


# 2008 USMHS Correspondence Analyses

Mette Posamentier, PhD, UTSW  
Hervé Abdi, PhD, UTD  
Robert Haley, MD, UTSW



## Correspondence Analyses

- Correspondence analysis is an exploratory data analytic technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns.
- As opposed to traditional hypothesis testing designed to verify *a priori* hypotheses about relations between variables, exploratory data analysis is used to identify systematic relations between variables when there are no (or rather incomplete) *a priori* expectations as to the nature of those relations.
- Correspondence analysis is also a (*multivariate*) *descriptive data analytic* technique. Even the most commonly used statistics for simplification of data may not be adequate for description or understanding of the data. Simplification of data provides useful information about the data, but that should not be at the expense of valuable information. Correspondence analysis remarkably simplifies complex data and provides a detailed description of practically every bit of information in the data, yielding a simple, yet exhaustive analysis.
- Correspondence analysis has several features that distinguish it from other techniques of data analysis. An important feature of correspondence analysis is the multivariate treatment of the data through simultaneous consideration of multiple categorical variables. The multivariate nature of correspondence analysis can reveal relationships that would not be detected in a series of pair wise comparisons of variable.

## Correspondence Analyses

- Another important feature is the graphical display of row and column points in biplots, which can help in detecting structural relationships among the variable categories and objects (*i.e.*, cases). Finally, correspondence analysis has highly flexible data requirements. The only strict data requirement is a rectangular data matrix with non-negative entries.
- A distinct advantage of correspondence analysis over other methods yielding joint graphical displays is that it produces two dual displays whose row and column geometries have similar interpretations, facilitating analysis and detection of relationships. In other multivariate approaches to graphical data representation, this duality is not present.
- *In a nutshell*, correspondence analysis (CA) may be defined as a special case of principal components analysis (PCA) of the rows and columns of a table, especially applicable to a cross-tabulation. However CA and PCA are used under different circumstances. Principal components analysis is used for tables consisting of continuous measurement, whereas correspondence analysis is applied to contingency tables (*i.e.* cross-tabulations). Its primary goal is to transform a table of numerical information into a graphical display, in which each row and each column is depicted as a point.
- Correspondence analysis shows how the variables are related, not just that a relationship exists.
- Extension: Discriminant Correspondence Analyses developed by Dr. Abdi.

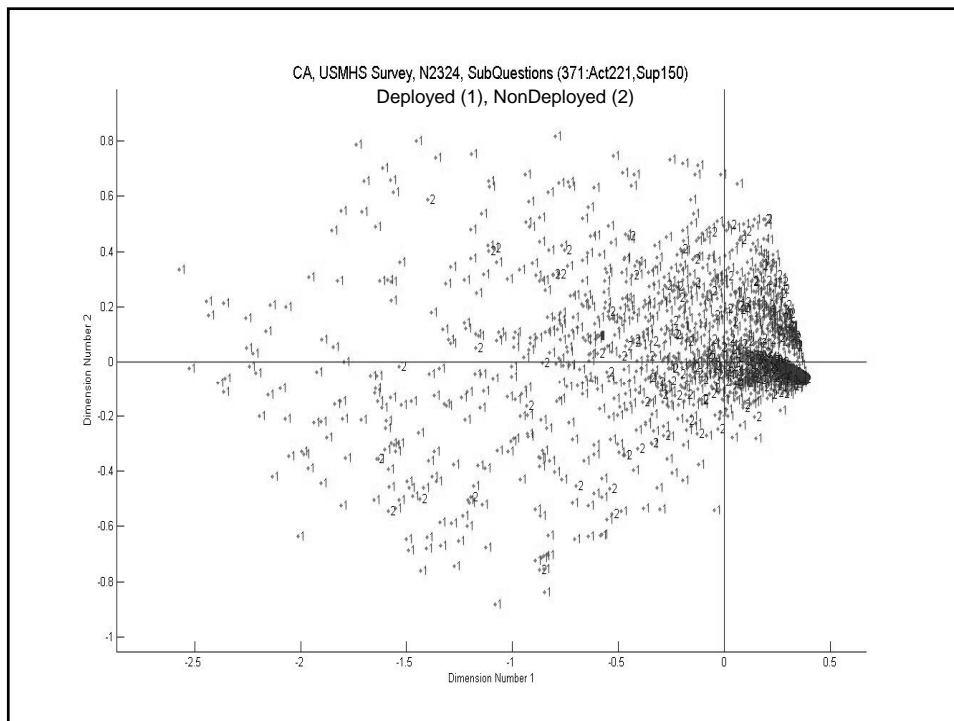
## Multiple Correspondence Analysis

S31e	S31f	S31g	S31h	S31j	S31k	S31o	S31p	S31r	S31u	S32a	S32b
0	1	1	1	1	1	1	1	1	1	0	1
0	1	0	1	0	0	1	1	1	1	0	1
1	1	1	1	0	0	0	0	0	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
0	1	0	0	1	1	1	1	1	1	0	1
0	1	0	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	1	1	1	1	1	1
0	1	1	1	0	0	1	1	1	0	1	1
0	1	0	1	1	1	1	1	1	1	1	1
0	1	0	1	0	0	1	1	1	1	1	1
1	1	0	0	1	1	1	1	1	0	1	1
1	1	0	1	0	0	1	1	1	1	1	0
0	1	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	0	0	1	1
1	1	1	1	0	0	1	1	0	0	1	1
0	1	1	1	1	1	1	1	1	1	0	0
0	1	0	0	1	1	0	1	1	1	1	0
0	1	0	0	0	1	1	1	1	1	0	1
0	0	1	1	1	1	0	1	0	1	1	0
0	0	0	1	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	1	0	0	1
0	1	0	0	0	0	1	1	1	1	1	1
0	1	0	0	1	1	0	1	0	1	0	1
0	1	1	1	1	1	1	1	1	0	1	0
1	1	0	1	1	1	1	1	1	1	1	1

