Dataset Integrity Check for Features and Outcomes of 899 Patients with Drug-induced Liver Injury: The DILIN Prospective Study Chalasani et al.

Prepared by Dominick Parisi
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
11 August 2020

Contents

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) has established the Drug-Induced Liver Injury Network (DILIN) to collect and analyze cases of severe liver injury caused by prescription drugs, over-the-counter drugs, and alternative medicines, such as herbal products and supplements. The diagnosis of drug-induced liver injury relies on evidence linking the injury to a specific drug or agent. DILIN uses expert opinion to determine the degree of association between the implicated medication(s) and the liver injury.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the following DILIN data package. The specific datasets used for this review are listed in Table A below.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Chalasani et al. [1] in Gastroenterology in 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], “Demographic and Selected Clinical Features for the 899 Subjects Studied- Comparisons by Patterns of Drug-Induced Liver Injury”, Table A lists the variables that were used in the replication, and Table B compares the results calculated from the archived data files to the results published in Table 1.

6 Conclusions

Results showed minor discrepancies determined to be within variation expected limits. The NIDDK repository is confident that the DILIN data files to be distributed are a true copy of the study data.

7 References

[1] Chalasani N, Bonkovsky HL, Fontana R, et al. Features and Outcomes of 899 Patients With Drug-Induced Liver Injury: The DILIN Prospective Study. Gastroenterology. 2015;148(7):1340-52.e7. doi:10.1053/j.gastro.2015.03.006

Table A: Variables used to replicate Table 1: Demographic and Selected Clinical Features for the 899 Subjects Studied. Comparisons by Patterns of Drug-Induced Liver Injury

Table Variable	**dataset.variable**
Drug-Induced Liver Injury type	hxinj2.FLWRATIO
Sex	demog.SEX
Race	demog.WHITE, demog.BLACK, demog.ASIAN
BMI	pex1.HT, pex1.HTUN, pex1.WT, pex1.WTUN
Prior drug allergies	medhx2.ALLERGY
Immuno-allergic features	sign.SGNRES
Preexisting Liver Disease	livhx.MEDLIVHX
Diabetes mellitus	medhx1.PASTHX
Jaundice	sign.SGNRES
Liver biochemistries (DILI recognition), ALT	hxinj2.ALTVAL
Liver biochemistries (DILI recognition), AP	hxinj2.AKPVAL
Liver biochemistries (DILI recognition), total bilirubin	hxinj2.STBVAL
Liver biochemistries (DILI recognition), INR	hxinj2.INRRATE
Liver biochemistries (Peak Value), ALT	hxinj2.ALTVAL
Liver biochemistries (Peak Value), AP	hxinj2.AKPVAL
Liver biochemistries (Peak Value), total bilirubin	hxinj2.STBVAL
Liver biochemistries (Peak Value), INR	hxinj2.INRRATE
Peripheral Eosinophilia	eos.result_in_standard_units
Causality Assessment	judgmnt.ASSESS
Severity of Liver Injury	judgmnt.SEVERITY
Death	ptcompl.DEATHDT
Liver-related Death	ptcompl.NLLRDTH
Liver Transplantation	hxinj2.LIVTRX, medout.LIVTRANS
Chronic Dili	chronic.CINEXC1-CINEXC10

Table B: Comparison of values computed in integrity check to reference article Table 1 values.

