



# Standard SDTM Methodology with Legacy Data

Nth Analytics

---

Mike Todd, President

[info@nthanalytics.com](mailto:info@nthanalytics.com)

908 672 5649

Barbara Costantini, Senior VP

[bcostantini@nthanalytics.com](mailto:bcostantini@nthanalytics.com)

609 406 0149

# Legacy Data

- Old studies never die ...
- Legacy studies are often required for submissions or pharmacovigilance
- Two types:
  - Well-documented
  - Less well-documented
- This presentation focuses on the worst case
  - Large pharmacovigilance project with lots of issues

# Legacy Study Scenarios

- Best case
  - Requires no/little additional Data Management intervention
  - Maps “easily” to SDTM format
- Worst case
  - Becomes a Data Management problem with high level of intervention prior to mapping to SDTM
  - May require heavy statistician involvement for validation
  - Problematic data, may be excluded or imputed, so long as a reasonable, well-documented process can be defined

# Case Study

# A Large, Messy Database

- Large pharmacovigilance project
  - Sponsor required to address questions from regulatory authorities regarding safety issues.
- Convert demography, concomitant medications, and labs to SDTM
  - Minimum set of SDTM files required: DM, SC, CM, LB
  - SUPPCM and SUPPLB also needed
- 42 studies from 1987 to 2001 comprising 400 sites, 5,000 subjects, 1.2M lab records, 92,000 medication records
  - Original CRFs and study reports unavailable for some studies

# Staff

- Production
  - Data Manager
  - Clinician
  - Programmer
- Validation
  - Statistician
- Coordinator
  - Project Manager

# Strategy

- Start with analysis datasets
  - Already pooled
  - Already gone through one round of cleaning
  - Some consistency issues resolved
  - Working with 42 individual trial datasets was infeasible
- Three-Stage Process
  1. Data Management
  2. Validation
  3. SDTM Mapping
- Once Data Management straightens out the data, SDTM is the easy part

# Why Convert to SDTM

- SDTM offers many added benefits beyond just cleaning the data
  - Integrate with current database, if necessary
  - Possible submissions to FDA
  - Ease of use
  - Not that hard, compared with data cleaning problems

# Conmeds - Problems

- Inconsistent data collection
- Inconsistent variable naming conventions
- Inconsistent/missing medication coding
  - Use of multiple medication dictionaries
- 92,000+ medication records in the file
  - 20,000 meds with a populated verbatim term
  - 56,000 meds with a dictionary term but no verbatim term
  - Remaining medications had neither verbatim nor dictionary term
    - Instead, there was a 3-digit code from an old internal dictionary (existing only in hard-copy)

# Conmeds - Solutions

- Create new dataset and new derived term to be autoencoded, using lowest level of granularity available on each record
  - verbatim, dictionary term, 3-digit code
- Recreate electronic version of old 3-digit codes/decodes dictionary.
- Use the latest WHO Drug Dictionary to recode all the medication terms.

# Labs

- Lab data was a mess
- The sponsor did not have a good “comfort level” in addressing safety issues
- Twenty-eight (28) lab analytes of importance
  - Out of 40 total

# Labs - Problems

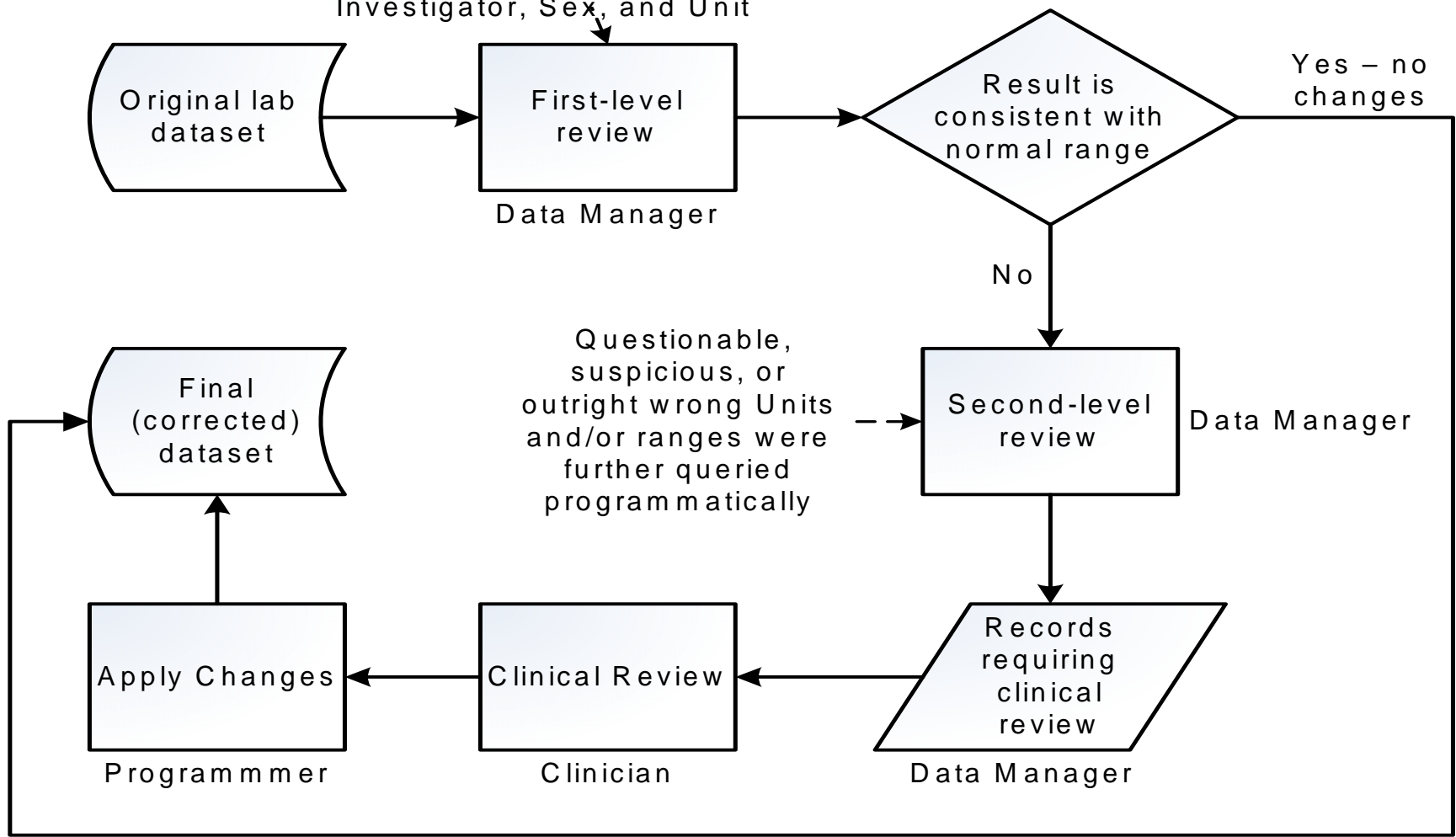
- Normal ranges missing, incomplete, wrong compared to result
  - 88,112 records had no range at all
  - 9,413 records had the low end of the range missing
  - 123 records had the high end of the range missing
- Normal range flags (H,L) inconsistent with ranges
- Implausible lab values
  - questionable, suspicious, or outright wrong
- Implausible units
- Questionable value/unit conversions

