Cluster Analysis of Service Data for Managed Care & Mental Health Systems Evaluation

H. Stephen Leff, Ph.D.
The Evaluation Center@HSRI

Marcus Lieberman
Responsive Methodology

Matthew M. Wise, M.P.H.
The Evaluation Center@HSRI

@address *the Evaluation Center@HSRI, 2336 Massachusetts Avenue, Cambridge, MA 02140*
@telephone *(617)876-0426* @fax *(617)492-7401* @email *TECENTER@hsri.org*

This publication was made possible by grant number SM97 001 from the Substance Abuse and Mental Health Services Administration's Center for Mental Health Services. These contents are solely the responsibility of the authors and do not necessarily represent the official views of the Substance Abuse and Mental Health Services Administration's Center for Mental Health Services.

Cluster Analysis of Service Data: An Introduction

Purpose

The purpose of this short working paper is to introduce the statistical procedure known as cluster analysis and illustrate its application to analyzing mental health service utilization data. While this exposition may not be sufficient to understand all the complexities of the procedure, it will enable people with a knowledge of elementary statistics to use software packages intelligently and perhaps, if they are motivated, to read one or more of the several texts on cluster analysis listed in the bibliography. This working paper is a document in process and we invite comments.

Why Use Cluster Analysis with Services Data?

Cluster analysis can be used to group persons into homogeneous subgroups based on similarities across variables. In this monograph, we illustrate how cluster analysis can be used to group persons based on the service packages they received. In the past, cluster analysis in mental health research primarily has been used to classify persons in terms of symptoms, traits or other intra-individual characteristics (Double, 1991; Furukawa, et al., 1992).

The service packages persons receive are important to study for several reasons. First, they are one indicator for assessing quality of care. Second, they are a starting point for designing benefit packages for health care planning. Finally, they can be used to measure the strength and integrity of interventions in outcomes research (Sechrest, et al., 1979). These are areas of research that have taken on additional importance with the emergence of managed care and health care reform.

Recently several researchers have used cluster methods to analyze mental health service data. Roth and colleagues (1992) cluster analyzed data on annual service utilization for persons in the Ohio public mental health system. Leff (in press) cluster analyzed monthly service utilization data for persons in capitated and fee for service programs in a public mental health system. Fisher and Altaffer (personal communication) analyzed annual service utilization data for persons served by another public mental health system. Leff and Wise (in press) applied cluster analysis to data on both services utilized and recommended in a public mental health system.

What is Cluster Analysis?

Cluster analysis is a data reduction technique. However, unlike factor analysis which produces underlying dimensions for a set of *variables*, usually of a much lower number than the original number of variables, cluster analysis is used most often to group *persons* who have similar profiles across a set of variables. Then these groups of persons may be compared on demographics, outcomes or program variables.

Typically, cluster analysis has been used in mental health research to classify persons in terms of attributes or symptoms. More recently, cluster analysis is being applied to mental health service utilization data which is the particular focus of this paper.

An Example

We will illustrate our discussion of cluster analysis using data from a statewide service planning and evaluation survey. In this example, recipients of mental health services were sampled from across a state. Case managers or other service providers (raters) completed forms for sampled service recipients and provided data on recipient demographics and clinical history as well as the services received and those ideally needed by these persons during a one month study period.

Some studies have clustered service data aggregated over longer periods of time (e.g., one year) (Roth et al., 1992). To the extent that the status of mental health service recipients varies over time, we would expect them to receive different services. Therefore, clusters for different periods of time should differ in their contents for persons who change in mental health status. Clusters covering longer time periods should be related to trait measures and variables like course, whereas clusters covering shorter periods of time should be related to state measures like level of functioning.

The study included data on 6,689 service recipients and 50 services. The list of services was based on a taxonomy of services maintained by the state mental health agency (SMHA) For each service, raters indicated the amounts of service recipients received and ideally needed (i.e., should have received). All amounts of service were converted into hours. Seventeen of the services were neither prescribed nor provided to any persons and were not considered for the cluster analysis, leaving thirty three services with data for at least some persons.

Analysis Procedure

Table 1 is an overview of the steps in a cluster analysis of service data. The points in Table 1 are discussed in detail below.