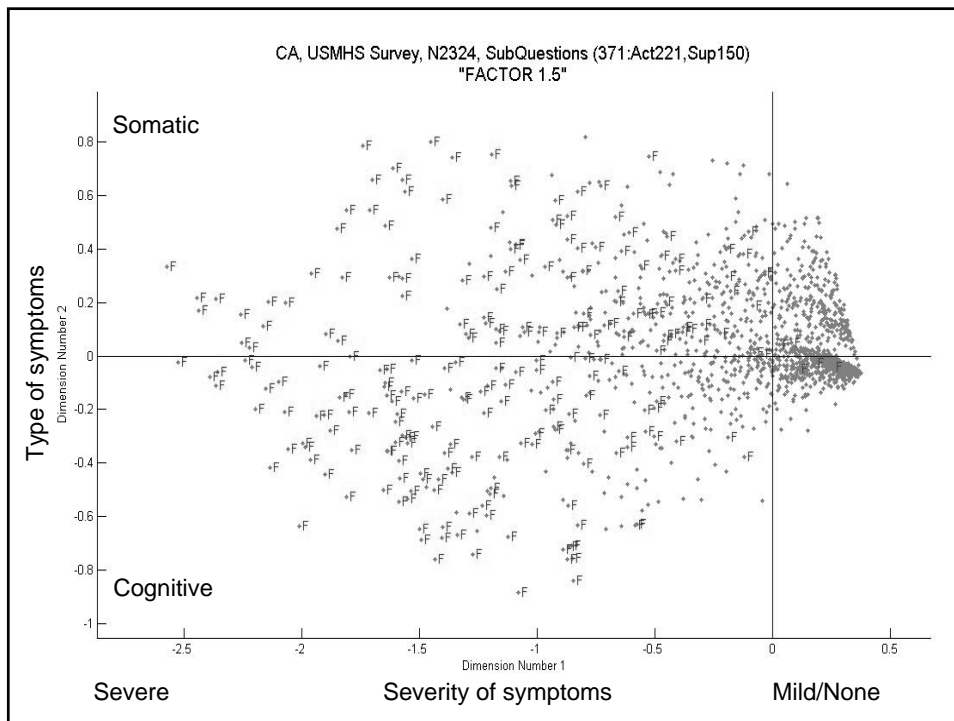
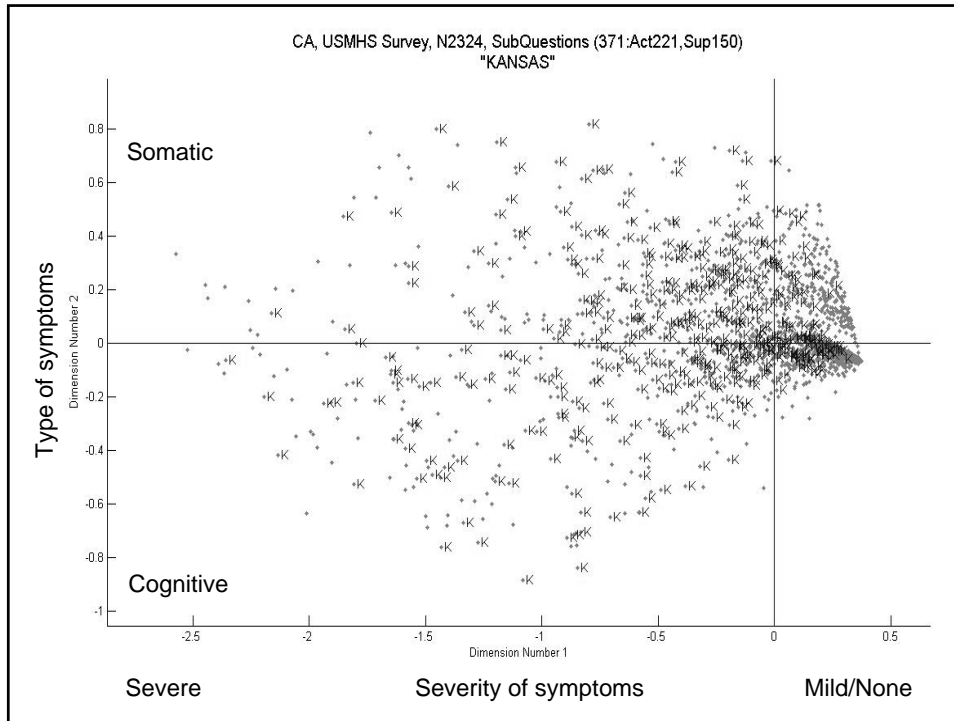
Matrix size: 2324 rows, 742 columns

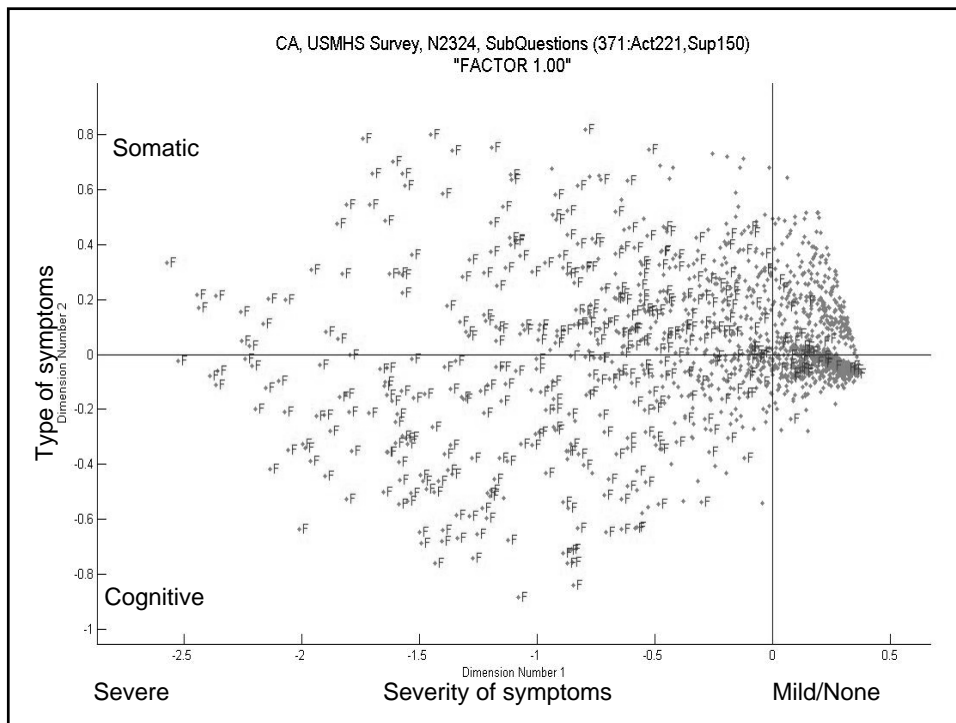
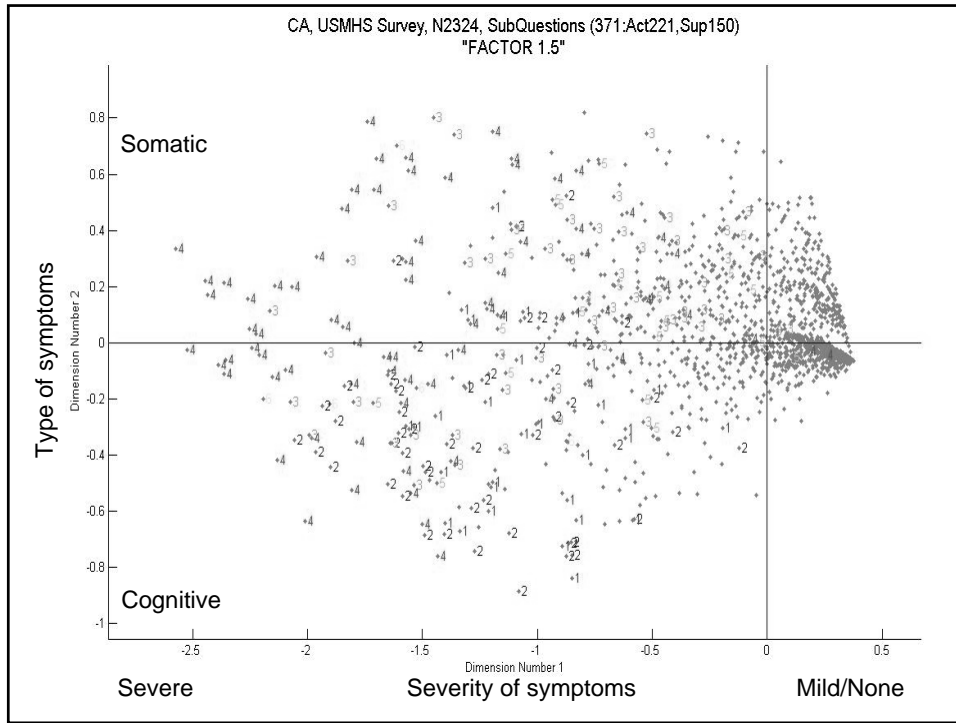
## Multiple Correspondence Analysis