	Entire Cohort (n=899) Manuscript	Entire Cohort (n=899) DSIC	Diff (n=0)	Hepatocellular (n=484) Manuscript	Hepatocellular (n=486) DSIC	Diff (n=2)
Age (years, mean [SD])	49 [17]	n/a	n/a	45 [17]	n/a	n/a
Females (%)	59	59	0	65	65	0
Self-reported race (%)						
• White	79	79	0	74	74.7	0.7
• Black or African-American	12	12	0	13.5	13.6	0.1
• Other/Multiracial	10	10	0	12	11.7	0.3
BMI (kg/m^2, mean [SD])	27 [6.5]	27 [6.5]	0 [0]	28 [6.8]	28 [6.8]	0 [0]
Prior drug allergies (%)	44	44	0	44	44	0
Immuno-allergic features (%)	15	15	0	12	12	0
Preexisting Liver Disease (%)	9.9	9.9	0	9.9	9.9	0
Diabetes mellitus (%)	25	26	1	23	25	2
Latency (days in median, IQR)	36 [19-88]	n/a	n/a	46 [22-104]	n/a	n/a
Jaundice (%)	70	70	0	65	65	0
Liver Biochemistries – DILI recognition						
• ALT (U/L, mean [SD])	825 [1105]	824 [1107]	1 [2]	1275 [1329]	1271 [1330]	4 [1]
• AP (U/L, mean [SD])	288 [254]	289 [257]	1 [3]	187 [110]	188 [111]	1 [1]
• Total bilirubin (mg/dl, mean [SD])	6.7 [6.6]	6.7 [6.7]	0 [0.1]	6.3 [6.6]	6.3 [6.6]	0 [0]
• INR	1.4 [1.0]	1.4 [1.0]	0 [0]	1.6 [1.2]	1.6 [1.2]	0 [0]
Liver Biochemistries – Peak values						
• ALT (U/L, mean [SD])	1008 [1221]	1003 [1217]	5 [4]	1510 [1431]	1498 [1426]	12 [5]
• AP (U/L, mean [SD])	406 [388]	406 [388]	0 [0]	271 [252]	272 [252]	1 [0]
• Total bilirubin (mg/dl, mean [SD])	13 [12]	13 [12]	0 [0]	12 [11.3]	12 [11.2]	0 [0.1]
• INR	1.7 [1.5]	1.7 [1.5]	0 [0]	1.8 [1.8]	1.8 [1.8]	0 [0]
Peripheral eosinophilia (>500/uL) (%)	11	11	0	7.2	7.4	0.2

	Entire Cohort (n=899) Manuscript	Entire Cohort (n=899) DSIC	Diff (n=0)	Hepatocellular (n=484) Manuscript	Hepatocellular (n=486) DSIC	Diff (n=2)
Improvement in liver tests – median days		n/a	n/a		n/a	n/a
• Peak ALT to below ULN	71			79		
• Peak AP to below ULN	96			48		
• Peak bilirubin to ≤ 1 mg/dL	70			66		
Causality Assessment (%)						
• Definite/Highly likely/Prob, n	235 / 466 / 198	239 / 466 / 188	4 / 0 / 10	120 / 255 / 109	123/ 253 / 107	3 / 2 / 2
• %	26 / 52 / 22	27 / 52 / 21	1 / 0 / 1	25 / 53 / 22	25 / 52 / 22	0 / 1 / 0
Severity of Liver Injury (%)		1 missing				
• Mild	24	24	0	29	29	0
• Moderate	21	21	0	15	16	1
• Moderate-hospitalized	29	29	0	26	26	0
• Severe	19	19	0	21	21	0
• Fatal	7	7	0	9	9	0
Death at any point (%)	6.2	6.3	0.2	5.4	5.4	0
• Percent Liver-related (%)	49	47	2	58	58	0
Liver Transplantation (%)	4	4	0	6.2	6.0	0.2
Chronic DILI (%)	17	17	0	13	13	0