Table 1
Overview of Steps in Cluster Analysis of Service Data

1. **Examine the frequency distributions of the service data**
2. **Reduce the service data**
3. **Choose a cluster method**
4. **Look for solutions giving clusters that are large enough for use in additional analyses**
5. **Perform one-way analysis of variance and multiple post-hoc comparisons to identify services that distinguish between clusters in order to interpret clusters**
6. **Compare the cluster results with the clusters resulting from another clustering method**
7. **Compare with theory based expectations**
8. **Test the stability of the solution by comparing cluster results for randomly selected subsamples**
9. **Examine the associations between cluster membership and other variables**

1. Examine the frequency distributions of all data.

Figure 1 shows the distributions of two of the several service variables to be included in the cluster analysis. It is apparent that the distributions are not normal, but resemble a "backwards J," where many people receive none of the service and a small proportion of outliers receives some amount of the service. This type of distribution is characteristic of service data. A method for clustering persons based on service data should produce meaningful clusters even when the data are distributed in this manner.

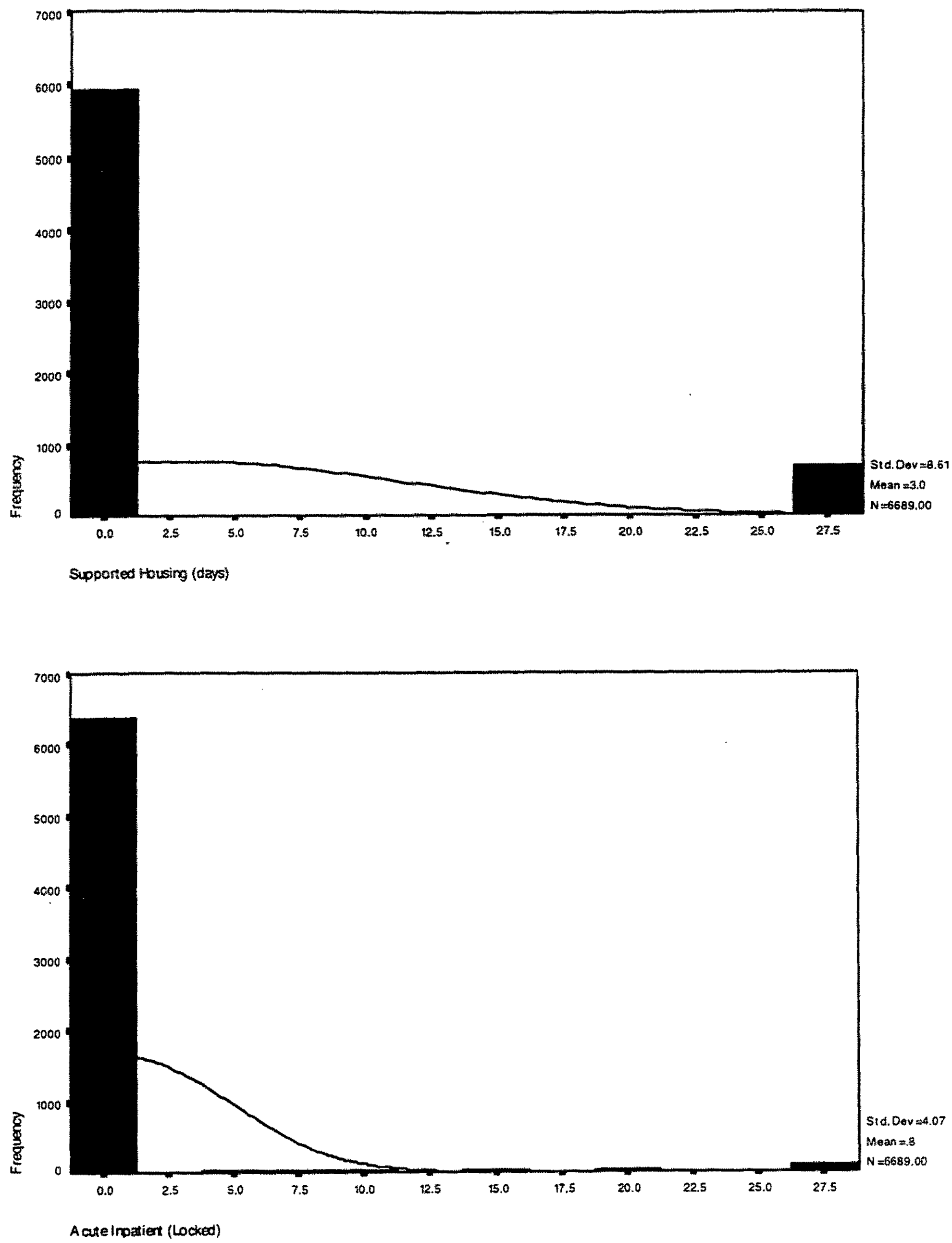

Figure 1: Typical "Backwards J" Distributions

2. Reduce the data on a conceptual basis.

First, we recommend removing services that are neither prescribed nor received since this finding casts doubt on their relevance for the service recipients under study. Second, we recommend collapsing low frequency variables that have conceptual similarity. Some investigators have factor analyzed their service data before they cluster analyzed it. This might be appropriate for attribute data (e.g., symptoms), since these data are usually based on multiple measures of similar variables that we would expect a factor analysis to group together. However, we do not recommend this approach for service data. We do not expect a person to receive similar services. Presumably, this would be wasteful. We do expect a person to receive complimentary services. Given this, service factors would be like clusters and using both approaches together would be redundant. Cluster analysis is preferred over factor analysis alone with service data because factor analysis does not readily allow for individuals to be placed in a single group.

3. Choose a cluster method, depending on the desired shape of the final clusters and psychometric considerations.

