Efficiently Converting Datasets to Vertical Structure using Metadata

Thevaki Mahadevan, Syne Qua Non, Norfolk, UK

Introduction

- Raw Data extracted from CRF in horizontal data structure gets converted to vertical structure by programmers for analysis purposes.
- By utilizing variable prefixes and metadata, via the proc contents procedure, horizontal datasets can then be converted into vertical structure via macro automation.
- Generally proc transpose or arrays are used for this purpose, but the advantages of the proc contents approach is that it avoids the need for manually entering each variable and their labels, hence saving time and improving quality and supporting traceability, in particular, when there are many parameters to transpose.

Trends in aCRF Data Mapping

- Variable names within a domain tend to have the same prefix letters. e.g. AIM domain shown in the aCRF page has the same prefix letters, AIMxxx, for all the domain specific variables.

Horizontal RAW Dataset

- As mapped in the aCRF, the same variable names are seen in the horizontally structured raw dataset and support an understanding of the relationship of aCRF to its raw dataset, aiding traceability.
- Generally as part of the mapping process, each variable collected is formatted to either two or three forms; the numeric, character and long/full (decoded/formatted) version.

Vertical Analysis Dataset

- The disadvantages of these approaches include:
  1. Requiring manual entry of variables and their attributes
  2. Will consume more time
  3. Has more chances for human error
  4. Transparency of the data is manually influenced

Metadata transposing approach minimises these disadvantages.

Typical Transposing Approaches

1. proc transpose
2. Arrays

An Array- Approach

- Generally proc transpose or arrays are used for this purpose
- The numeric, character and long/full (decoded/formatted) version.

Proc transpose- Approach

- In general proc transpose or arrays are used for this purpose
- As mapped in the aCRF page has the same prefix letters, AIMxxx, for all the domain specific variables.

Vertical RAW Dataset

- As mapped in the aCRF, the same variable names are seen in the horizontally structured raw dataset and support an understanding of the relationship of aCRF to its raw dataset, aiding traceability.

Conclusion

- Several ways to transpose a Horizontally structured dataset to Vertical but they all require a good understanding of the dataset metadata.
- Several ways to transpose a Horizontally structured dataset to Vertical but they all require a good understanding of the dataset metadata.

When there are many variables to transpose, a data dependent automated approach, has many advantages over other manually inputted approaches.