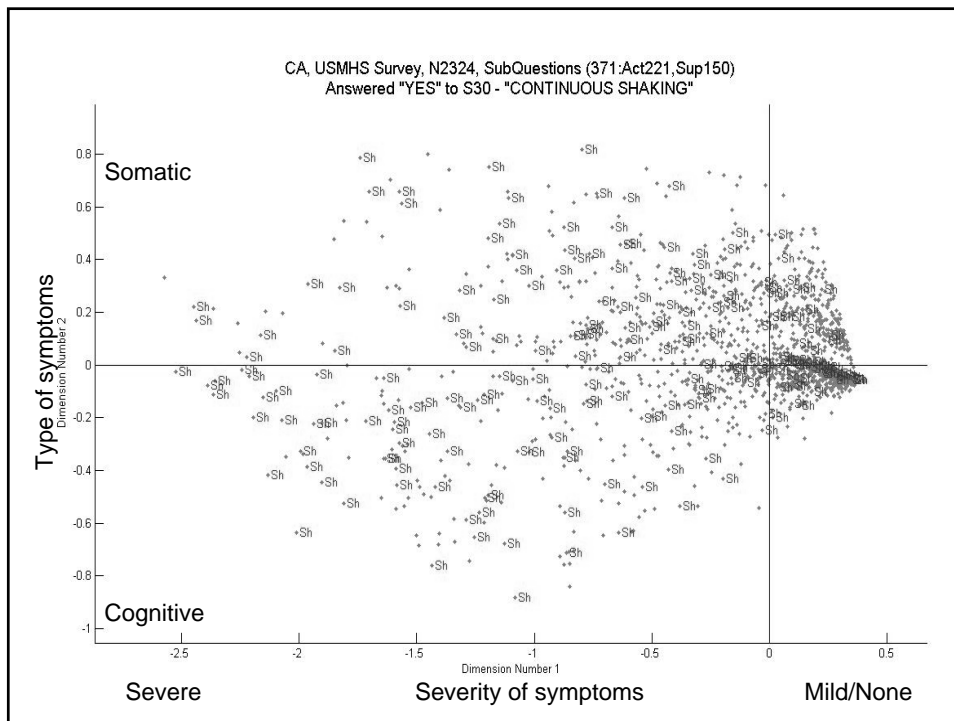
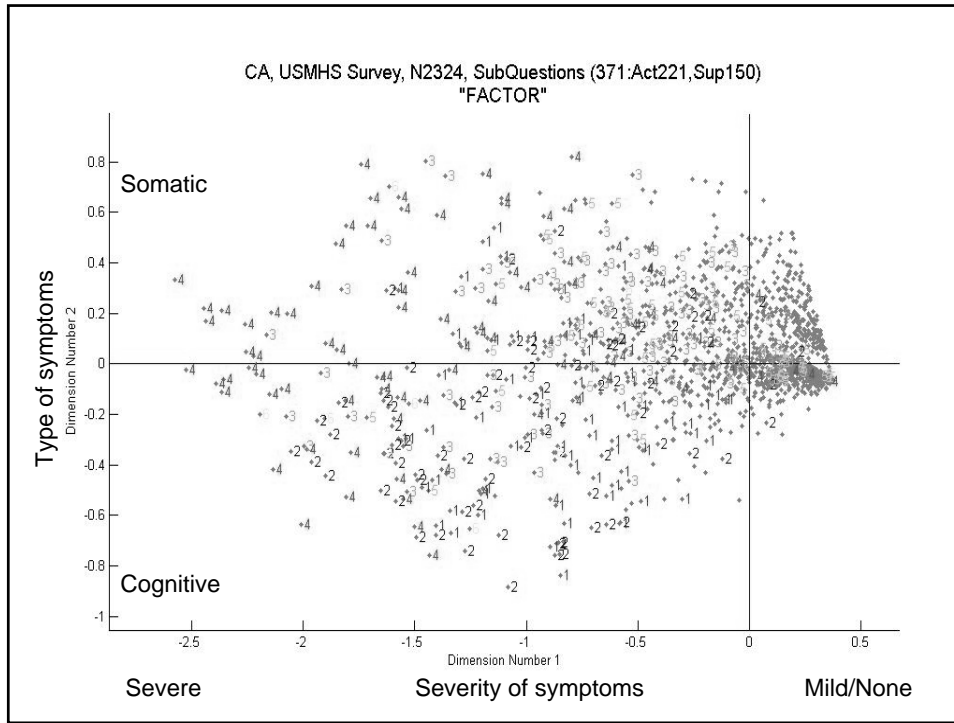
- Sub-questions or follow-up questions on USHMS
- 371 questions
  - Questions endorsed by <5% of respondents were coded as “supplementary”.
- 1 if respondent answered “YES”
- 0 if respondent answered “NO”
- 2324 respondents
  - Random 1/3 of survey respondents