A cluster method should be chosen based on the shape and size of the clusters desired and the psychometric properties of the data to be clustered. As we describe below, there are a variety of cluster methods available. We recommend k-means, a centroid method, for use with service data for three reasons: (1) Centroid clusters are composed of members who have high degrees of mutual similarity to each other. In contrast, linkage clusters, another common cluster type, are composed of members who are more like one member in the cluster. The idea that we are looking for persons whose service profiles are mutually similar to each other's (rather than to one other member's in a cluster) is intuitively appealing. (2) Previous research suggests that centroid clusters are not as affected by outliers as other cluster methods (Milligan, 1980). Since service data distributions are often characterized by outlier groups (high utilizers), the centroid method may be most appropriate for service data. (3) The k-means approach can be implemented with large samples on desktop computers. Despite the forgoing, we have found little guidance in the literature as to how other methods might apply to specific problems. This is an area in which evaluators applying cluster methods might seek the assistance of a statistician.

Types of Clusters

This discussion is very technical. We have tried to simplify it as much as possible. However, evaluators who are not statistically inclined may want to skim this section. It is not necessary to fully understand it to apply k-means cluster analysis to service data.

Clusters can be thought of as a shape in p-dimensional space, where p is the number of variables or attributes in a subject's profile. Clusters can be determined to be compact (spherical or ellipsoidal) or they can be extended (serpentine, chained, connected).

All of a *compact* cluster's members have high mutual similarity, the elements are located within a circumscribed distance of one another, and are more like every other member in a cluster than like members of other clusters. The clusters are roughly spherical in shape.

All of an *extended* cluster's members are more like *one* member in the cluster than like any other member not in the cluster (e.g., the member nearest or farthest according to some Euclidean distance measure). The clusters tend to be long, serpentine or amoebae-like in shape.

The Clustering Procedure

Clustering procedures can be *hierarchical* or *non-hierarchical.* Figure 2 illustrates the relationship of various clustering procedures.

Hierarchical. In one hierarchical method, the agglomerative method, one begins with each element as a cluster, then merges the two most similar, then the next two most similar until there is a single cluster. In another hierarchical method, the divisive method, one begins with all elements in a single cluster then searches for an element most dissimilar from the others to create a second cluster. The method continues until each element forms its own cluster. One applies a rule to the "tree"

