MWSUG 2017 - Paper RF03

No News Is Good News:

A Smart Way to Impute Missing Clinical Trial Lab DATA

Ming Yan, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

In clinical trials, specific laboratory microscopic UA, RBC morphology, and special WBC subordinate tests are reported to sponsors ONLY if an abnormality is observed. The normal results which are not explicitly reported, however, need to be in place in order to compute a percentage of abnormalities for each lab test in the subject population. This macro starts from SDTM LB domain that only stores observed abnormal lab tests and converts it to ADaM with all the normal test results filled out for all patients at all time-point.

BACKGROUND

Morphology DATA is widely collected in Oncology clinical trials and urine microscopic example is commonly collected for compounds where there are kidney function/inflammation concerns. Cell morphology, for example, is a collection of subtype tests on coloring, cell size, cell shape, etc., for Reb Blood Cell (RBC), White Blood Cell (WBC) primitives/Atypical cells, and platelet. The laboratories may report sponsors only the subtypes when an abnormality is observed. A common practice at the laboratory is that technicians look at blood sample under microscope and count the subtype that can be seen. What is seen is reported and what is not seen is not reported to the sponsor. Since all the subtypes that are seen are abnormal, the subtype analytes in the GLS (Generic Lab System) data are always abnormal ("Bad news"). What is absent in the GLS data set ("No news") is considered normal ("Good news"). If there is no abnormal finding at all for a patient, an over-arching lab test code is used to indicate all the sub-ordinary tests are normal for that patient.

When analyzing cell morphology and urine microscopic exams result, the absent subtype tests must be first imputed to "normal" before percentage of abnormalities could be calculated. For example,

% of patients with abnormal finding of "HYPOCHROMIA" @ Visit 5

= (# of patients with "HYPOCHROMIA" reported explicitly as present @ Visit 5 in GLS DATA set) / (# of patients with Hematology Cell Morphology test performed @ Visit 5).

EXAMPLES

The program reads in SDTM DATA sets in following structure. All the subtype tests share the same LBTESTCD and LBTEST value as the over-arching test. LBSTRESC stores the name of the abnormal subtype tests (Visit 6). When all subtypes are normal, LBSTRESC='NORMAL' (Visit 1, 2, 5).

SUBJID	VISITNUM	VISIT	LBDTM	LBREFID	LBTEST	LBTESTCD	lbstresc
10322	1	V1 Screening	11AUG16:21:54:26	5000001182	RBC Morphology	HMT71	NORMAL
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	RBC Morphology	HMT71	NORMAL
10322	5	V5	29SEP16:13:37:35	5000001184	RBC Morphology	HMT71	NORMAL
10322	6	V6	05DEC16:09:44:14	5000001185	RBC Morphology	HMT71	Anisocytosis
10322	6	V6	05DEC16:09:44:14	5000001185	RBC Morphology	HMT71	Polychromatophilia
10322	6	V6	05DEC16:09:44:14	5000001185	RBC Morphology	HMT71	Schistocytosis
10322	6	V6	05DEC16:09:44:14	5000001185	RBC Morphology	HMT71	Elliptocytosis
10322	6	V6	05DEC16:09:44:14	5000001185	RBC Morphology	HMT71	Tear Drop Cells

We need to perform two analyses starting from the DATA above. One is to count percentage of patients with abnormal over-arching lab, the other is to count percentage of patients with abnormal subtype.

METHODOLOGY

The program below is used to detect any abnormal subtype, and fill up all absence with “Normal” for all patients. The only exception is, if a patient didn’t conduct any lab draw at a certain visit, that visit is not filled. In the example above, the patient didn’t conduct visit 3 or 4.

Pre-process:

```
DATA LB2;
   SET LB1;
   PARCAT3=LBTEST;                                                                    1
   *** Subtype ***;
   if LBSTRESC not in (‘ABNORMAL’,’NORMAL’) then do;                                  2
      AVALC='ABNORMAL';
      PARAM = LBSTRESC;
      if SCAN(PARAM,3,' ')>'' and SUBSTR(SCAN(PARAM,3,' '),1,1)^='(' then
PARAMCD=SUBSTR(SCAN(PARAM,1,' '),1,4)||SUBSTR(SCAN(PARAM,2,' '),1,1)||SUBSTR(SCAN(PARAM,3,'
'),1,3);                                                                              3
      else if SCAN(PARAM,2,' ')>'' and SUBSTR(SCAN(PARAM,2,' '),1,1)^='(' then
PARAMCD=SUBSTR(SCAN(PARAM,1,' '),1,4)||SUBSTR(SCAN(PARAM,2,' '),1,4);
      else PARAMCD=SUBSTR(PARAM,1,8);
   End;
   *** Over-arching ***;
   else if LBSTRESC in (‘ABNORMAL’,’NORMAL’) then do;                                 4
      AVALC=LBSTRESC;
      PARAMCD=LBTESTCD; PARAM=LBTEST;
   End;
RUN;
```