	Cholestatic (n=210) Manuscript	Cholestatic (n=216) DSIC	Diff (n=6)	Mixed (n=205) Manuscript	Mixed (n=195) DSIC	Diff (n=10)
Age (years, mean [SD])	54 [16]	n/a	n/a	50 [17]	n/a	n/a
Females (%)	51	51	0	52	52	0
Self-reported race (%)						
• White	84	83	1	84	85	1
• Black or African-American	11	12	1	8	8	0
• Other/Multiracial	5	5	0	8	7	1
BMI (kg/m^2, mean [SD])	27 [6.3]	27 [6.2]	0 [0.1]	27 [6.1]	27 [6.2]	0 [0.1]
Prior drug allergies (%)	47	46	1	42	43	1
Immuno-allergic features (%)	17	17	0	21	20	1
Preexisting Liver Disease (%)	13.3	13.4	0.1	6.3	6.2	0.1
Diabetes mellitus (%)	33	34	1	20	21	1
Latency (days in median, IQR)	31 [16-72]	n/a	n/a	31 [18-51]	n/a	n/a
Jaundice (%)	78	79	1	75	74	1
Liver Biochemistries – DILI recognition						
• ALT (U/L, mean [SD])	202 [160]	203 [161]	1 [1]	379 [226]	380 [224]	1 [2]
• AP (U/L, mean [SD])	497 [371]	490 [370]	7 [1]	306 [206]	312 [217]	6 [11]
• Total bilirubin (mg/dl, mean [SD])	7.8 [7.3]	7.9 [7.3]	0.1 [0]	6.4 [6.0]	8.5 [22.7]	2.1 [16.7]
• INR	1.2 [0.5]	1.2 [0.5]	0 [0]	1.3 [0.6]	1.3 [0.6]	0 [0]
Liver Biochemistries – Peak values						
• ALT (U/L, mean [SD])	339 [445]	339 [441]	0 [4]	506 [439]	511 [441]	5 [2]
• AP (U/L, mean [SD])	682 [532]	678 [526]	4 [6]	440 [317]	438 [321]	2 [4]
• Total bilirubin (mg/dl, mean [SD])	15 [12.5]	15 [12.6]	0 [0.1]	13.3 [12.1]	13.1 [12.2]	0.2 [0.1]
• INR	1.6 [1.15]	1.6 [1.13]	0 [0.02]	1.5 [1.1]	1.5 [1.2]	0 [0.1]
Peripheral eosinophilia (>500/uL) (%)	14.6	14.7	0.1	15.8	15.6	0.2

	Cholestatic (n=210) Manuscript	Cholestatic (n=216) DSIC	Diff (n=6)	Mixed (n=205) Manuscript	Mixed (n=195) DSIC	Diff (n=10)
Improvement in liver tests – median days		n/a	n/a		n/a	n/a
• Peak ALT to below ULN	113			59		
• Peak AP to below ULN	183			90		
• Peak bilirubin to ≤ 1 mg/dL	77.5			74		
Causality Assessment (%)						
• Definite/Highly likely/Prob, n	47 / 109 / 54	47 / 118 / 48	0 / 9 / 6	68 / 102 / 35	68 / 95 / 32	0 / 7 / 3
• %	22 / 52 / 26	22 / 55 / 22	0 / 3 / 4	33 / 50 / 17	35 / 49 / 16	2 / 1 / 1
Severity of Liver Injury (%)						
• Mild	15	15	0	20	21	1
• Moderate	27	27	0	29	28	1
• Moderate-hospitalized	33	33	0	34	33	1
• Severe	21	22	1	13	13	0
• Fatal	4	4	0	4	5	1
Death at any point (%)	9	9	0	5.4	5.6	0.2
• Percent Liver-related (%)	56	50	6	18	18	0
Liver Transplantation (%)	2.9	1.4	1.5	0	0.5	0.5
Chronic DILI (%)	31	31	0	14	13	1