# Labs - Solutions

- Data Management review
  - Determine extent of the problem
  - Ranges and units of measure fixed where possible
  - Determined when rule-based changes could be applied
  - Remaining issues passed to clinician for review
- Clinical review
  - Final approval on proposed changes
  - Provided referenced normal ranges to replace missing ranges
- Programmer
  - Applied changes from data manager and clinician
  - Derived standard lab values/units from the corrected data
- Statistician
  - Validated the whole thing
- Complete audit trail maintained

# Lab Review Process

Review minimum and maximum low and high normal range vs. result by Analyte, Study, Investigator, Sex, and Unit



# Labs - First Level Review

- The original lab value was the bedrock on which everything depended.
  - Only units and normal ranges were subject to change after review and authorization by the clinician.
  - No lab values were changed
- The magnitude of the range was compared with the magnitude of the lab value.
  - by analyte, investigator, study, unit and sex
- The unit was reviewed for appropriateness with the range, the result and the analyte
- Units and/or ranges that were questionable, suspicious, or outright wrong examined in more detail.
  - Identify patient numbers and the context and extent of the problem.
  - This tabulation formed the basis of the second level review.

# Labs - Second Level Review

- Lab data were transferred to Excel for clinical review.
- Possible actions (if any), included:
  - No change
  - Correct the units
  - Apply generic normal ranges
  - Exclude the value as implausible

# Lab Data - Programming Steps

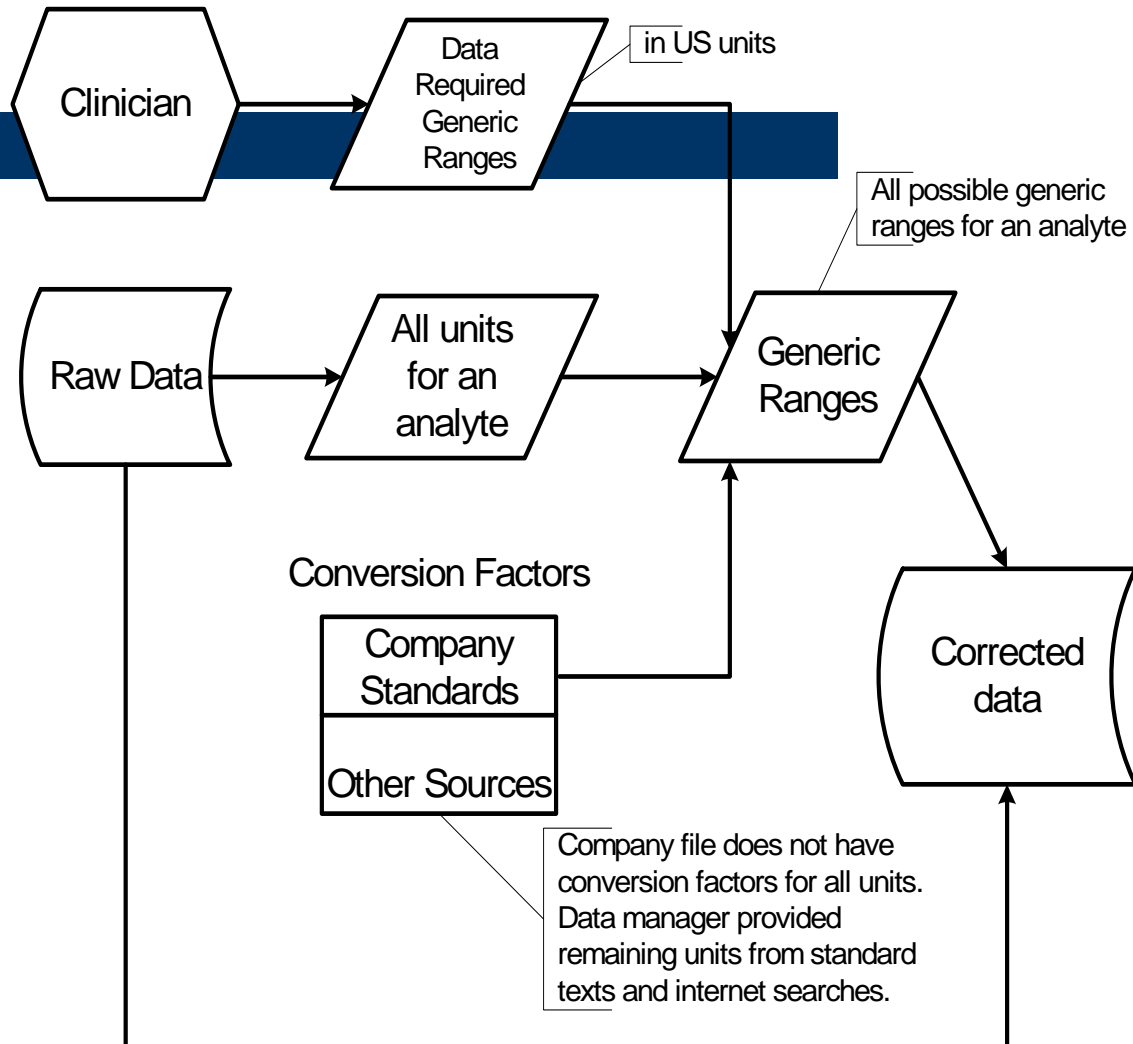
1. Programmers applied imputed values to the database
2. Missing ranges were replaced with generic ranges programmatically
3. Programmers converted to standard units and assigned high/low normal range flag after all imputations and generic ranges were applied.
4. Clinicians identified implausible values to be flagged in the database and excluded from analysis.