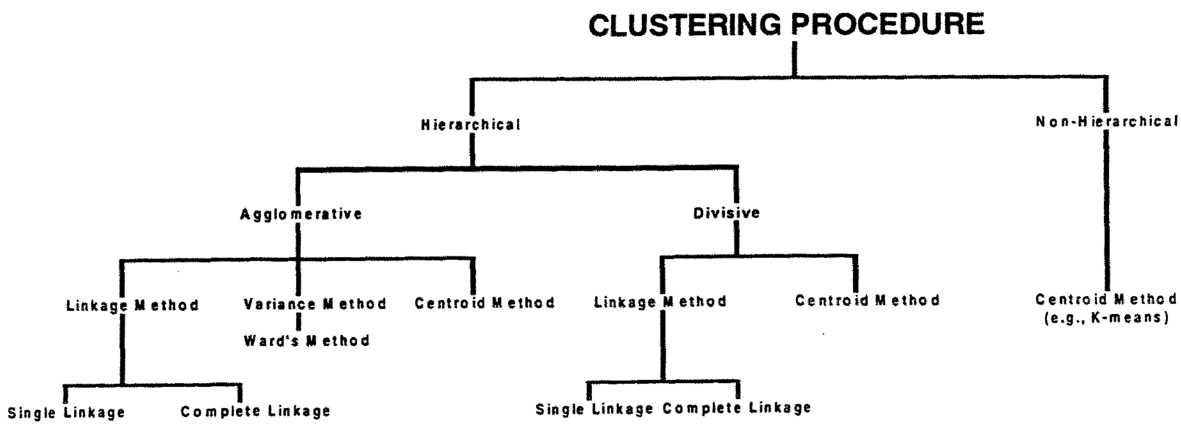

Figure 2: Relationship of various types of cluster formation algorithms

diagram formed by the procedure to determine the optimal number of clusters. Often, this rule involves inspection of the cluster coefficients displayed in the output listing that represents the squared Euclidean distance between clusters (see below). When the increase between two adjacent steps becomes large, relative to previous steps, the number of clusters is optimal.

Non-Hierarchical. In the non-hierarchical approach, there is only one pass through the data, and each element is assigned a cluster membership based on its profile's similarity to the clusters already existing. This approach allows persons to leave one cluster and join another as the clusters are formed. The procedure, for example, k-means, uses a fixed number of clusters and attempts to find the solution that maximizes the difference between clusters using an ANOVA-like test.

Measuring the Similarity Between Two Profiles

In cluster analysis procedures, indices of similarity such as distances or correlations can still be computed even if the scales are not truly interval.

One commonly used "distance" between two profiles is the Euclidean measure for distance. For two persons, one calculates the difference between the persons' scores on a variable, squares that difference, then performs the same procedure on every other variable in their profiles. All the squared differences are summed, and the square root of that sum is the Euclidean distance between the profiles.

In symbolic terms,

$$D = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2}$$

where x_i are the p values for one person, and y_i are the p values for the other person.

D is also affected by three components:

1. The Elevation: the mean level of the scores
2. The Scatter: the range or variation in the scores
3. The Shape: the configuration of high and low scores

If scores across the profiles are mean deviated, the elevations for profiles become zero. If scores are standardized, all profiles' scatter is the same.

The D measure could represent a large difference between two profiles on a single variable or the sum of many small differences on all the variables.

The cluster coefficient is identical to the Euclidean distance between clusters. It is a measure of the extent to which a clustering algorithm is responding to outlier data. This measure can be computed at each step in a hierarchical analysis. When there is a jump in the cluster coefficient, this suggests that there has been a response to outlier data and that subsequent clusters may be less valid than previous ones. Unfortunately, the cluster coefficient is only produced with hierarchical methods and is not applicable to k-means.

There are other distance measures as well, such as the "city-block" metric where the absolute value of the differences between scores is summed, rather than the squared difference.

There are also other similarity measures that may be appropriate as well. The Pearson Product-Moment Correlation Coefficient between persons across variables can be used as well as the various rank correlation coefficients. For non-parametric data, Kendall's Tau and the Contingency Coefficient (based on chi-square, but restricted to a range from 0 to 1) have also been used in cluster analysis.

If two profiles are parallel in some sense, the correlation coefficient between them will be unity, but if the profiles are "far" apart, the similarity in shape may not be a sufficient criterion for them to be considered similar.

A Display of the Clustering Process – the Dendrogram

The steps in a hierarchical solution can be illustrated in a diagram known as a *dendrogram*. It shows the clusters joining at each step. At the bottom of the dendrogram, every case is a cluster. As clusters are joined, lines connecting them appear until, at the top of the dendrogram, there is a single cluster. The researcher inspects the chart and cluster coefficients and makes a decision about the optimal number of clusters based on cluster sizes and coefficients. This procedure requires large amounts of remote access computer memory and can only be performed with small samples of persons (e.g., < 200) if personal computers are being used.

When the procedure proceeds from every case being a cluster to a single cluster the procedure is called *agglomerative*. When the procedure begins with a single cluster, divides that into two and ends with every case a cluster, the method is called *divisive*.

Clustering Methods

There are a variety of methods for clustering data. As described above, the choice of a method may be based on conceptual reasons, the distribution of the data or computational practicality. The following is a brief discussion of the most commonly used methods

Linkage Methods

Single Linkage (Nearest Neighbor). The first two cases have the smallest distance or largest similarity between them. The next case to be added to a cluster has the smallest distance or largest similarity to a case in the cluster. The resulting solution places emphasis on detecting extended or snake-like clusters rather than compact ones.

Complete Linkage (Furthest Neighbor). The distance between clusters is the distance between their farthest points. This method is strongly biased toward producing clusters with roughly equal diameters. The shape can be severely distorted by moderate outliers.