Attachment A: SAS Code

```
/***********/
/* Formats */
/***********/
%include '/prj/niddk/ims_analysis/DILIN_PRO/private_created_data/pro_formats.edit.sas';

PROC FORMAT;
  VALUE FLOWRATE
    1 = 'HEPATOCELLULAR'
    2 = 'CHOLESTATIC'
    3 = 'MIXED';
RUN;

/**********************/
/* Filename statements */
/**********************/
LIBNAME SASDATA '/prj/niddk/ims_analysis/DILIN/private_orig_data/';
LIBNAME SASDATA2 '/prj/niddk/ims_analysis/DILIN/private_orig_data/DILIN_InForm_Pro_Retro/data';

/*******************/
/* Import datasets */
/*******************/
DATA POPULATION;
  SET SASDATA.pop;
RUN;

PROC IMPORT
  DATAFILE='/prj/niddk/ims_analysis/DILIN/private_orig_data/eos.csv'
  DBMS=csv
  OUT=EOS (RENAME=DEIDENTIFIED_SUBJECT_NUMBER=DEIDNUM);
RUN;
```

```
/*************************************************************/
/* Prepare variables needed to replicate manuscript numbers */
/*************************************************************/
*** Peripheral eosinophilia;
PROC SORT DATA=POPULATION;
  BY DEIDNUM;
RUN;

DATA EOSINOPHIL (KEEP=UNIQUEID ABS_EOS EOS500N);
  MERGE EOS        (IN=INEOS WHERE=(VISIT_NUMBER IN(17598,17601)))
        POPULATION (IN=INPOP KEEP=DEIDNUM UNIQUEID);
  BY DEIDNUM;
  IF INEOS AND INPOP;

  RETAIN ABS_EOS;
  IF FIRST.DEIDNUM THEN ABS_EOS=0;

  IF RESULT_IN_STANDARD_UNITS>ABS_EOS THEN ABS_EOS=RESULT_IN_STANDARD_UNITS;

  IF LAST.DEIDNUM;

  IF ABS_EOS>500 THEN EOS500N=1;
  ELSE IF .<ABS_EOS<=500 THEN EOS500N=0;

  LABEL ABS_EOS='Absolute eosinophil count (/uL)'
        EOS500N='Eosinophil>500/ul (num)';
RUN;

/*
PROC MEANS DATA=EOSINOPHIL MEAN STD MEDIAN P25 P75 MIN MAX;
  VAR ABS_EOS;
RUN;

PROC FREQ DATA=EOSINOPHIL;
  TABLE EOS500N;
RUN;
*/
```

```
*** Jaundice;
PROC SORT DATA=SASDATA2.SIGN OUT=SIGN;
  BY UNIQUEID;
RUN;

DATA SIGNS (KEEP=UNIQUEID JAUNDICE FEVER RASH);
  SET SIGN;
  BY UNIQUEID;
  RETAIN JAUNDICE FEVER RASH;
  IF FIRST.UNIQUEID THEN DO;
    JAUNDICE=0;
    FEVER=0;
    RASH=0;
    END;
  IF SGNRES=1 THEN JAUNDICE=1;
  IF SGNRES=5 THEN FEVER=1;

  IF SGNRES=8 THEN RASH=1;
  IF LAST.UNIQUEID;
RUN;

*** Allergy;
proc sort data=sasdata2.medhx2 out=medhx2 nodupkey;
  by uniqueid;
run;

*** Liver transplant;
DATA HXINJ2;
  SET SASDATA2.HXINJ2;
RUN;

DATA MEDOUT;
  SET SASDATA2.MEDOUT;
RUN;

PROC SORT DATA=HXINJ2 NODUPKEY;
  BY UNIQUEID;
RUN;
```

```
PROC SORT DATA=MEDOUT NODUPKEY;
  BY UNIQUEID;
  WHERE VISIT='vsMONTH6';
RUN;

DATA LIVER_TRANSPLANT;
  MERGE HXINJ2 (IN=INHXINJ2 KEEP=UNIQUEID LIVTRX)
        MEDOUT (IN=INMEDOUT KEEP=UNIQUEID LIVTRANS);
  BY UNIQUEID;

  IN_HXINJ2=(INHXINJ2);
  IN_MEDOUT=(INMEDOUT);

  IF LIVTRX=1 OR LIVTRANS=1 THEN LIVER_TRANSPLANT=1;
  ELSE LIVER_TRANSPLANT=0;
RUN;

*** ATL, ALK, BILI, INR values;
PROC SORT DATA=SASDATA2.LABFLOW out=LABFLOW;