# Generic Normal Ranges

- Since more than 50% of ranges were missing from lab records, the clinician established generic ranges that would be used in the absence of the laboratory-specified normal range.
  - References included the Merck manual and Harriet Lane Handbook (for pediatric ranges)
- Due to sheer volume of issues, the project team agreed that researching the original normal ranges through extensive review of the original CRFs for each subject affected was infeasible given the resources available.

# Generic Normal Ranges

- Replace missing normal ranges with generic ranges provided by the clinician if either low or high or both are missing,
- All normal range flagging is done on the original results in the original units.
- Clinician provides generic ranges in US units. However, final normal ranges have to be in original units, and these are programmatically converted back to the original units.



# Implausible Values

- After all possible imputations were applied, some results simply did not make sense.
  - Questionable values were presented to the clinician to review for plausibility.
  - A flag was set in the database on each record for which there was an implausible value.
- Results that were ambiguous, i.e. the value could have been a reasonable outlier, or the unit of measure could have been in error, were left as is.
- Results that had no unit of measure and no normal range were flagged as implausible (i.e. unusable for analysis).

# Implausible Values

| Analyte | Unit   | Study   | Subject | Result | Normal Range |      | Comment  |
|---------|--------|---------|---------|--------|--------------|------|--|
|         |        |         |         |        | Low          | High |  |
| SODIUM  | mmol/L | STUDY1  | S1-001  | 4.1    |              |      | Possibly a potassium value.                            |
| SODIUM  | mmol/L | STUDY1  | S1-005  | 11.3   | .            | .    | Possible data entry error or result from another test. |
| ALBUM   | g/dL   | STUDY3  | S3-006  | 0.47   | 3.2          | 5.5  | Possible data entry error.                             |
| POTASS  | mmol/L | STUDY4  | S4-034  | 141    | .            | .    | Likely sodium value.                                   |
| POTASS  | mEq/L  | STUDY5  | S5-011  | 25.2   | 3.5          | 5.3  | Obvious error of some sort.                            |
| POTASS  | mEq/L  | STUDY6  | S6-039  | 20.2   | 3.5          | 5.3  | Obvious error of some sort.                            |
| POTASS  | mmol/L | STUDY4  | S4-112  | 74     | .            | .    | Sodium and potassium values likely switched.           |
| POTASS  | mEq/L  | STUDY8  | S8-309  | 17.4   | 3.5          | 5.3  | Obvious error of some sort.                            |
| POTASS  | mmol/L | STUDY9  | S9-237  | 149    | .            | .    | Likely sodium value.                                   |
| POTASS  | mg/L   | STUDY9  | S9-247  | 100    | .            | .    | Unit and value make no sense.                          |
| POTASS  | mmol/L | STUDY9  | S9-651  | 243.4  | .            | .    | Sodium and potassium values likely switched.           |
| POTASS  | mmol/L | STUDY9  | S9-112  | 35     | .            | .    | Possible data entry error.                             |
| POTASS  | mEq/L  | STUDY9  | S9-126  | 0.7    | .            | .    | Obvious error of some sort.                            |
| ALT     | U/L    | STUDY11 | S11-212 | 2.1    | 0            | 39   | Possible data entry error.                             |
| ALT     | IU/L   | STUDY11 | S11-335 | 0.5    | 0            | 31   | Error of some sort.                                    |

# Validation

- Obviously, such a multi-layered cleaning and imputation process requires extensive validation
- Validation performed by statistician
  - Statisticians are detailed oriented, and are experts in findings errors in data
  - Statisticians have an entirely different perspective than data managers, which is a plus when independent confirmation is required
- The entire process was independently programmed using different methodology to the extent possible.
  - All validation was performed using SAS Enterprise Guide
- The goal was to confirm, through independent programming, that all of the records that should have changed did change, and records that should not have changed did not change