Average Linkage (Between Groups). Known as the Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA), this method uses as the distance between clusters, the average of the distances between all pairs of cases where one member of the pair is from each cluster. This technique tends to join clusters with small variances and is slightly biased toward producing clusters with the same variance.

Variance Methods

Ward's Method. The means of all variable within clusters are calculated. Then for each new case, the squared Euclidean distance from the cluster means are computed. A combination is made with the case that results in the smallest increase in the overall sum of the squared within cluster distances. For this reason, Ward's method is sometimes called a variance method. Ward's Method tends to join clusters with small numbers of observations and is strongly biased toward producing clusters with roughly the same number of observations. It is also very sensitive to outliers.

Centroid Method

The distance between clusters is calculated by using the distance between their means for all variables. The centroid of two clusters joined is a weighted sum of the centroids of each cluster where the weights are related to the size of the clusters. This method is more robust to outliers than most other hierarchical methods; i.e., the value of the centroid is less affected by the outliers. However, the centroid method might not perform as well as some others because the distance at which clusters are combined can change with each step and actually can decrease. This means that clusters that are joined in later steps may be more alike than those joined at earlier steps, which might be undesirable.

K-means is a centroid method. It has the advantages of centroid methods cited above. It has the added virtue of being computable on a desktop computer for a large number of persons, and is readily accessible through programs like SPSS (QuickCluster) and SAS (FasCluster). In our example, the sample was so large as to preclude a hierarchical analysis including all service recipients. That analysis required calculating similarity indices for all possible pairs of persons. At this writing, there is no desktop computer that can hold that much information in it's RAM. One drawback of k-means is that you cannot examine the clustering process step by step. The number of clusters produced is the number specified by the analyst. To explore the results of several numbers of clusters you must perform several analyses and there is no cluster set of coefficients for each run. Nevertheless, it is likely that in most cases that the k-means solution will give similar results to other centroid methods.

4. Look for solutions giving clusters that are large enough for subsequent analyses.

The purpose of a cluster analysis is often to investigate the relationships between service clusters and variables related to service recipient characteristics or service system organization and financing. For such analyses, clusters must be large enough for investigators to examine their association with other variables. Evaluators will want to decide on a cluster solution that will result in enough persons per cluster for subsequent analyses without sacrificing too much specificity.

Evaluating Clustering Results: Cluster Size

By running multiple analyses, the investigator can compare the cluster sizes resulting from schemes using various numbers of clusters. Starting with the largest conceivable number of clusters and working downward may be the most efficient way to approach this task; however, since the final decision as to the "correct" number of clusters is subjective, some investigators may have other preferences.

Table 2 shows the results of several centroid (k-means) cluster analyses performed on the data. The numbers of service recipients in each cluster are given for solutions of 5 through 10 clusters. In this example, after six clusters, the numbers in clusters fall below 100 and after seven clusters, the numbers can be very small, even a single person.

Table 2: K-Means Cluster Analysis: Cluster Census by Analysis

	1	2	3	4	5	6	7	8	9	10
5 Clusters	144	188	347	5484	491					
6 Clusters	341	188	4952	554	486	133				
7 Clusters	539	178	4928	58	129	486	336			
8 Clusters	557	328	693	350	4132	122	471	1		
9 Clusters	548	55	327	474	1	130	687	340	4092	
10 Clusters	542	682	176	37	328	4064	321	13	450	41

Cluster Evolution

Table 3 describes the way in which the cluster procedure "de-constructs" existing clusters when forced to arrive at a solution of more clusters. These paths were constructed using contingency tables where two solutions, e.g., six vs. seven cluster solutions were compared and the migration of service recipients from one cluster to another can be noted. This table was created by saving a cluster membership variable for each analysis.

Table 3: K-means Cluster Evolution

	6 Cluster	7 Cluster	8 Cluster	9 Cluster	10 Cluster	11 Cluster	12 Cluster
1	334	333	331	329	332	331	329
		1	**1**	**1**	**1**	**1**	**8**
2	396	396	397	397	394	394	394
		6	**4**	**6**	**10**	**10**	**10**
3	133	132	133	133	132	132	132
		2	**3**	**8**	**6**	**5**	**5**
4	398	387	381	380	317	318	317
		3	**5**	**5**	**5**	**11**	**9**
5	208	208	207	207	206	205	205
		5	**6**	**2**	**9**	**7**	**7**
6	5220	4855	4762	4825	4756	4668	4658
		4	**2**	**4**	**3**	**8**	**12**
7		378	378	378	376	376	376
			8	**7**	**7**	**2**	**6**
8			100	13	154	148	147
			7	**3**	**2**	**4**	**4**
9				21	21	21	21
				9	**4**	**9**	**3**
10					1	1	1
					8	**3**	**2**
11						95	94
						6	**11**
12							13
							1

5. Perform one-way analysis of variance and multiple post-hoc comparisons across the clusters on all services and use the results to interpret clusters. Service means can be displayed on tables or graphs.