  BY UNIQUEID;
RUN;

data LABFLOW;
  merge labflow (in=in_lab)
        hxinj2  (in=in_hx keep=UNIQUEID ONSETDT);
  by UNIQUEID;
  IF IN_HX AND IN_LAB;
run;

PROC SORT DATA=LABFLOW;
  BY UNIQUEID FCLABDT SEQNO;
RUN;

DATA LAB_DILI_ONSET;
  SET LABFLOW;
  BY UNIQUEID FCLABDT SEQNO;
  WHERE FCLABDT=ONSETDT;

  IF LAST.UNIQUEID;
RUN;
```

```
*** Flow ratio;
data labflow_nomiss;
  set labflow;
  if FLWRATIO=. then delete;

proc sort data=labflow_nomiss;
  by UNIQUEID FCLABDT;

proc sort data=hxinj2;
  by UNIQUEID;

data hx_lab_prelim;
  merge hxinj2 (in=in_hx keep=UNIQUEID ONSETDT) labflow_nomiss(in=in_lab);
  by UNIQUEID;
  hx_ok=in_hx;
  lab_ok=in_lab;
  retain OUTPUT_DATE OUTPUT_FLAG EARLY_TO_ONSET;
  if first.UNIQUEID then do;
    OUTPUT_FLAG=0;
    OUTPUT_DATE=.;
    EARLY_TO_ONSET=ONSETDT-FCLABDT;
  end;
  if FCLABDT>=ONSETDT and OUTPUT_FLAG=0 then do;
    OUTPUT_DATE=FCLABDT;
    OUTPUT_FLAG=1;
  end;
  if OUTPUT_FLAG=1 and FCLABDT ne OUTPUT_DATE then do;
    OUTPUT_FLAG=2;
    OUTPUT_DATE=.;
  end;

data hx_lab_prelim_b;
  set hx_lab_prelim;
  if OUTPUT_FLAG=1 then output;

proc sort data=hx_lab_prelim_b;
  by UNIQUEID FLWRATIO;
```

```
data hx_lab(keep=UNIQUEID EARLY_TO_ONSET FLWRATIO);
  set hx_lab_prelim_b;
  by UNIQUEID;
  if last.UNIQUEID then output;


data hx_lab(drop=FLWRATIO);
  set hx_lab;
  if FLWRATIO>=5 then FLOW_NUM=1;
  else if 0<=FLWRATIO<=2 then FLOW_NUM=2;
  else if 2<FLWRATIO<5 then FLOW_NUM=3;
run;

*** Peak ALT, ALK, BILI, INR;
PROC SORT DATA=LABFLOW;
  BY UNIQUEID FCLABDT;
RUN;

PROC SORT DATA=SASDATA2.STDLAB2 OUT=STDLAB2;
  BY UNIQUEID;
RUN;

data labflow_PEAKS (keep=UNIQUEID ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK);
  set LABFLOW (WHERE=(VISIT IN('vsPREISV','vsBASE','vsMONTH6') AND FCLABDT>=ONSETDT) KEEP=UNIQUEID VISIT FCLABDT
ONSETDT ALTVAL AKPVAL STBVAL STBUNIT INRRATE);
  by UNIQUEID;
  length ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK 8.;
  retain ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK;
  if first.UNIQUEID then do;
    ALT_PEAK=.;
    ALK_PEAK=.;
    BILI_PEAK=.;
    INR_PEAK=.;
  end;
  ALT_PEAK=max(ALTVAL,ALT_PEAK);
  ALK_PEAK=max(AKPVAL,ALK_PEAK);
  IF STBUNIT^=4 THEN BILI_PEAK=max(STBVAL,BILI_PEAK);
  INR_PEAK=max(INRRATE,INR_PEAK);

  if last.UNIQUEID then output;
```

```
data stdlab2_PEAKS (keep=UNIQUEID ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK LABUNT17);
  set stdlab2 (WHERE=(VISIT IN('vsPREISV','vsBASE','vsMONTH6')) KEEP=UNIQUEID VISIT LABVAL15 LABVAL17 LABUNT17 LABVAL20
LABVAL21);
  by UNIQUEID;
  length ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK 8.;
  retain ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK;
  if first.UNIQUEID then do;
    ALT_PEAK=.;