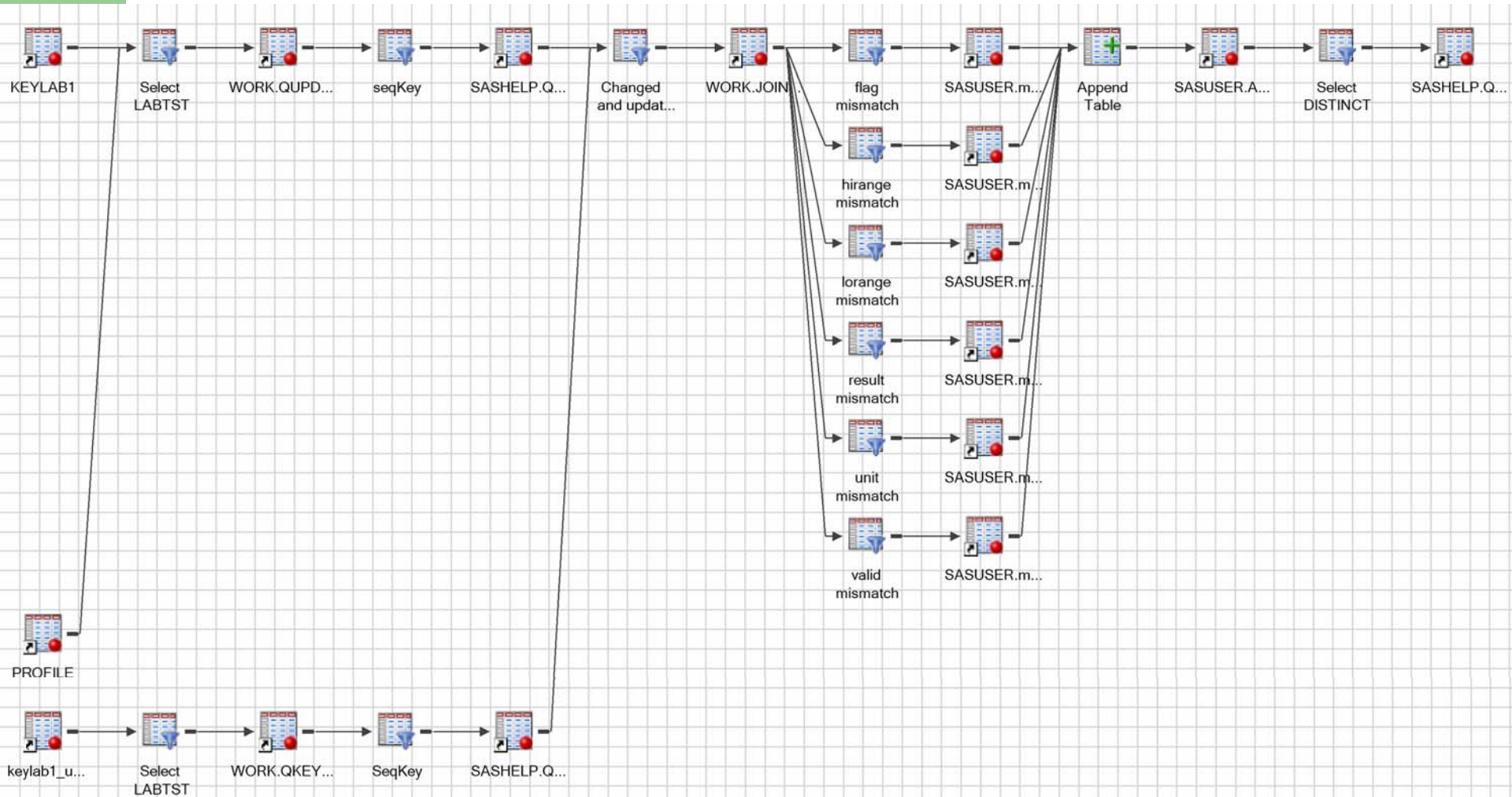
# Validation Steps

1. Identify all records that are changed comparing original vs. final datasets. Potential changes fell into 6 categories:
  - Normal range flags
  - Normal range upper/lower limits
  - Results
  - Units
  - Invalid data flags
  - Conversions
- Note that results should not have changed. The validation step simply verified that there were in fact no changes to the RESULT variable

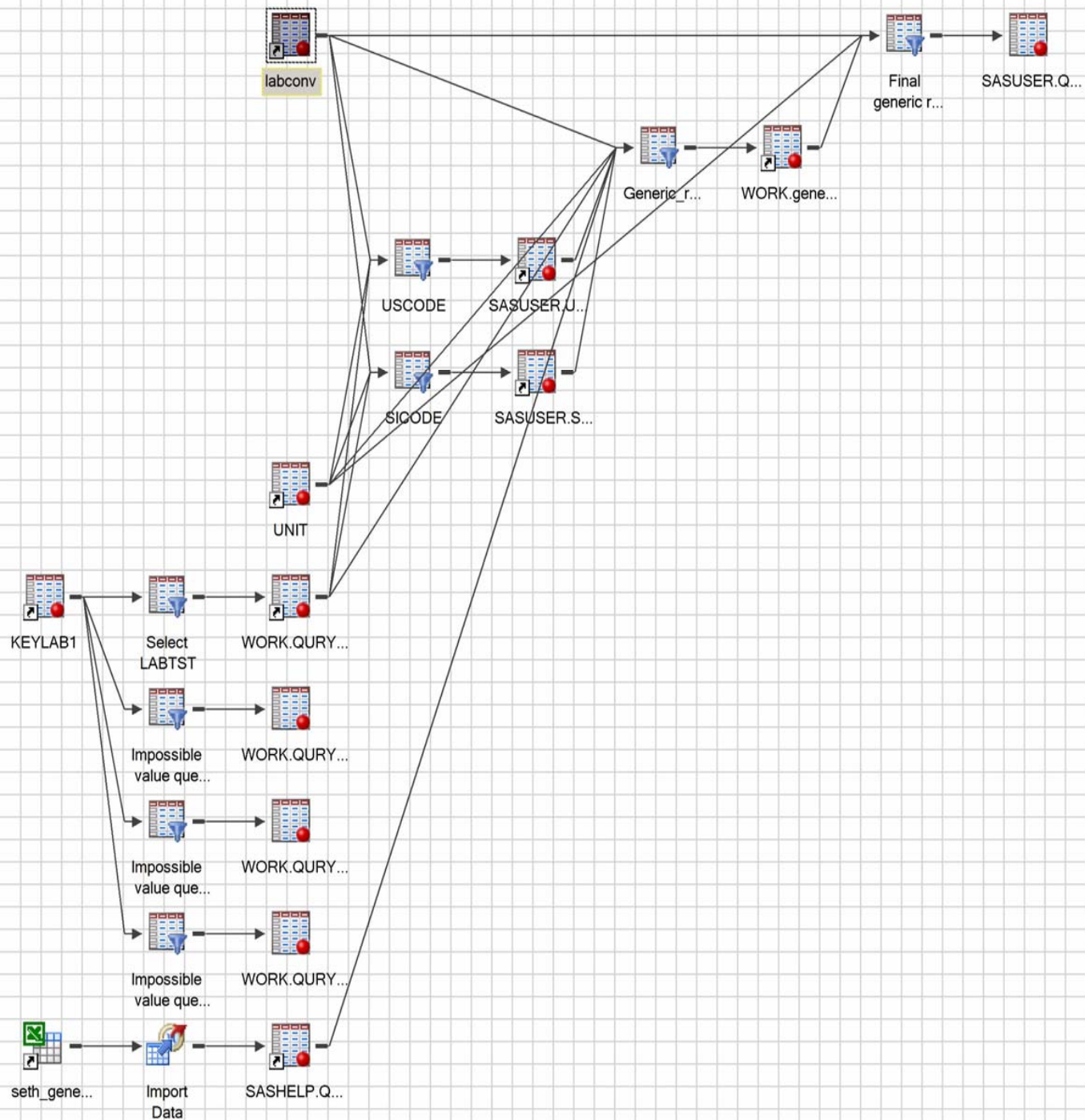
## Validation Steps (cont)

2. Independently create a dataset containing the generic normal ranges, using ranges provided in the emails from the clinician as the source
3. Independently program high/low flags, normal range limits, results, units, valid flags, and converted values and units
4. Compare the results, and send any discrepancies back to data manager and programmer for resolution. Repeat until all discrepancies are resolved.