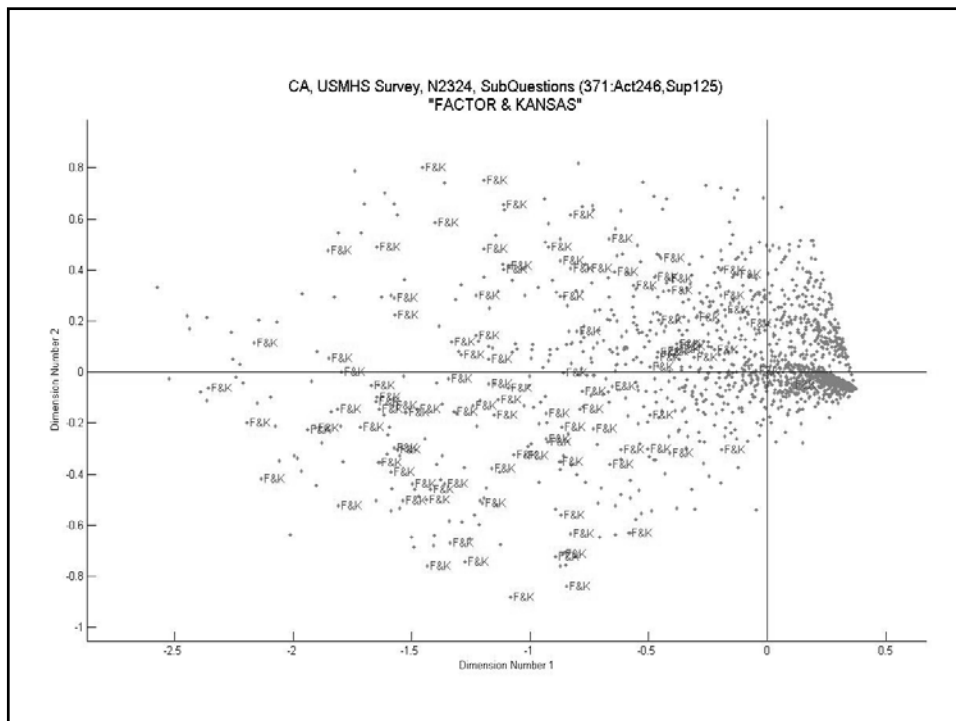




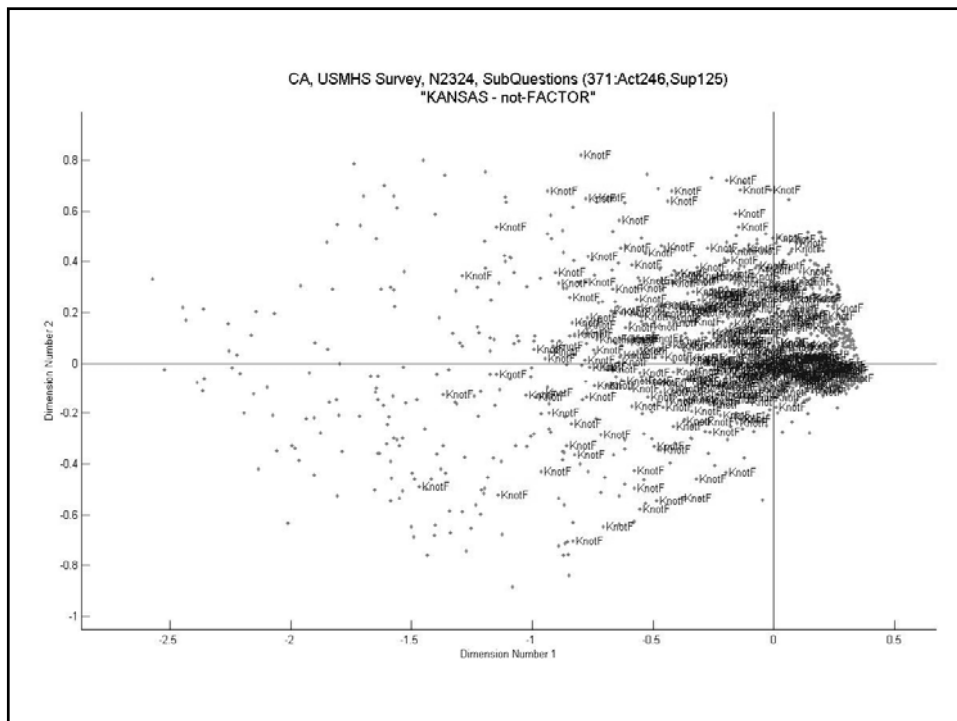
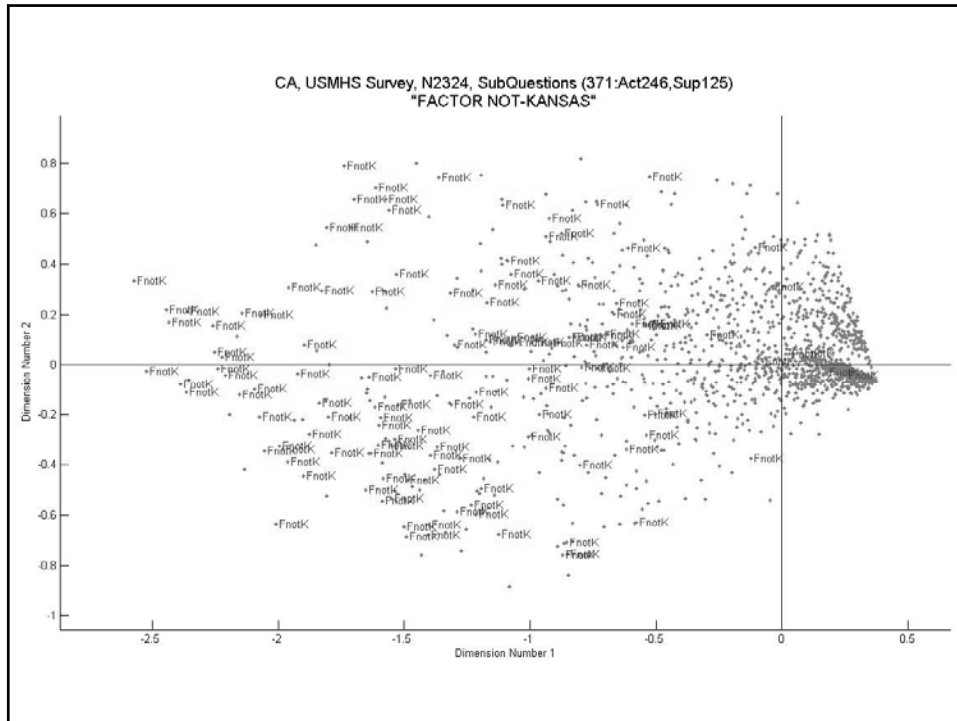


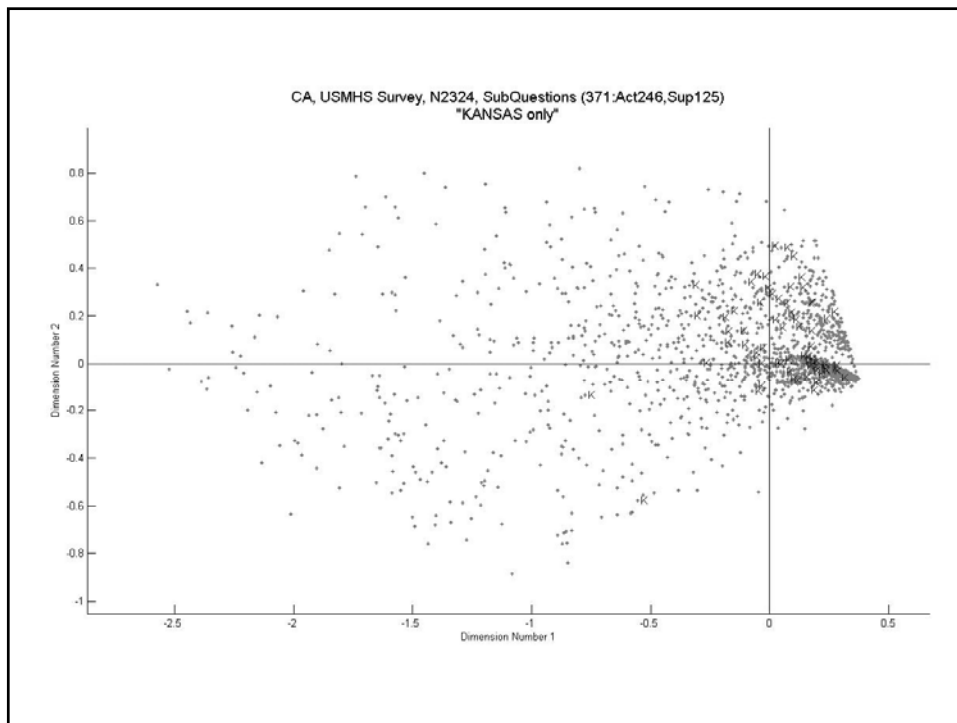
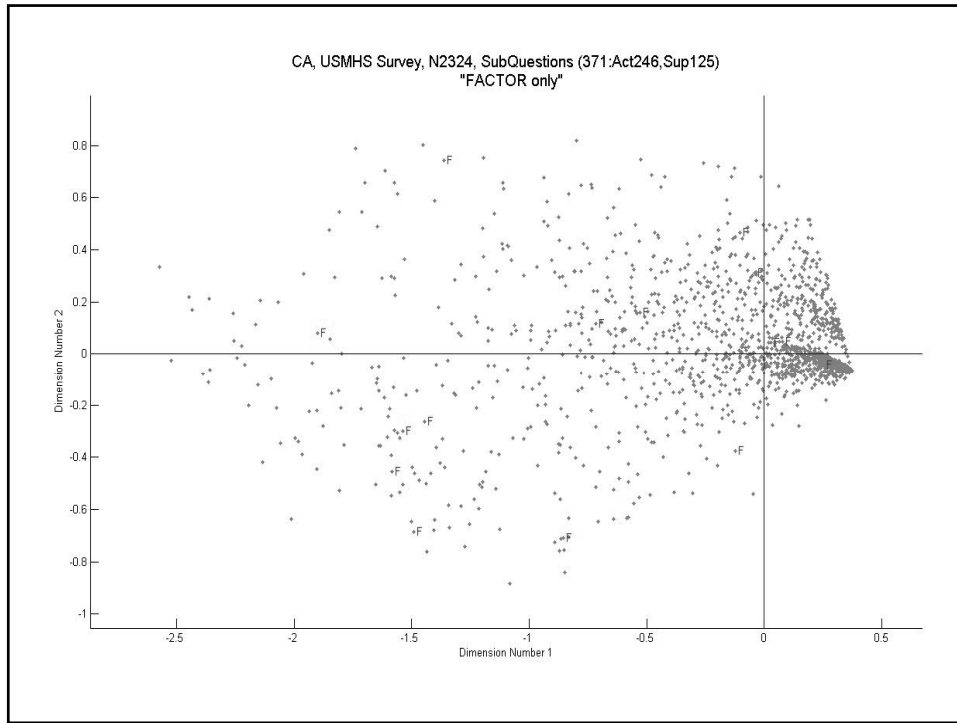
## Comparison of Surveys