    ALK_PEAK=.;
    BILI_PEAK=.;
    INR_PEAK=.;
  end;
  ALT_PEAK=max(LABVAL20,ALT_PEAK);
  ALK_PEAK=max(LABVAL21,ALK_PEAK);
  BILI_PEAK=max(LABVAL17,BILI_PEAK);
  INR_PEAK=max(LABVAL15,INR_PEAK);

  if last.UNIQUEID then output;
RUN;

DATA PEAKS;
  MERGE LABFLOW_PEAKS (IN=INLABFLOW RENAME=(ALT_PEAK=ALT_PEAK_LABFLOW ALK_PEAK=ALK_PEAK_LABFLOW
BILI_PEAK=BILI_PEAK_LABFLOW INR_PEAK=INR_PEAK_LABFLOW))
        STDLAB2_PEAKS (IN=INSTDLAB2 RENAME=(ALT_PEAK=ALT_PEAK_STDLAB2 ALK_PEAK=ALK_PEAK_STDLAB2
BILI_PEAK=BILI_PEAK_STDLAB2 INR_PEAK=INR_PEAK_STDLAB2));
  BY UNIQUEID;

  ALT_PEAK=MAX(ALT_PEAK_LABFLOW,ALT_PEAK_STDLAB2);
  ALK_PEAK=MAX(ALK_PEAK_LABFLOW,ALK_PEAK_STDLAB2);
  BILI_PEAK=MAX(BILI_PEAK_LABFLOW,BILI_PEAK_STDLAB2);
  INR_PEAK=MAX(INR_PEAK_LABFLOW,INR_PEAK_STDLAB2);
RUN;
```

```
*** Liver disease;
PROC SORT DATA=SASDATA2.LIVHx OUT=LIVHX;
  BY UNIQUEID;
RUN;

data livdis (keep=uniqueid LIVER_DISEASE);
  set livhx (WHERE=(VISITID=17598));
  by UNIQUEID;
  length LIVER_DISEASE 8.;
  retain LIVER_DISEASE;
  if first.UNIQUEID then LIVER_DISEASE=0;
  if MEDLIVHX in(1 2 3 4 5 6 7 8 10 11 12 13 16) then LIVER_DISEASE=1;
  if last.UNIQUEID then output;
run;

*** judgment: SEVERITY, ASSESS, JUDGE_FLAG***;
*** FORM: keep reassessments when there are multiples ***;
PROC SORT DATA=SASDATA2.judgmnt OUT=JUDGMENT;
  BY UNIQUEID;
RUN;

data judgment(keep=UNIQUEID FORM SEVERITY ASSESS JUDGE_FLAG);
  set judgment;
  where severity^=.;
  if SEVERITY in(1 2 3) then JUDGE_FLAG=0;
  else if SEVERITY in(4 5) then JUDGE_FLAG=1;
  else abort;
  if REVIEW=4 and strip(FORM)="INITIAL_ASSESSMENT" then output;

proc sort data=judgment nodupkeys;
  by _ALL_;

proc sort data=judgment;
  by uniqueid;
```

```
*** chronic dili;
data chronic (keep=UNIQUEID CHRONIC CINEXC:);
  set sasdata2.chronic;
  if SUM(CINEXC1,CINEXC2,CINEXC3,CINEXC4,CINEXC5,CINEXC9)>0 AND SUM(CINEXC6,CINEXC7,CINEXC8,CINEXC10)=0 then CHRONIC=1;
  ELSE IF CINEXC1=0 AND CINEXC2=0 AND CINEXC3=0 AND CINEXC4=0 AND CINEXC5=0 AND
          CINEXC6=0 AND CINEXC7=0 AND CINEXC8=0 AND CINEXC9=0 AND CINEXC10=0 THEN CHRONIC=0;
  ELSE CHRONIC=9;
RUN;

*** Death;
DATA PTCOMPL;
  SET SASDATA2.PTCOMPL;

  IF DEATHDT^=. THEN DECEASED=1;
  ELSE DECEASED=0;
RUN;

/*********************************************************/
/* Create dataset to use to replicate manuscript numbers */
/*********************************************************/
proc sort data=population;
  by uniqueid;
run;

PROC SORT DATA=SASDATA2.DEMOG OUT=DEMOG;
  BY UNIQUEID;
RUN;

PROC SORT DATA=SASDATA2.PEX1 OUT=PEX1;
  BY UNIQUEID;
RUN;

PROC SORT DATA=SASDATA2.MEDHX1 OUT=MEDHX1 NODUPKEY;
  BY UNIQUEID;
  WHERE PASTHX=1;
RUN;

PROC SORT DATA=EOSINOPHIL;
  BY UNIQUEID;
RUN;
```

```
DATA POPULATION;
  MERGE POPULATION (IN=INPOP)