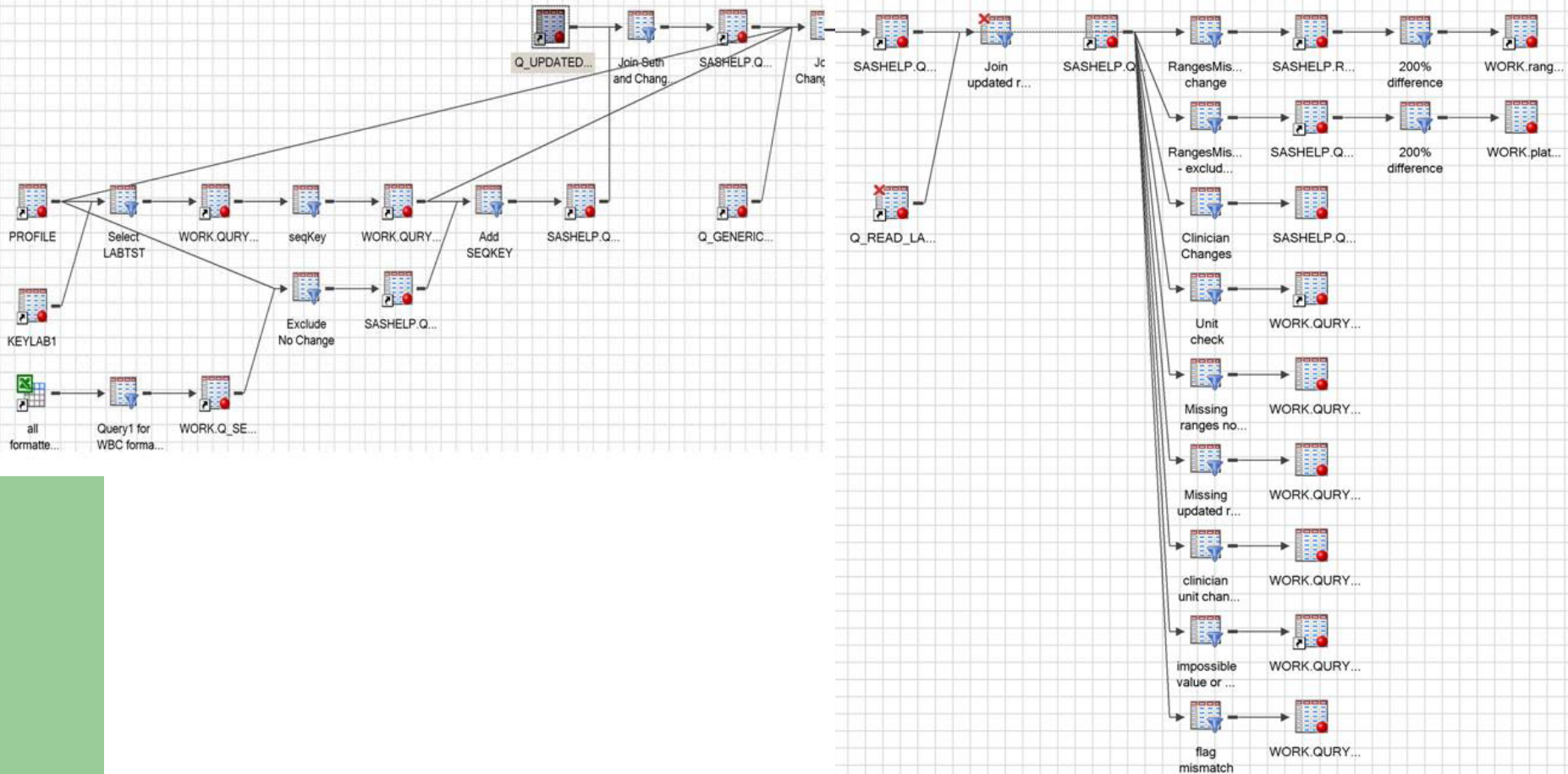
# Differences between new and old datasets