Analysis of Variance

Clusters will be useful to the degree that they are interpretable. In order to interpret service :lusters, one-way analysis of variance (ANOVA) can be computed to determine those services that ignificantly distinguish among clusters. Using cluster membership to assign cases to groups the

ANOVA procedure can be used to compare the mean number of hours of each service among clusters. When the overall F-score for a service is significant, the Student-Newman-Keuls (SNK) procedure can be used to differentiate among clusters on that service. SNK is a multiple comparisons procedure that arranges groups means from smallest to largest and sets the range that is used to test for a significant difference between means on the basis of the number of steps between the two means being tested. For example, the output in Table 4 compares 6 clusters on the mean hours of Locked Inpatient Care provided:

Table 4: Output from Multiple Range Test (SNK) in SPSS for Windows 6.2

Variable AD1 (d) Locked I/P
By Variable PACKPROV Service Package: Provided

Multiple Range Tests: Student-Newman-Keuls test with significance level .050

The difference between two means is significant if
MEAN(J)-MEAN(I) >= 33.3950 * RANGE * SQRT(1/N(I) + 1/N(J))
with the following value(s) for RANGE:

Step	2	3	4	5	6
RANGE	2.81	3.33	3.65	3.87	4.04

(*) Indicates significant differences which are shown in the lower triangle

Mean	PACKPROV	Grp 6	Grp 5	Grp 3	Grp 8	Grp 4	Grp 1
4.4458	Grp 6						
6.0583	Grp 5						
8.1208	Grp 3						
9.8182	Grp 8						
10.7417	Grp 4						
621.6555	Grp 1	*	*	*	*	*	

Table 4 indicates that Group 1 used more hours of locked inpatient service than any of the other groups, and that all differences are statistically significant *within the cluster analysis*. Note that although there were differences between other groups, none were statistically significant.

Table 5 shows the results of one-way ANOVAs and multiple post-hoc comparisons for a seven-cluster analysis. The cells contain mean hours of service provided for 33 services by cluster. Each cluster is represented by a column of data. Figures in each column in bold show services that distinguish the cluster for that column from most other clusters in the SNK comparison. Underlined figures represent services that distinguishes the cluster from fewer other clusters. For example, the service shown in row 1 of the table, Locked Inpatient Care, distinguish cluster 7 from *all* other clusters. However, it distinguishes clusters 1, 2, 5 and 6 only from clusters 3 and 4.