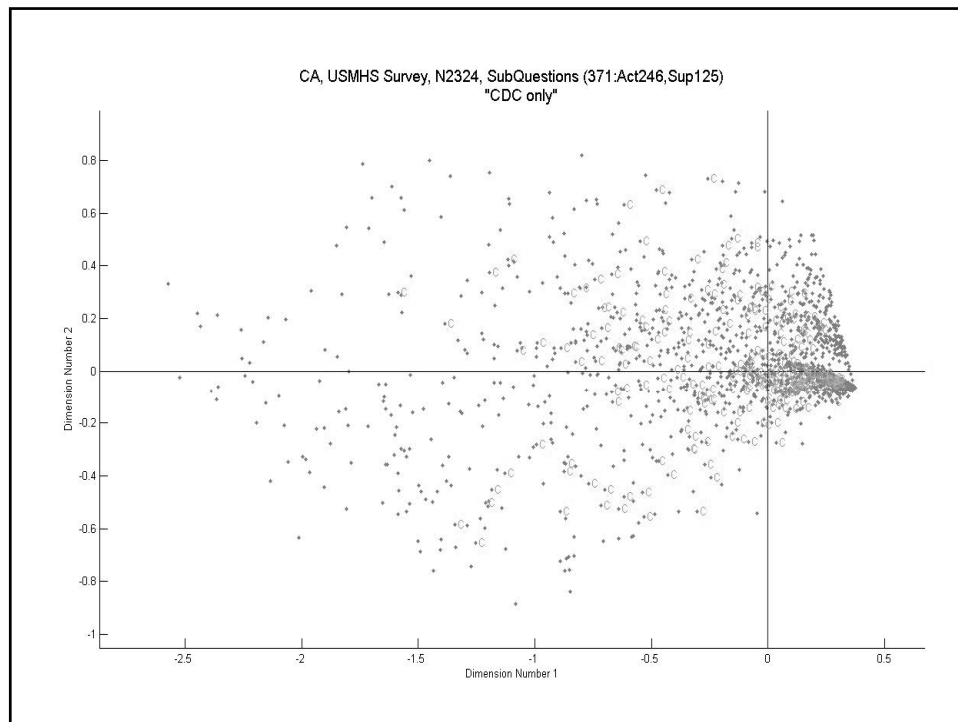
- Factor (Haley)
- Kansas (Steele)
- CDC (Fukuda)
- Factor & Kansas
- Factor, notKansas
- Kansas, notFactor
- Factor only
- Kansas only
- CDC only







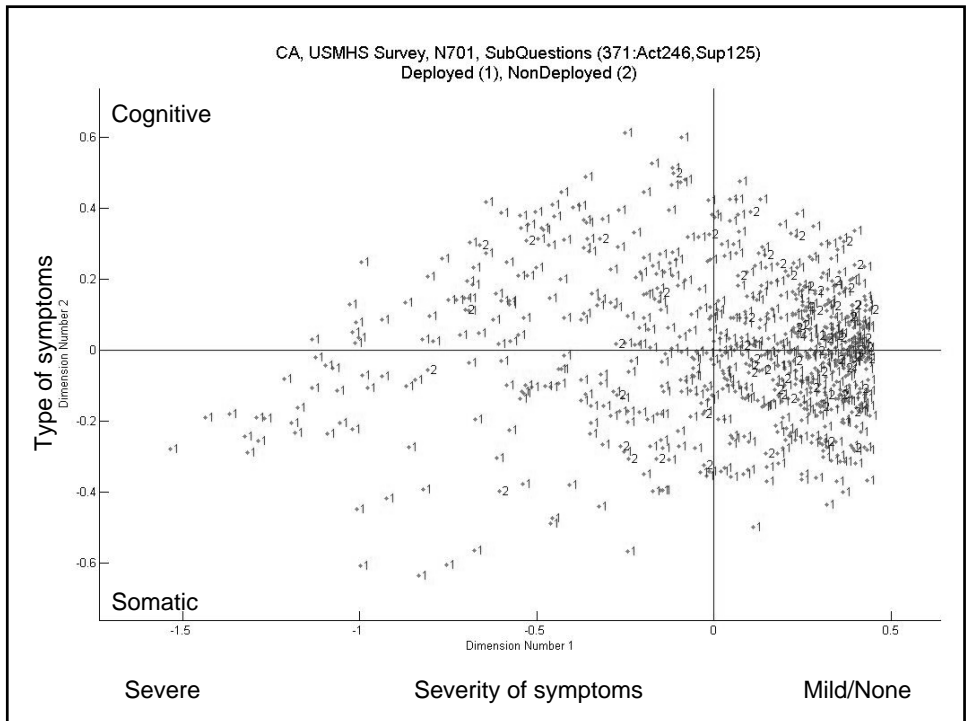
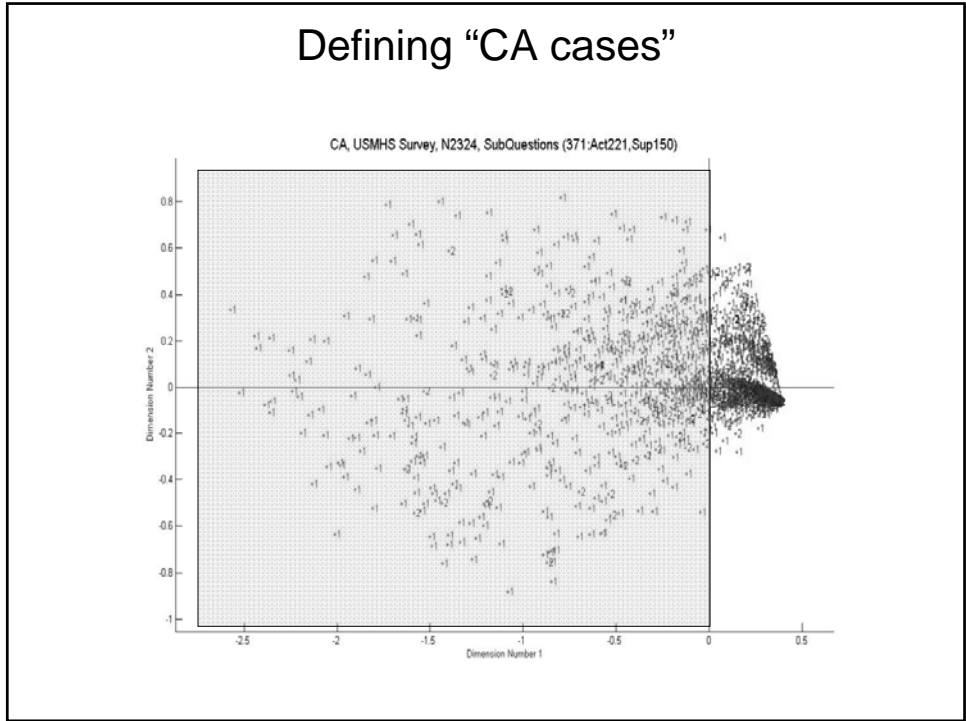