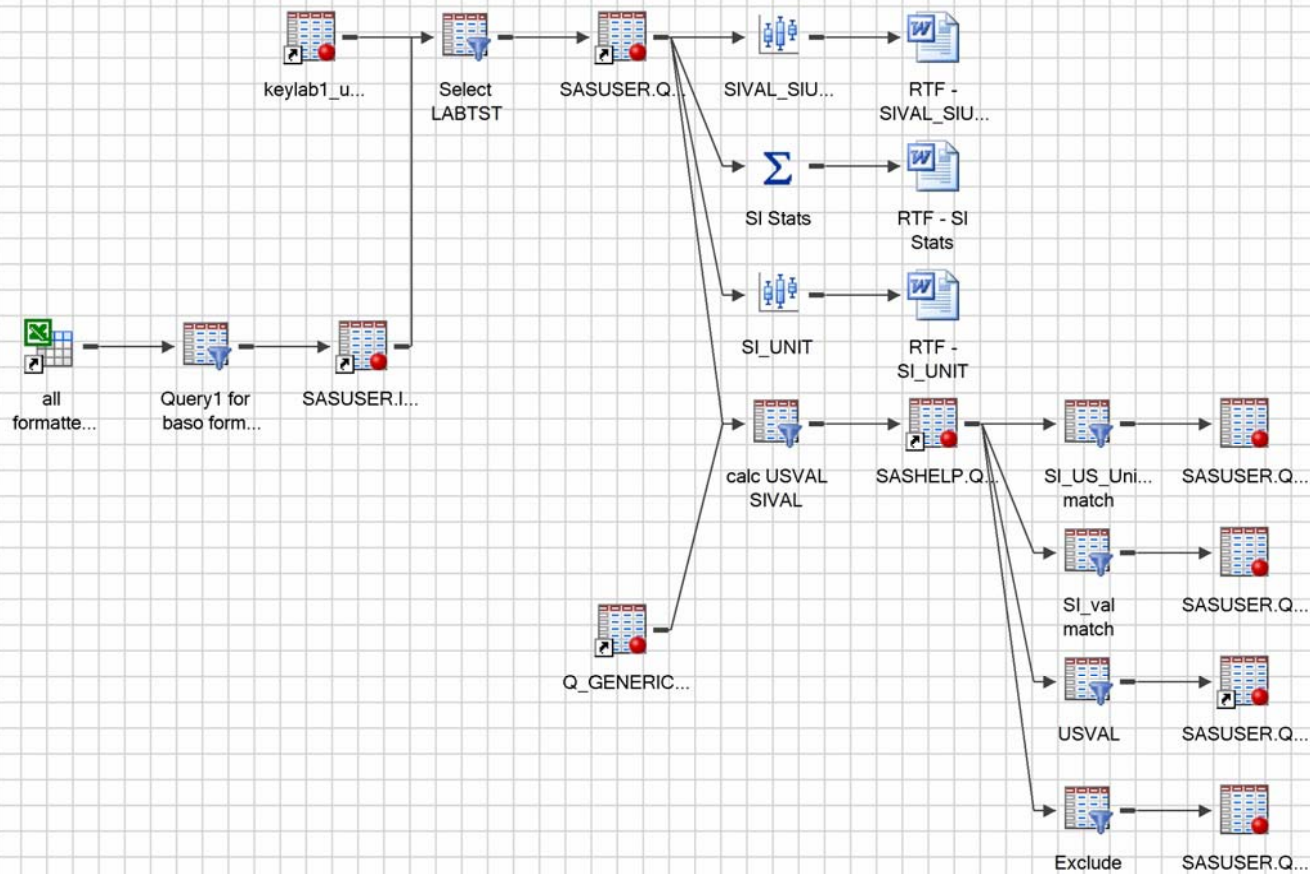
# Recreate Generic Normal Ranges



# Changes vs. Independently Programmed Results



# Check Conversions



# SDTM Mapping

# Domains

- DM
- SC
- CM
- SUPPCM
- LB
- SUPPLB

# Strategy

- Map the minimum number of variables
  - Required
  - Expected
  - Permissible if absolutely necessary (or easy)
- Analysis dataset strategy

# DM

All DM variables mapped directly from analysis dataset.

| Variable Name | Variable Label                    | Type | Mapping                        | Core |
|---------------|-----------------------------------|------|--------------------------------|------|
| STUDYID       | Study Identifier                  | Char | LEGACY_DEMOG.PNO               | Req  |
| DOMAIN        | Domain Abbreviation               | Char | = 'DM'                         | Req  |
| USUBJID       | Unique Subject Identifier         | Char | LEGACY_DEMOG.UPATNO            | Req  |
| SUBJID        | Subject Identifier for the Study  | Char | LEGACY_DEMOG.UPATNO            | Req  |
| RFSTDTC       | Subject Reference Start Date/Time | Char | %ISODATE(LEGACY_DEMOG.FRSTDOS) | Exp  |
| RFENDTC       | Subject Reference End Date/Time   | Char | = ''                           | Exp  |
| SITEID        | Study Site Identifier             | Char | LEGACY_DEMOG.INO               | Req  |
| INVID         | Investigator Identifier           | Char |                                | Perm |
| INVNAM        | Investigator Name                 | Char | LEGACY_DEMOG.INVNAME           | Perm |
| BRTHDTC       | Date/Time of Birth                | Char |                                | Perm |
| AGE           | Age in AGEU at RFSTDTC            | Num  | LEGACY_DEMOG.AGE               | Exp  |
| AGEU          | Age Units                         | Char | = 'YEARS'                      | Exp  |
| SEX           | Sex                               | Char | LEGACY_DEMOG.SEX               | Req  |
| RACE          | Race                              | Char | LEGACY_DEMOG.RACE              | Exp  |
| ETHNIC        | Ethnicity                         | Char |                                | Perm |
| ARMCD         | Planned Arm Code                  | Char | LEGACY_DEMOG.TRT_CODE          | Req  |
| ARM           | Description of Planned Arm        | Char | LEGACY_DEMOG.TRTF              | Req  |
| COUNTRY       | Country                           | Char | LEGACY_DEMOG.CNTRY             | Req  |
| DMDTC         | Date/Time of Collection           | Char |                                | Perm |
| DMDY          | Study Day of Collection           | Num  |                                | Perm |

# SC

- Baseline weight goes in SC
- Repeat for baseline height
- SCDTC – date not in analysis dataset, leave blank

| Variable Name | Variable Label                           | Type | Comments                         | Core |
|---------------|--|------|----------------------------------|------|
| STUDYID       | Study Identifier                         | Char | LEGACY_DEMOG.PNO                 | Req  |
| DOMAIN        | Domain Abbreviation                      | Char | ='SC'                            | Req  |
| USUBJID       | Unique Subject Identifier                | Char | LEGACY_DEMOG.UPATNO              | Req  |
| SCSEQ         | Sequence Number                          | Num  | %SEQ                             | Req  |
| SCSTAT        | Status of SC Measurement                 | Char |                                  | Perm |
| SCDY          | Study Day of Examination                 | Num  |                                  | Perm |
| SCGRPID       | Group ID                                 | Char |                                  | Perm |
| SCSPID        | Sponsor-Defined Identifier               | Char |                                  | Perm |
| SCTESTCD      | Subject Characteristic Short Name        | Char | ='BASEWGT'                       | Req  |
| SCTEST        | Subject Characteristic                   | Char | ='BASELINE WEIGHT'               | Req  |
| SCSAT         | Category for Subject Characteristic      | Char |                                  | Perm |
| SCSCAT        | Subcategory for Subject Characteristic   | Char |                                  | Perm |
| SCORRES       | Result or Finding in Original Units      | Char | LEGACY_DEMOG.BASEWGT             | Exp  |
| SCORRESU      | Original Units                           | Char | ='KG'                            | Perm |
| SCSTRESC      | Character Result/Finding in Std Units    | Char | PUT(LEGACY_DEMOG.BASEWGT,BEST8.) | Exp  |
| SCSTRESN      | Numeric Result/Finding in Standard Units | Num  | LEGACY_DEMOG.BASEWGT             | Perm |
| SCSTRESU      | Standard Units                           | Char | ='KG'                            | Perm |
| SCREASND      | Reason Not Performed                     | Char |                                  | Perm |
| SCDTC         | Date/Time of Collection                  | Char | =''                              | Exp  |
| SCEVAL        | Evaluator                                | Char |                                  | Perm |
| SCGRPID       | Group ID                                 | Char |                                  | Perm |
| SCSPID        | Sponsor-Defined Identifier               | Char |                                  | Perm |