Table 5: 7 Cluster Analysis (K-Means): Mean Hours of Service Provided

groups	1	2	3	4	5	6	7
rvices	n=539	n=178	n=4928	n=58	n=129	n=486	n=336
	hours	hours	hours	hours	hours	hours	hours
) Locked I/P	23.6	53.66	6.63	12	34.98	66.91	**540.86**
) Unlocked I/P	9.22	19.69	2.59	12.83	**329.86**	13.38	9.71
) High Intensity Residential	8.28	7.28	1.77	0	15.63	**420.74**	6.36
) Moderate Intensity Residenti	**430.71**	35.46	3.67	0.41	15.63	5.98	3.21
) Low Intensity residential	7.93	9.03	**25.94**	0	5.21	8.3	2.43
) supportive housing	10.08	15.19	**48.04**	0.41	12.17	4.16	13.51
) Substance abuse residential	0	**73.89**	0.04	0	0	0.35	2
) Crisis respite	2.63	1.23	1.22	0.41	1.49	0.49	**4.64**
Crisis/emergency Outpatient	**0.52**	0.31	0.23	0.47	0.19	0.12	0.4
Evaluation/Diagnosis	0.1	0.26	0.08	0.19	0.14	0.08	**0.58**
Individual Counseling	**1.68**	0.86	1.42	1.38	1.21	1.11	1.15
family counseling	0.03	0.02	**0.1**	0	0.12	0.06	0.08
Group counseling	**1.31**	1.34	0.45	0.21	0.56	0.8	0.91
Day Hospital	**6.37**	1.98	2	0	6.59	**7.37**	4.53
med eval/maintenance	**0.96**	0.62	0.62	0.32	0.54	**1.09**	0.77
Substance Abuse/op detox	0.3	**2.51**	0.18	0.48	0.58	0.15	0.3
Voc assessment	0.66	0.05	0.29	0.07	0.47	**1.26**	0.24
Skills training	4.76	0.12	1.37	0	1.12	**5.78**	0.31
Clubhouse	**18.38**	8.68	4.96	0	0.71	5.91	1.61
Job Dev/Ind. & grp supp emp	**3.14**	1.21	1.39	0	0	1.82	0.35
Education	0.59	1	0.58	0	0.4	0.49	0.07
Case management	**2.38**	2.12	0.88	1.24	0.97	1.91	1.69
Indiv support	0.77	1.12	0.55	0.6	0.33	0.5	0.2
Drop in	**8.43**	2.71	2.26	6.1	2.29	**6.74**	1.51
Family support	0.67	0.48	1.95	0	2.26	1.31	0.39
Recreation/socialization	2.33	0.89	1.04	1.21	0.98	**3.81**	0.63
Medical IP/Dental OP	2.8	0.44	1.92	0.1	0.38	**6.1**	3.96
Self help	0.6	**2.81**	0.46	0.36	0.53	0.23	0.52
Guardian/Rep. Payee	0.3	0.29	0.07	0.05	0.09	**0.37**	0.2
ransportation	**2.03**	0.71	0.58	0.26	0.19	**2.51**	0.58
iomeless	31.21	74.29	4.22	**463.86**	20.84	16.49	4.29
Other legal assistance	0.12	0.06	0.05	0.14	0.13	0.08	0.13
ail	1.25	3.91	1.15	0	0	1.38	0.57

In order to interpret and name a cluster, we recommend focusing on those services that the ANOVA post-hoc comparisons indicate most distinguish the cluster from others. Below we list the names assigned to the clusters in our example, and the services that most distinguish them:

1. **Psycho-Social Rehabilitation Services Cluster**
   - Moderate Intensity Residential
   - Crisis/Emergency Outpatient
   - Individual Counseling
   - Group Counseling
   - Day Hospital
   - Med. Eval./Maintenance
   - Clubhouse
   - Job Development/Indiv. Group Supported Employment
   - Case Management
   - Drop In
   - Transportation
2. **Substance Abuse Services Cluster**
   - Substance Abuse Residential
   - Substance Abuse/OP Detox
   - Self Help
3. **Minimal Services Cluster**
   - Low Intensity Residential
   - Supportive Housing
   - Family Counseling
4. **Homeless Services Cluster**

Homeless Services

5. **Acute Psychiatric Inpatient Services Cluster**

   Unlocked IP

6. **High Intensity Residential Services Cluster**

   High Intensity Residential

   Day Hospital

   Med. Eval./Maintenance

   Vocational Assessment

   Skills Training

   Drop In

   Recreation/Socialization

   Medical IP/Dental OP

   Guardian/Representative Payee

   Transportation

7. **Services for Persons At Risk Cluster**

   Locked IP

   Crisis Respite

   Evaluation/Diagnosis

6. Compare the results with another clustering method.

As noted, different clustering methods can be expected to yield different results. It may be desirable to compare several methods. Cluster methods can be simply compared by classifying persons by cluster memberships resulting from different methods and then cross tabulating one membership variable by another.

The next steps all pertain to estimating the validity of the service clusters obtained:

7. Compare expected and observed clusters.

It is useful, in evaluating clusters, to have some idea of the types of clusters you expect to find. For example, with our service data, on the basis of our knowledge of services for persons with severe and persistent mental illness, we expected to find a cluster composed of acute psychiatric services and one composed of psychosocial rehabilitation services. From previous research, we also expected to find a cluster of persons receiving few services (Roth et al., 1992; Leff, in press). The clusters found confirmed these expectations. Although, we were prepared to see these expectations refuted, we would have reviewed the soundness of methods before accepting such results. Consistency between expected and observed clusters might be considered a form of construct validity.

8. Test the stability of the solution by comparing cluster results for randomly selected subsamples.

"Hold-out" approaches in which clusters are obtained for some subsamples and then validated on others is the most rigorous approach to this form of validation. However, it is also possible to generate clusters for all persons in a sample and then look for these clusters in randomly drawn subsamples. This is like assessing the external validity of the clusters obtained.