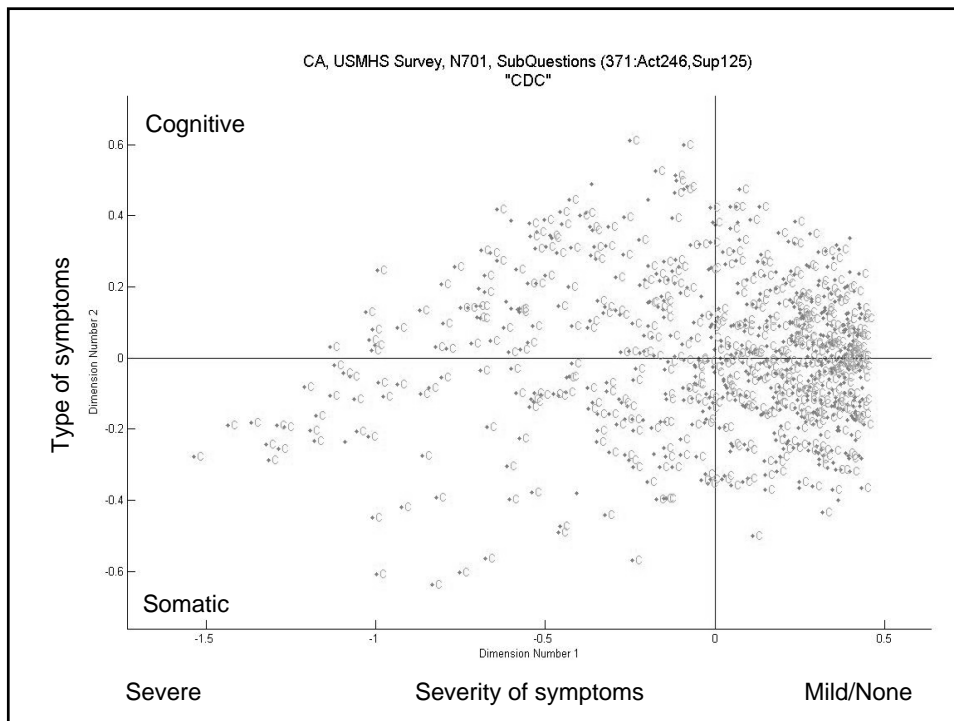
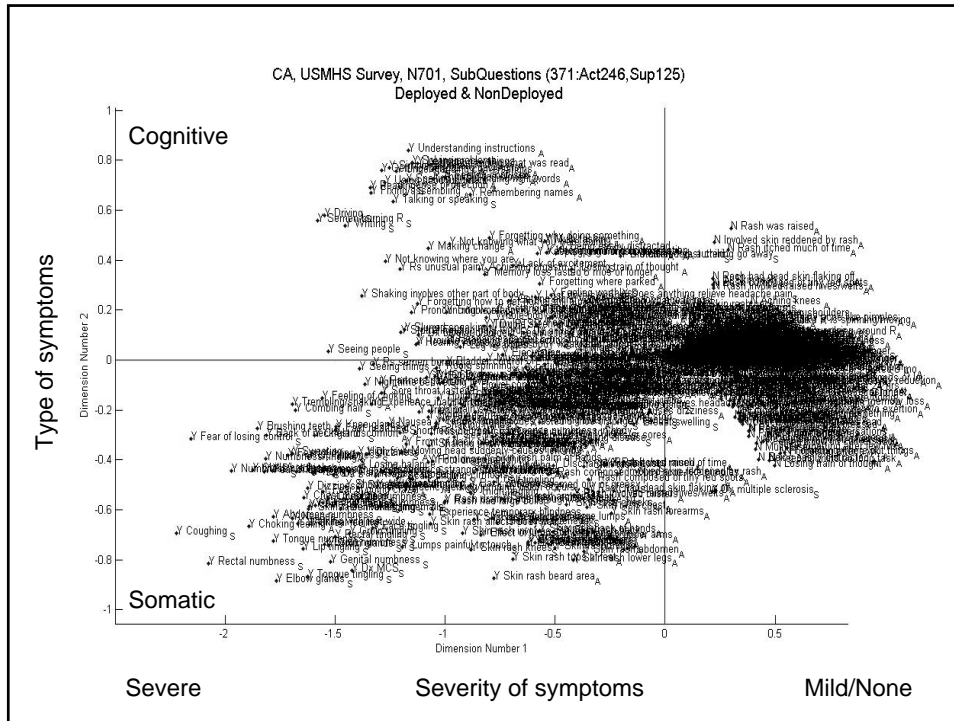


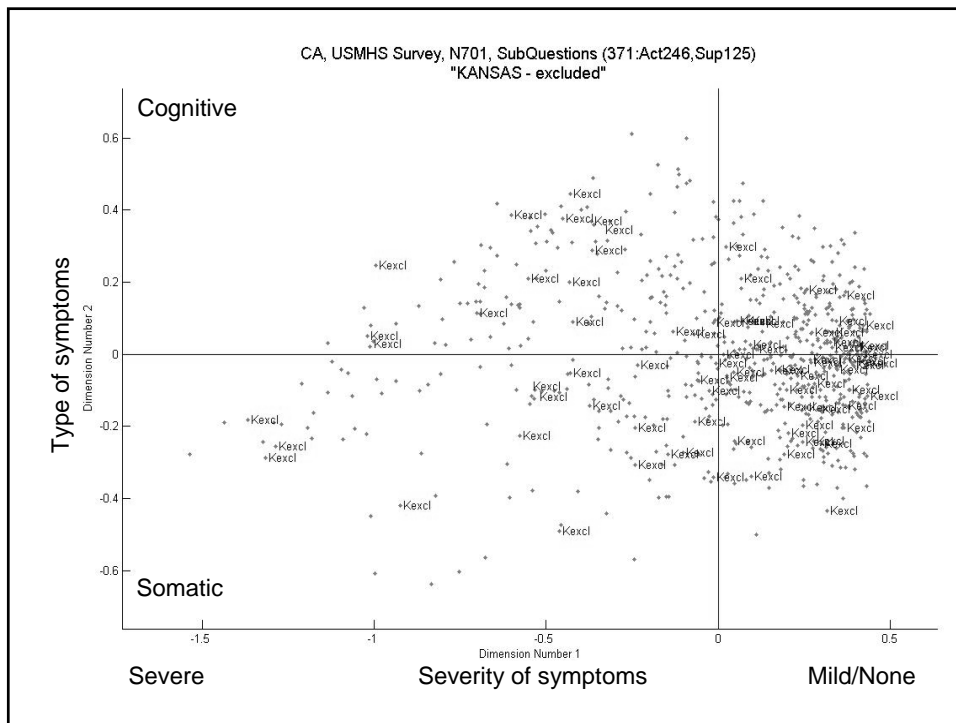
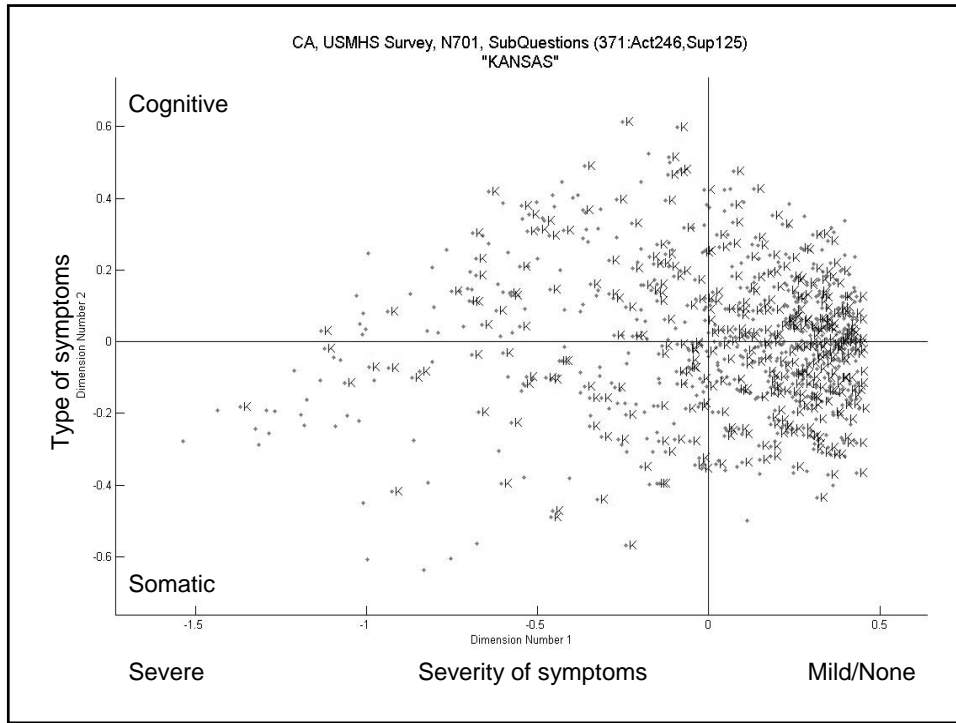
## Multiple Correspondence Analysis

