Complicated Data?
SAS User Formats Will Simplify Your Code
Complicated Data? SAS User Formats Will Simplify Your Code
Margaret Burgess and Stephanie Nelson
Optum

ABSTRACT
The International Classification of Diseases (ICD) is used for medical coding in the U.S. It includes medical diagnosis
coding and inpatient hospital procedure coding. On October 1, 2015, health care professionals changed from using the
9th revision (ICD-9) to the 10th revision (ICD-10). When analyzing administrative claims data for a research study that
includes both ICD-9 and ICD-10 codes and a time period that spans over October 1, 2015, we have identified
three main areas that present challenges. Below are examples of these three challenges and our suggested solutions to
overcome them.

CHALLENGE 1
The same diagnosis code value can appear in ICD-9 and ICD-10 with different meanings. An example is shown in Table 1
below.

<table>
<thead>
<tr>
<th>Code</th>
<th>Version</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>E8981</td>
<td>ICD-9</td>
<td>ACCIDENT CAUSED BY OTHER BURNING MATERIALS</td>
</tr>
<tr>
<td>E8981</td>
<td>ICD-10</td>
<td>POSTPROCEDURAL HEMORRHAGE OF AN ENDOCRINE SYSTEM OR STRUCTURE FOLLOWING A PROCEDURE</td>
</tr>
<tr>
<td>V293</td>
<td>ICD-9</td>
<td>OBSERVATION FOR SUSPECTED GENETIC OR METABOLIC CONDITION</td>
</tr>
<tr>
<td>V293</td>
<td>ICD-10</td>
<td>MOTORCYCLE RIDER (DRIVER) (PASSENGER) INJURED IN UNSPECIFIED NONTRAFFIC ACCIDENT</td>
</tr>
</tbody>
</table>

CHALLENGE 2
Specifying a range of ICD procedure codes can return inaccurate results. Procedure codes, which are character values,
are sorted by SAS as character values. Consequently the ICD-9 range ‘0015’ – ‘0016’ shown in Table 2, would now
include a long list of ICD-10 procedure codes, some of which are displayed in Table 3.

<table>
<thead>
<tr>
<th>Code</th>
<th>Version</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0015</td>
<td>ICD-9</td>
<td>HIGH-DOSE INFUSION INTERLEUKIN-2 [IL-2]</td>
</tr>
<tr>
<td>0016</td>
<td>ICD-9</td>
<td>PRESSURIZED TREATMENT OF VENOUS BYPASS GRAFT [CONDUIT] WITH PHARMACEUTICAL SUBSTANCE</td>
</tr>
<tr>
<td>0017</td>
<td>ICD-9</td>
<td>INFUSION OF VASOPRESSOR AGENT</td>
</tr>
</tbody>
</table>

CHALLENGE 3
A claims analysis study that includes both ICD-9 and ICD-10 codes can lead to overly complicated programming logic.
This complicated logic includes “if-then-else” analysis along with the interrogation of code versions, to insure proper
interpretation of the claims data.

If diag_ver_cd_1='9' then do……; end;
Else if diag_ver_cd_1='10' then do……; end;

Tables 1, 2, and 3 provide examples of code versions and their definitions.
Complicated Data? SAS User Formats Will Simplify Your Code
Margaret Burgess and Stephanie Nelson
Optum

METHODS

Our first suggestion is to derive a new variable and add it to your analytic SAS dataset. This is illustrated below in Table 4 with a diagnosis code. This new variable value holds the original diagnosis code, prefixed with a version identifier.

<table>
<thead>
<tr>
<th>Version</th>
<th>Original Diagnosis Code</th>
<th>New Variable – Diagnosis Code with a Version Prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9</td>
<td>E8981</td>
<td>D09_E8981</td>
</tr>
<tr>
<td>ICD-10</td>
<td>E8981</td>
<td>D10_E8981</td>
</tr>
</tbody>
</table>

Creating this new variable now allows the programmer to create one SAS user Format, and bypass the more complicated logic required using the correct SAS user format with the corresponding diagnosis code version.

PROC FORMAT;
VALUE $MYFMT
‘D09_E8981’ = ‘ACCIDENT CAUSED BY OTHER BURNING MATERIALS’
‘D10_E8981’ = ‘POSTPROCEDURAL HEMORRHAGE OF AN ENDOCRINE SYSTEM ORGAN OR STRUCTURE FOLLOWING A PROCEDURE’
‘D09_V293’ = ‘OBSERVATION FOR SUSPECTED GENETIC OR METABOLIC CONDITION’
‘D10_V293’ = ‘MOTORCYCLE RIDER (DRIVER) (PASSENGER) INJURED IN UNSPECIFIED NONTRAFFIC ACCIDENT’
‘P09_0015’ = ‘HIGH-DOSE INFUSION INTERLEUKIN-2 [IL-2]’
‘R_0499’ = ‘AMBULATORY SURGICAL CARE-OTHER AMBULATORY SURGICAL CARE’
‘H_S0187’ = ‘TAMOXIFEN CITRATE, ORAL’
;

METHODS (continued)

In other words, this:

```sas
if code_type = "9" then do;
category = put(diag, $fmtnine.);
end;
else if code_type = "10" then do;
category = put(diag, $fmtten.);
end;
category = put(diag, $myfmt.);
```

Becomes this simpler logic without the need for if –then –else logic:

```sas
category = put(diag, $myfmt.);
```
Complicated Data? SAS User Formats Will Simplify Your Code
Margaret Burgess and Stephanie Nelson
Optum

RESULTS

1. Using a combination of a new variable (code value with prefixed version) along with a SAS user format removes any ambiguity in interpreting claims data containing a mix of ICD-9 and ICD-10 codes.

2. Analyzing claims data using ranges of alpha numeric values across versions can still be used without the risk of including unwanted values in your specified ranges.

3. Incorporating the version along with the code value, in a new variable, reduces the complexity of SAS coding required.

CONCLUSION

SAS user formats can be used to:

- Simplify programming logic
- Remove ambiguity in code sets that vary across versions
- Generally make writing SAS programs and understanding SAS programs used to analyze claims data contain a mix of codes and code versions, much easier.
SAS® GLOBAL FORUM 2018

April 8 - 11 | Denver, CO
Colorado Convention Center

#SASGF