9. Compare theoretically expected and observed associations between cluster membership and other variables.

A cluster analysis is often a step in testing theorized relationships between service profiles and other variables not used in the cluster analysis (e.g., sociodemographic, clinical, and program). If these relationships conform to theoretical expectations, this could be considered a form of construct validity. If clusters and theory do not match, the appropriateness of the theory and the cluster method should both be reconsidered.

Using Cluster Membership in Additional Data Analyses

Once a set of clusters has been adopted, it is possible to create a nominal cluster membership variable. Using this variable, evaluators can explore the association between cluster membership and: variables such as sociodemographic and clinical variables; program variables related to the organizing and financing of mental health care; and outcome variables. If the unit costs for services in clusters are known, the costs of clusters can also be estimated.

In investigating the associations between cluster membership and other variables, it is important to consider the time period covered by the services in the cluster. As noted above, clusters covering longer time periods should be related to trait measures and variables like course, whereas clusters covering shorter periods of time should be related to state measures like level of functioning.

The associations between cluster membership and individual variables can be explored through measures of association. Discriminant function analysis can be used to determine the combinations of sociodemographic and other variables that best predict cluster membership.

Reporting Results

Blashfeld (1980) posits five criteria as minimum requirements for any journal article using cluster analysis, and these should be borne in mind from the very start of the analytic process:

1. *An unambiguous description of the cluster analytic method should be provided.*
2. *The choice of the similarity measure (or statistical criterion if an iterative procedure is used) should be clearly specified.*
3. *The computer program used to perform the cluster analytic method should be stated.*
4. *The procedure used to determine the number of clusters should be explained.*
5. *Adequate evidence of the validity of a cluster analytic solution should be provided before the solution is published.*

The first three of these criteria can be easily satisfied through unambiguous reference to the documentation provided with the statistical software employed. In the case of the example just described, we used the QUICK CLUSTER procedure in SPSS for Windows, which employs k-means method of cluster formation using squared Euclidean distances to measure similarity (Norušis/SPSS Inc., 1993).

Because there are no mechanistic rules for choosing a number of clusters for a given data set, the criteria affecting this decision must be carefully described. Attention should be paid to the replicability of the technique used. In the example described above, an arbitrary minimum number of cases (100) for "useful" clusters was selected.

Finally, as Blashfeld notes, the issue of cluster validation should be addressed. However here, too, there are no simple rules or procedures to follow. Cluster analysis will generate an unambiguous solution for data even if the data consist of random numbers. Further, the use of ANOVA and multiple range tests, while useful for distinguishing clusters by service, leaves room for reading meaning into randomness. Thinking of initial cluster analyses as "exploratory" and replications as "confirmatory" seems prudent. Everitt (1974) proposes replication across parallel data sets, across different cluster analytic methods and across a different collection of variables as three general procedures to validate a cluster solution. Procedures for measuring construct, concurrent and predictive validity should also be implemented.

Attending to the details of the first four criteria during the analysis makes it possible for researchers to communicate their results in such a way as to allow for replication of their efforts. With replication of studies and other forms of validation, cluster analyses contribute to an ongoing research process.

References

Anderberg, M.R., *Cluster Analysis for Applications*, Academic Press, New York, 1973.

Blashfeld, R.K., Propositions Regarding the Use of Cluster Analysis in Clinical Research. *Journal of Consulting and Clinical Psvchology* 1980, 48, 4: (456 – 459).

Double, D.B., A Cluster Analysis of Manic States. *Comprehensive Psychiatry* 32, 3 (May/June), 1991: 187 – 194.

Everitt, B., *Cluster Analysis*, John Wiley and Sons, New York, 1974.

Furukawa, T, Sumita, Y. A cluster-analytically derived subtyping of chronic affective disorders. *Acta Psychiatr Scand* 1992, 85:177 – 182.

Hartigan, J.A., *Clustering Algorithms*, John Wiley and Sons, New York, 1975.

Lorr, M., *Cluster Analysis for Social Scientists*, Jossey-Bass, San Francisco, 1983.

Milligan, G. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325 – 342.

Norušis/SPSS Inc., *SPSS: SPSS for Windows Professional Statistics Release 6.0.* SPSS Inc., 1993.

Roth, D., Champney, T. F., Clark, J. A., Cogswell, S. H., Lauber, E. G. & Wemmerus, V. A. Services in Systems: Impact on Client Outcomes. *New Research in Mental Health*, 10. Ohio Department of Mental Health, Columbus, 1992.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., Yeaton, W. Some Neglected Problems in Evaluation Research: Strength and Integrity of Treatments. *Evaluation Studies Review Annual* 4, 1979.