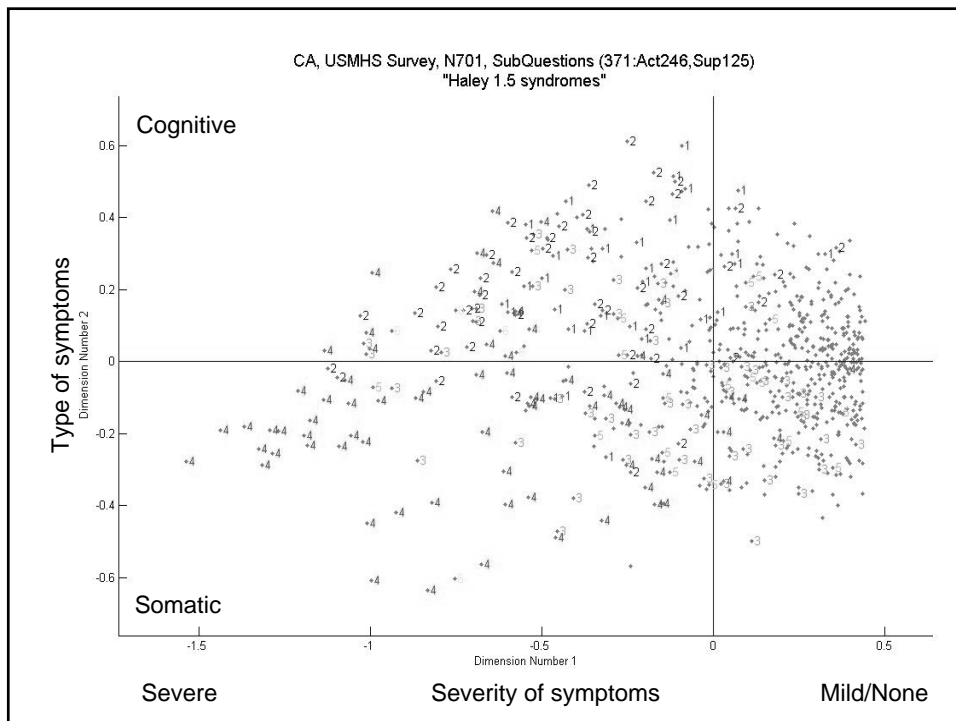
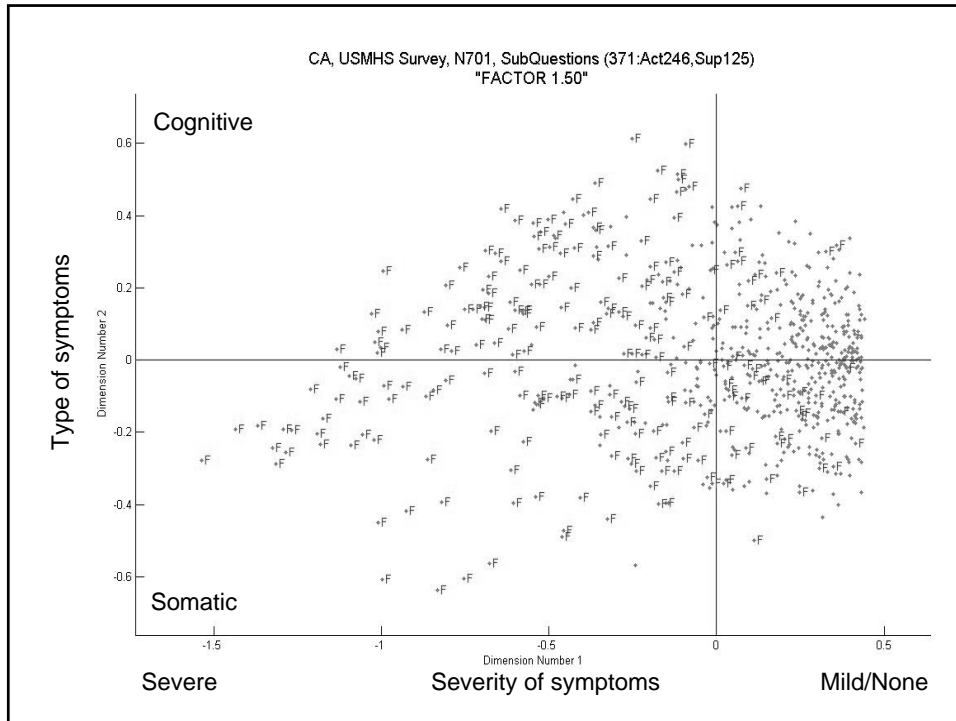
- 701 respondents
  - CA defined cases
  - Negative loading of MCA of larger sample
- Sub-questions or follow-up questions on USHMS
- 371 questions
- 1 if respondent answered "YES"
- 0 if respondent answered "NO"

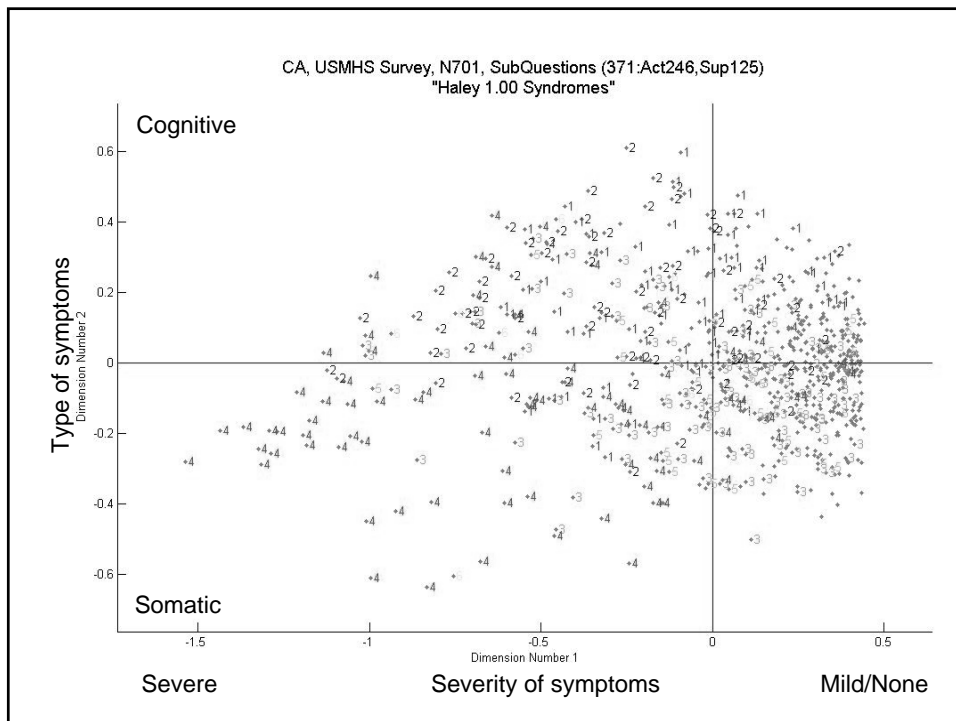
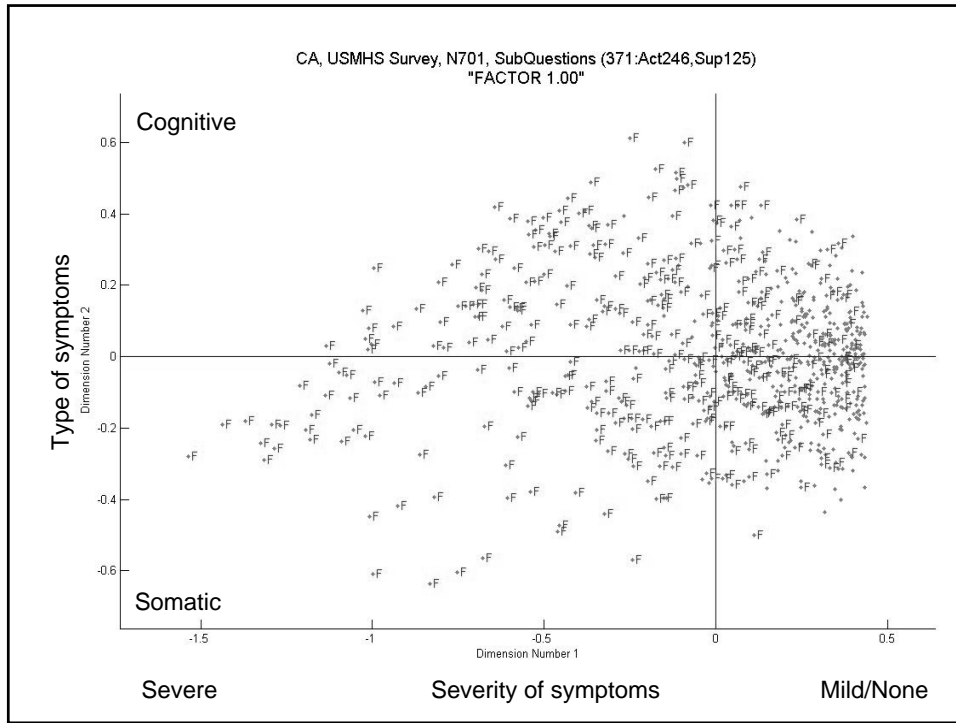
## Defining "CA cases"



