS1='RENAL FAILURE';
77         if DTHCAUS=9 then DTHCAUS1='SEPSIS';
78         if DTHCAUS=10 then DTHCAUS1='STROKE';
79
80
81
```

```
        if DTHCAUS=11 then DTHCAUS1='SUDDEN CARDIAC DEATH';
    drop DTHCAUS; rename DTHCAUS1=DTHCAUS;
    format DTHCAUS1 $50.;
run;



/************** Adding Labels **************/
data Death.Final;
    set coding;
    Label   USUBJID = "Unique Subject Identifier"
            EOSSTT  = "End of Study Status"
            DCSREAS = "Reason for Discontinuation from Study"
            CREATBL = "Baseline Creatinine (mg/dL)"
            GFREBL  = "Baseline GFRE (mL/min/1.73m2)"
            DTHDT   = "Date of Death"
            DTHCAUS = "Cause of Death"
            EOSDT   = "End of Study Date"
            TRTSDT  = "Date of First Exposure to Treatment";

run;


/************** Create the Listing **************/
%let outpath=/folders/myfolders/Portfolio3/;
ods pdf file="&outpath\DeathReport.pdf" style=Meadow bookmarkgen=Yes;
ods noproctitle;
options nodate; /*removes the date*/
options label;

Title "Reason for Discontinue";
ods proclabel "Deaths vs Completion";
PROC FREQ DATA=Death.Final order=freq;
TABLE DCSREAS /nocum nofreq nopercent;
RUN;

Title "Number of Deaths per Year";
ods proclabel "Number of Deaths per year";
proc freq data=Death.Final;
    table DTHDT /nocum nopercent plots=freqplot; /*this can also be done using the Proc SGPLOTS step*/
    format DTHDT year.;
    where DTHDT is not missing;
run;



proc sort data=death.final;
by dthcaus;
run;

Title "Creatinine and GFRE for those with deaths reported";
ods proclabel "Creatinine and GFRE for those with deaths reported";
proc means data=death.final mean median maxdec=1;
var CREATBL GFREBL;
by dthcaus;
where dthcaus is not missing;
run;

Title "Cummulative Cause of Death";
ods proclabel "Cummulative Cause of Death";
proc freq data=Death.Final;
    table dthcaus /nocum nopercent plots= cumfreqplot; /*this can also be done using the Proc SGPLOTS step*/
    where dthcaus is not missing;
run;

ods pdf close;


/*Eport final to excel file*/

proc sort data=death.final;
by USUBJID ;
run;

proc export data=Death.Final
    outfile="&outpath\Death_Report.xls"
    dbms=xls
    replace;
run;
```

```
164
165
166 /************** TESTING DATASET **************/
167
168 /* data temp; */
169 /*  set Death.Clean; */
170 /*      where GEBEBL>110; */
171
```