1. Store over-arching test name in PARCAT3 to group all subtype tests.
2. All subtype test that is reported is abnormal. The subtype name is stored in LBSTRESC in SDTM. For these records, set AVALC to ‘ABNORMAL’ and copy test name from LBSTRESC to PARAM.
3. Construct unique PARAMCD by concatenating the first few letters from every word in PARAM.
4. All over-arching test’s value is either ‘NORMAL’ or ‘ABNORMAL’. For these records, copy result from LBSTRESC to AVALC and copy test name/code from LBTEST/LBTESTCD to PARAM/PARAMCD respectively.

Fill in records for normal subtype

```
PROC SORT DATA=LB2 OUT=ABN(keep=PARAMCD PARAMCD2 PARAM) NODUPKEY;                     5
   WHERE AVALC NOT IN ('NORMAL','') and PARAM NOT IN ('RBC Morphology','WBC
Morphology','Microscopic (urine sediment)','');
   by PARAM;
RUN;
PROC SORT DATA=LB2 OUT=FRAME0 NODUPKEY;
   by SITEID subjid VISITNUM VISIT LBDTM LBREFID PARCAT3;                             6
RUN;
PROC SQL;                                                                             7
   CREATE TABLE FRAME AS
      SELECT l.*, r.PARAMCD, r.PARAM
          FROM FRAME0(KEEP=SITEID SUBJID VISITNUM VISIT LBDTM LBREFID PARCAT3) AS L, ABN AS R
             ORDER BY SITEID, SUBJID, VISITNUM, VISIT, LBDTM, LBREFID, PARAMCD;
quit;

PROC SORT DATA=LB2; by SITEID subjid VISITNUM VISIT LBDTM LBREFID PARAMCD; RUN;
PROC SORT DATA=LB2 OUT=ABN2;
   by SITEID subjid VISITNUM VISIT LBDTM LBREFID PARAMCD;
   where AVALC not in ('NORMAL','');
RUN;
DATA NORMS;
   MERGE FRAME(in=a) ABN2 (in=b);
   BY SITEID subjid VISITNUM VISIT LBDTM LBREFID PARAMCD;
   if a and not b;
   AVALC='NORMAL';                                                                    8
RUN;
DATA LB3;
   SET LB2(in=a) NORMS(in=b);                                                         9
   by SITEID subjid VISITNUM VISIT LBDTM LBREFID;
```

```
RUN;
```

5. Identify all abnormal subtypes.
6. FRAME0 contains one record for each patient each lab drawn.
7. Set up a shell, so that if any patient ever has any abnormal sub-type, that sub-type is filled for all patients at all lab draws. If any patient skipped a visit, the skipped visit is not to be filled in with "Normal" values.
8. For any sub-type test that is not reported as abnormal, fill up AVALC with 'NORMAL'.
9. Pool over-arching test (LB2), abnormal sub-type (LB2), and normal sub-type labs (NORMS) together.

Output final DATA sets

```
DATA OVERARCH SUBTYPE;                                                    10
  SET lb3;
  if PARAM=PARCAT3 then output overArch;
  else output SUBTYPE;              OVERARCH
RUN;
```

10. Store over-arching lab and sub-type lab in different DATA set for analysis.

FINAL OUTPUTS

The program creates below output DATA sets.

1. DATA set OVERARCH contains over-arching tests only, which only reports overall "Normal" or "Abnormal" without low-level information.

SUBJID	VISITNUM	VISIT	LBDTM	LBREFID	param	paramcd	AVALC
10322	1	V1 Screening	11AUG16:21:5...	5000001182	RBC Morphology	HMT71	NORMAL
10322	2	V2 Baseline	08OCT16:12:4...	5000001183	RBC Morphology	HMT71	NORMAL
10322	5	V5	29SEP16:13:3...	5000001184	RBC Morphology	HMT71	NORMAL
10322	6	V6	05DEC16:09:4...	5000001185	RBC Morphology	HMT71	ABNORMAL

2. DATA set SUBTYPE contains all subtype tests. Only visit 2 through 6 are displayed below for brevity. The patient doesn't have visit 3 or 4 in SDTM DATA, hence visit 3 and 4 records are not created in SUBTYPE.
   This patient has only 5 abnormal subtype tests but 10 tests are inserted as "Normal" for visit 5. This is because in addition to what we see in the example, some other patients have 5 other abnormal subtype tests. Percentage of abnormality needs to be calculated for any abnormal subtype, so as long as any patient reports any abnormal subtype, we fill in that subtype for the entire population.