# CM – Variables Mapped

| Variable Name | Variable Label                         | Type | Mapping   | Core |
|---------------|--|------|---|------|
| STUDYID       | Study Identifier                       | Char | LEGACY_DEMOG.PNO  | Req  |
| DOMAIN        | Domain Abbreviation                    | Char | = 'CM'  | Req  |
| USUBJID       | Unique Subject Identifier              | Char | LEGACY_DEMOG.UPATNO   | Req  |
| CMSEQ         | Sequence Number                        | Num  | %SEQ  | Req  |
| CMTRT         | Reported Name of Drug, Med. Or Therapy | Char | LEGACY_CM.PSEUDOVB  | Req  |
| CMDECOD       | Standardized Medication Name           | Char | LEGACY_CM.PREFTERM  | Perm |
| CMCAT         | Category for Medication                | Char | CASE WHEN LEGACY_CM.AEDYN='Y' THEN 'AED'<br>ELSE 'NON-AED' END  | Perm |
| CMINDC        | Indication                             | Char | LEGACY_CM.INDICT  | Perm |
| CMDOSTXT      | Dose Description                       | Char | LEGACY_CM.TOTDOSE   | Perm |
| CMROUTE       | Route of Administration                | Char | LEGACY_CM.ROUTE   | Perm |
| CMSTDTC       | Start Date/Time of Medication          | Char | %ISODATE(LEGACY_CM.CSTARTDT)  | Perm |
| CMENDTC       | End Date/Time of Medication            | Char | %ISODATE(LEGACY_CM.CSTOPDT)   | Perm |
| CMSTDY        | Study Day of Start of Medication       | Num  | %RELDAY(LEGACY_CM.CSTOPDT)  | Perm |
| CMENDY        | Study Day of End of Medication         | Num  | %RELDAY(LEGACY_CM.CSTARTDT)   | Perm |
| CMDUR         | Duration of Medication                 | Char | CASE WHEN LEGACY_CM.DUR IS NOT NULL THEN<br>='Y' THEN TRIM(INPUT(LEGACY_CM.DUR,8.))    '<br>DAYS' ELSE '' END | Perm |

# CM – Variables Not Mapped

| Variable Name | Variable Label                     | Type | Mapping | Core |
|---------------|------------------------------------|------|---------|------|
| CMGRPID       | Group ID                           | Char |         | Perm |
| CMSPID        | Sponsor-Defined Identifier         | Char |         | Perm |
| CMMODIFY      | Modified Reported Name             | Char |         | Perm |
| CMSCAT        | Subcategory for Medication         | Char |         | Perm |
| CMOCCUR       | CM Occurrence                      | Char |         | Perm |
| CMSTAT        | Concomitant Medication Status      | Char |         | Perm |
| CMREASND      | Reason Medication Not Collected    | Char |         | Perm |
| CMCLAS        | Medication Class                   | Char |         | Perm |
| CMCLASD       | Medication Class Code              | Char |         | Perm |
| CMDOSE        | Dose per Administration            | Num  |         | Perm |
| CMDOSU        | Dose Units                         | Char |         | Perm |
| CMDOSFRM      | Dose Form                          | Char |         | Perm |
| CMDOSFRQ      | Dosing Frequency per Interval      | Char |         | Perm |
| CMDOSTOT      | Total Daily Dose Using CMDOSU      | Num  |         | Perm |
| CMDOSRGM      | Intended Dose Regimen              | Char |         | Perm |
| CMSTRF        | Start Relative to Reference Period | Char |         | Perm |
| CMENRF        | End Relative to Reference Period   | Char |         | Perm |
| VISITNUM      | Visit Number                       | Num  |         | Perm |

# SUPPCM

- ATC codes (1 to 10) and Level 1, 2, 3, and 4 decodes for each ATC code go in SUPPCM

| Variable Name | Variable Label              | Type | Mapping                          | Core |
|---------------|-----------------------------|------|----------------------------------|------|
| STUDYID       | Study Identifier            | Char | LEGACY_DEMOG.PNO                 | Req  |
| RDOMAIN       | Related Domain Abbreviation | Char | ='CM'                            | Req  |
| USUBJID       | Unique Subject Identifier   | Char | LEGACY_DEMOG.UPATNO              | Req  |
| IDVAR         | Identifying Variable        | Char | ='CMSEQ'                         | Req  |
| IDVARVAL      | Identifying Variable Value  | Char | %SUPPSEQ                         | Req  |
| QNAM          | Qualifier Variable Name     | Char | ='LEVEL1_1'                      | Req  |
| QLABEL        | Qualifier Variable Label    | Char | ='LEVEL 1 DECODE FOR ATC CODE 1' | Req  |
| QVAL          | Data Value                  | Char | LEGACY_CM.LEVEL1_1               | Req  |
| QORIG         | Origin                      | Char | ='DERIVED'                       | Req  |
| QEVAL         | Evaluator                   | Char |                                  | Perm |

# SUPPCM

- Prior medication flag

| Variable Name | Variable Label              | Type | Mapping              | Core |
|---------------|-----------------------------|------|----------------------|------|
| STUDYID       | Study Identifier            | Char | LEGACY_DEMOG.PNO     | Req  |
| RDOMAIN       | Related Domain Abbreviation | Char | = 'CM'               | Req  |
| USUBJID       | Unique Subject Identifier   | Char | LEGACY_DEMOG.UPA TNO | Req  |
| IDVAR         | Identifying Variable        | Char | = 'CMSEQ'            | Req  |
| IDVARVAL      | Identifying Variable Value  | Char | %SUPPSEQ             | Req  |
| QNAM          | Qualifier Variable Name     | Char | = 'PRIOR'            | Req  |
| QLABEL        | Qualifier Variable Label    | Char | = 'PRIOR MEDICATION' | Req  |
| QVAL          | Data Value                  | Char | LEGACY_CM.PREVMED    | Req  |
| QORIG         | Origin                      | Char | = 'DERIVED'          | Req  |
| QEVAL         | Evaluator                   | Char |                      | Perm |