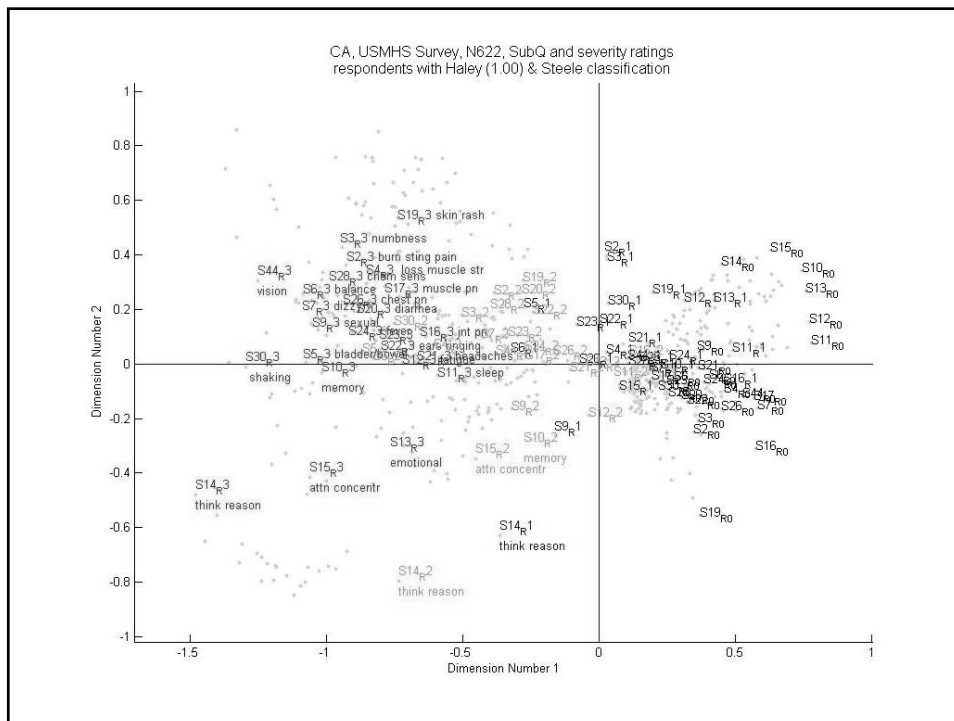
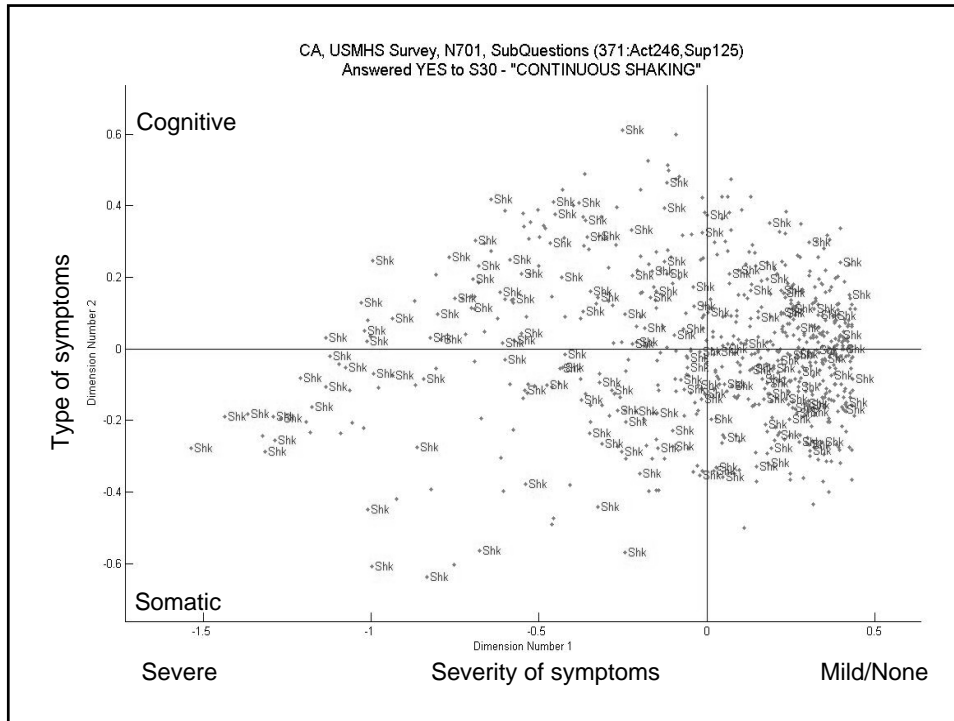


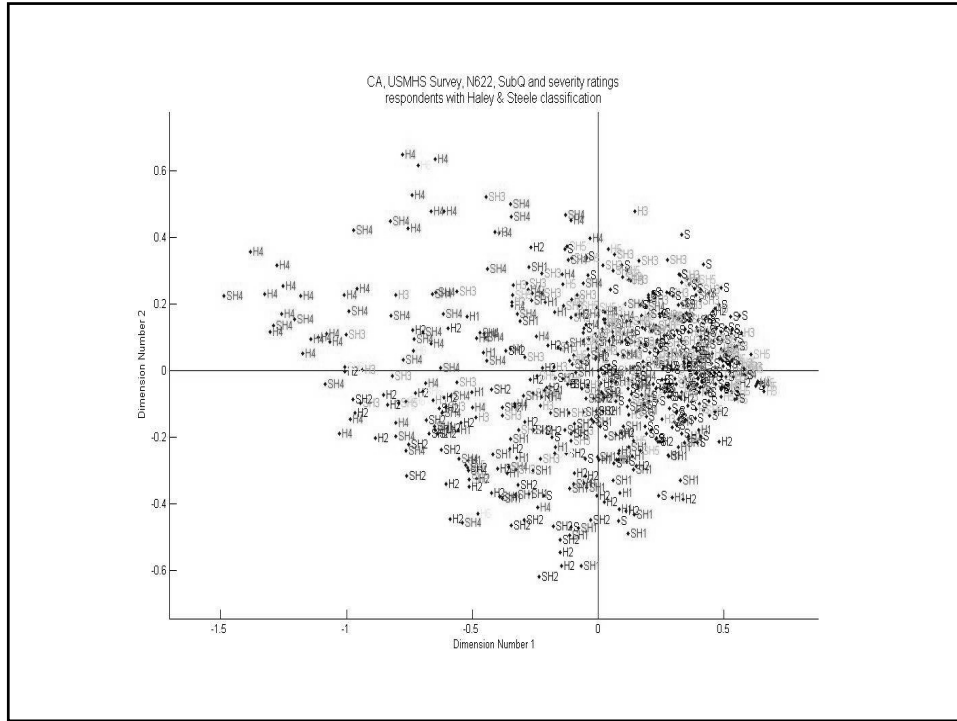












**EXPOSURE  
RISK FACTORS**