SUBJID	VISITNUM	VISIT	LBDTM	LBREFID	param	paramcd	PARCAT3	AVALC	LBSPID
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Anisocytosis	Anisocyt	RBC Morphology	NORMAL	1
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Elliptcytosis	Elliptcy	RBC Morphology	NORMAL	2
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Hypochromia	Hypochro	RBC Morphology	NORMAL	3
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Macrocytosis	Macrocyt	RBC Morphology	NORMAL	4
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Microcytosis	Microcyt	RBC Morphology	NORMAL	5
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Poikilocytosis	Poikiloc	RBC Morphology	NORMAL	6
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Polychromia	Polychro	RBC Morphology	NORMAL	7
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Schistocytosis	Schistoc	RBC Morphology	NORMAL	8
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Target cells	Targcell	RBC Morphology	NORMAL	9
10322	2	V2 Baseline	08OCT16:12:40:26	5000001183	Tear drop cells	Teardcel	RBC Morphology	NORMAL	10
10322	5	V5	29SEP16:13:37:35	5000001184	Anisocytosis	Anisocyt	RBC Morphology	NORMAL	1
10322	5	V5	29SEP16:13:37:35	5000001184	Elliptcytosis	Elliptcy	RBC Morphology	NORMAL	2
10322	5	V5	29SEP16:13:37:35	5000001184	Hypochromia	Hypochro	RBC Morphology	NORMAL	3
10322	5	V5	29SEP16:13:37:35	5000001184	Macrocytosis	Macrocyt	RBC Morphology	NORMAL	4
10322	5	V5	29SEP16:13:37:35	5000001184	Microcytosis	Microcyt	RBC Morphology	NORMAL	5
10322	5	V5	29SEP16:13:37:35	5000001184	Poikilocytosis	Poikiloc	RBC Morphology	NORMAL	6
10322	5	V5	29SEP16:13:37:35	5000001184	Polychromia	Polychro	RBC Morphology	NORMAL	7
10322	5	V5	29SEP16:13:37:35	5000001184	Schistocytosis	Schistoc	RBC Morphology	NORMAL	8
10322	5	V5	29SEP16:13:37:35	5000001184	Target cells	Targcell	RBC Morphology	NORMAL	9
10322	5	V5	29SEP16:13:37:35	5000001184	Tear drop cells	Teardcel	RBC Morphology	NORMAL	10
10322	6	V6	05DEC16:09:44:14	5000001185	Anisocytosis	Anisocyt	RBC Morphology	ABNORMAL	1
10322	6	V6	05DEC16:09:44:14	5000001185	Elliptcytosis	Elliptcy	RBC Morphology	ABNORMAL	2
10322	6	V6	05DEC16:09:44:14	5000001185	Polychromia	Polychro	RBC Morphology	ABNORMAL	3
10322	6	V6	05DEC16:09:44:14	5000001185	Schistocytosis	Schistoc	RBC Morphology	ABNORMAL	4
10322	6	V6	05DEC16:09:44:14	5000001185	Tear drop cells	Teardcel	RBC Morphology	ABNORMAL	5
10322	6	V6	05DEC16:09:44:14	5000001185	Hypochromia	Hypochro	RBC Morphology	NORMAL	6
10322	6	V6	05DEC16:09:44:14	5000001185	Macrocytosis	Macrocyt	RBC Morphology	NORMAL	7
10322	6	V6	05DEC16:09:44:14	5000001185	Microcytosis	Microcyt	RBC Morphology	NORMAL	8
10322	6	V6	05DEC16:09:44:14	5000001185	Poikilocytosis	Poikiloc	RBC Morphology	NORMAL	9
10322	6	V6	05DEC16:09:44:14	5000001185	Target cells	Targcell	RBC Morphology	NORMAL	10

CONCLUSION

Since labs only report abnormal subtype tests for microscopic UA, RBC morphology, and special WBC, all the non-reported subtypes are assumed to be “Normal”. This paper presents a program to impute the absent subtypes to be normal at ADaM level to expedite the TFL creation.

ACKNOWLEDGMENTS

Thanks to my mentor Cindy Lee, my supervisor Hangtao Xu, and other colleagues at Eli Lilly and Company, who reviewed this paper and provided helpful suggestions.

CONTACT INFORMATION

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

Ming (Janice) Yan
Eli Lilly and Company
317-433-3783
yan_ming_x1@lilly.com