        HXINJ2 (KEEP=UNIQUEID)
        DEMOG (IN=INDEMOG KEEP=UNIQUEID SEX WHITE BLACK ASIAN)
        PEX1  (IN=INPEX1 KEEP=UNIQUEID HTUN HT WTUN WT VISIT WHERE=(VISIT='vsBASE'))
        MEDHX1 (IN=INMEDHX)
        JUDGMENT (IN=INJUDG KEEP=UNIQUEID SEVERITY ASSESS)
        MEDHX2   (KEEP=UNIQUEID ALLERGY)
        LIVDIS   (IN=INLIVDIS KEEP=UNIQUEID LIVER_DISEASE)
       HX_LAB
       LAB_DILI_ONSET
       PEAKS
       CHRONIC
       SIGNS   (KEEP=UNIQUEID JAUNDICE FEVER RASH)
       PTCOMPL (KEEP=UNIQUEID DECEASED NLLRDTH NONDILI)
       LIVER_TRANSPLANT (KEEP=UNIQUEID LIVER_TRANSPLANT)
       EOSINOPHIL
       ;
  BY UNIQUEID;
  IF INPOP AND INDEMOG;

  if htun=1 then bmi_height=ht/100;
  else if htun=2 then bmi_height=ht/39.3701;
  if wtun=1 then bmi_weight=wt;
  else if wtun=2 then bmi_weight=wt/2.20462;
  if bmi_height>. and bmi_weight>. then bmi=bmi_weight/(bmi_height*bmi_height);

  IF ^INLIVDIS THEN LIVER_DISEASE=0;
  DIABETES=(PASTHX^=.);
  IMMUNO=(SUM(FEVER,RASH,EOS500N)>1);
RUN;

TITLE2 'Table 1 manuscript checks';
TITLE3 'Entire cohort';
PROC FREQ DATA=POPULATION;
  TABLE FLOW_NUM SEX WHITE * BLACK * ASIAN /LIST MISSING;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;
```

```
PROC MEANS DATA=POPULATION;
  VAR BMI;
RUN;

PROC FREQ DATA=POPULATION;
  TABLE ALLERGY IMMUNO /MISSING;
  TABLE LIVER_DISEASE;
  TABLE DIABETES JAUNDICE /MISSING;
RUN;

PROC MEANS DATA=POPULATION;
  VAR ALTVAL AKPVAL;
RUN;

PROC MEANS DATA=POPULATION;
  VAR STBVAL;
  WHERE STBUNIT^=4;
RUN;

PROC MEANS DATA=POPULATION;
  VAR INRRATE;
RUN;

PROC MEANS DATA=POPULATION;
  VAR ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK;
RUN;

PROC FREQ DATA=POPULATION;
  TABLE EOS500N ASSESS SEVERITY DECEASED NLLRDTH LIVER_TRANSPLANT CHRONIC;
RUN;

TITLE3 'By classification';
PROC FREQ DATA=POPULATION;
  TABLE SEX * FLOW_NUM;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;

PROC SORT DATA=POPULATION;
  BY FLOW_NUM;
RUN;
```

```
PROC FREQ DATA=POPULATION;
  TABLE WHITE * BLACK * ASIAN /LIST MISSING;
  BY FLOW_NUM;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;

PROC MEANS DATA=POPULATION;
  VAR BMI;
  CLASS FLOW_NUM;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;

PROC FREQ DATA=POPULATION;
  TABLE (ALLERGY IMMUNO) * FLOW_NUM /MISSING;
  TABLE LIVER_DISEASE * FLOW_NUM;
  TABLE (DIABETES JAUNDICE) * FLOW_NUM /MISSING;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;

PROC MEANS DATA=POPULATION;
  VAR ALTVAL AKPVAL STBVAL INRRATE ALT_PEAK ALK_PEAK BILI_PEAK INR_PEAK;
  CLASS FLOW_NUM;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;

PROC FREQ DATA=POPULATION;
  TABLE (EOS500N ASSESS SEVERITY DECEASED NLLRDTH LIVER_TRANSPLANT CHRONIC) * FLOW_NUM;
  FORMAT FLOW_NUM FLOWRATE.;
RUN;
```