# LB – Variables Mapped

| Variable Name | VariableLabel                           | Type | Comments  | Core |
|---------------|---|------|---|------|
| STUDYID       | Study Identifier                        | Char | LEGACY_DEMOG.PNO  | Req  |
| DOMAIN        | Domain Abbreviation                     | Char | = 'LB'  | Req  |
| USUBJID       | Unique Subject Identifier               | Char | LEGACY_DEMOG.UPATNO   | Req  |
| LBSEQ         | Sequence Number                         | Num  | %SEQ  | Req  |
| LBTESTCD      | Lab Test or Examination Short Name      | Char | LEGACY_LB.LABTST  | Req  |
| LBTEST        | Lab Test or Examination Name            | Char | LEGACY_LB.LABDESC   | Req  |
| LBCAT         | Category for Lab Test                   | Char | CASE WHEN LEGACY_LB.LABCAT='C' THEN 'CHEMISTRY' WHEN LEGACY_LB.LABCAT='H' THEN 'HEMATOLOGY' WHEN LEGACY_LB.LABCAT='U' THEN 'URINALYSIS' ELSE '' END | Exp  |
| LBORRES       | Result or Finding in Original Units     | Char | PUT(LEGACY_LB.RESULT,BEST8.)  | Exp  |
| LBORRESU      | Original Units                          | Char | LEGACY_LB.UNIT  | Exp  |
| LBORNRO       | Normal Range Lower Limit of Orig Units  | Char | PUT(LEGACY_LB.LORANGE,BEST8.)   | Exp  |
| LBORNRI       | Normal Range Upper Limit of Orig Units  | Char | PUT(LEGACY_LB.HIRANGE,BEST8.)   | Exp  |
| LBSTRESC      | Character Result/ Finding in Std Format | Char | PUT(LEGACY_LB.SIVAL,BEST8.)   | Exp  |
| LBSTRESN      | Numeric Result/ Finding in Std Units    | Num  | LEGACY_LB.SIVAL   | Exp  |
| LBSTRESU      | Standard Units                          | Char | LEGACY_LB.SIUNIT  | Exp  |
| LBSTNRLO      | Normal Range Lower Limit of Std Units   | Num  | =.  | Exp  |
| LBSTNRHI      | Normal Range Upper Limit of Std Units   | Num  | =.  | Exp  |
| LBNRIND       | Reference Range Indicator               | Char | LEGACY_LB.FLAG  | Exp  |
| LBBLFL        | Baseline Flag                           | Char | LEGACY_LB.BASEFLAG  | Exp  |
| VISITNUM      | Visit Number                            | Num  | LEGACY_LB.VISITNUM  | Req  |
| LBDC          | Date/Time of Specimen Collection        | Char | %ISODATE(LEGACY_LB.SAMPDT)  | Exp  |
| LB DY         | Study Day of Specimen Collection        | Num  | LEGACY_LB.STUDYDAY  | Perm |

# LB – Variables Not Mapped

| Variable Name | VariableLabel                           | Type | Comments | Core |
|---------------|---|------|----------|------|
| LBGRPID       | Group ID                                | Char |          | Perm |
| LBREFID       | Specimen ID                             | Char |          | Perm |
| LBSPID        | Sponsor-Defined Identifier              | Char |          | Perm |
| LBSCAT        | Subcategory for Lab Test                | Char |          | Perm |
| LBSTNRC       | Reference Range for Char Rslt-Std Units | Char |          | Perm |
| LBSTAT        | Lab Status                              | Char |          | Perm |
| LBREASND      | Reason Test Not Done                    | Char |          | Perm |
| LBNAM         | Vendor Name                             | Char |          | Perm |
| LBLOINC       | LOINC Code                              | Char |          | Perm |
| LBSPEC        | Specimen Type                           | Char |          | Perm |
| LBSPCCND      | Specimen Condition                      | Char |          | Perm |
| LBMETHOD      | Method of Test or Examination           | Char |          | Perm |
| LBDRVFL       | Derived Flag                            | Char |          | Perm |
| LBFAST        | Fasting Status                          | Char |          | Perm |
| LBTOX         | Toxicity                                | Char |          | Perm |
| LBTOXGR       | Standard Toxicity Grade                 | Char |          | Perm |
| VISIT         | Visit Name                              | Char |          | Perm |
| VISITDY       | Planned Study Day of Visit              | Num  |          | Perm |
| LBENDTC       | End Date/Time of Specimen Collection    | Char |          | Perm |
| LBTPPT        | Planned Time Point Name                 | Char |          | Perm |
| LBTPPTNUM     | Planned Time Point Number               | Num  |          | Perm |
| LBELTM        | Elapsed Time from Reference Point       | Char |          | Perm |
| LBTPTREF      | Time Point Reference                    | Char |          | Perm |

# SUPPLB

- Impossible value flag

| Variable Name | Variable Label              | Type | Mapping                  | Core |
|---------------|-----------------------------|------|--------------------------|------|
| STUDYID       | Study Identifier            | Char | LEGACY_DEMOG.PNO         | Req  |
| RDOMAIN       | Related Domain Abbreviation | Char | ='LB'                    | Req  |
| USUBJID       | Unique Subject Identifier   | Char | LEGACY_DEMOG.UPATNO      | Req  |
| IDVAR         | Identifying Variable        | Char | ='LBSEQ'                 | Req  |
| IDVARVAL      | Identifying Variable Value  | Char | %SUPPSEQ                 | Req  |
| QNAM          | Qualifier Variable Name     | Char | ='VALIDFG'               | Req  |
| QLABEL        | Qualifier Variable Label    | Char | ='IMPOSSIBLE VALUE FLAG' | Req  |
| QVAL          | Data Value                  | Char | LEGACY_LB.VALID          | Req  |
| QORIG         | Origin                      | Char | ='DERIVED'               | Req  |
| QEVAL         | Evaluator                   | Char | ='SPONSOR'               | Perm |

# SUPPLB

- Impossible value flag reason

| Variable Name | Variable Label              | Type | Mapping                          | Core |
|---------------|-----------------------------|------|----------------------------------|------|
| STUDYID       | Study Identifier            | Char | LEGACY_DEMOG.PNO                 | Req  |
| RDOMAIN       | Related Domain Abbreviation | Char | = 'CM'                           | Req  |
| USUBJID       | Unique Subject Identifier   | Char | LEGACY_DEMOG.UPATNO              | Req  |
| IDVAR         | Identifying Variable        | Char | = 'LBSEQ'                        | Req  |
| IDVARVAL      | Identifying Variable Value  | Char | %SUPPSEQ                         | Req  |
| QNAM          | Qualifier Variable Name     | Char | = 'VALIDCOM'                     | Req  |
| QLABEL        | Qualifier Variable Label    | Char | = 'IMPOSSIBLE VALUE FLAG REASON' | Req  |
| QVAL          | Data Value                  | Char | LEGACY_LB.COMMENT                | Req  |
| QORIG         | Origin                      | Char | = 'DERIVED'                      | Req  |
| QEVAL         | Evaluator                   | Char | = 'SPONSOR'                      | Perm |

# Summary

- Enter process took 5 months
- Resources:
  - Data Manager: 100%
  - Clinician: 25%
  - Programmer: 50%
  - Project Manager: 25%
  - Statistician: 100% (for one month)

# Discussion

- Messy legacy studies require a lot of data management expertise before automated methods can be used
  - A range of problems in the legacy data had to be resolved to make the data usable prior to mapping to SDTM
- While the approach used in cleaning the conmed and lab data might be applicable to other types of data, clearly caution must be taken, as each type of data typically has its own set of unique problems
- Legacy conversion to SDTM may appear on the surface to be costly, however, non-conversion can be more costly
  - The ability to respond to regulatory requests can be greatly reduced if working with multiple “old” databases in varying structures.