NAVIGANT CONSULTING

CLUSTER ANALYSIS

California Peer Grouping

Skilled Nursing Facilities

April 2005

Navigant Consulting, Inc.
633 West Fifth Street 60
Los Angeles, CA 90071
www.navigantconsulting.com

NAVIGANT CONSULTING

PEER GROUPING CLUSTER ANALYSIS

Table of Contents

NAVIGANT CONSULTING

PEER GROUPING CLUSTER ANALYSIS

I. Data Analysis

A. Data Received

A list of 1,033 Skilled Nursing Facilities (NF-Bs) was received, representing 51 of the 58 counties statewide. Seven counties in California have no Medi-Cal skilled nursing days. These counties were excluded from the peer grouping analysis. Costs used in the analysis were based on as-submitted OSHPD data for reports ending in calendar year 2003. Direct care costs were selected to peer group the NF-B population, since these costs represent the majority of provider costs.

The data file received contained the following relevant fields:

Table 1

Field Name	*Field Description*
Facility ID	Unique OSHPD number for each facility
Facility Name	Legal facility name
City	City of operation
County	County of operation
Days in the Period	Number of days in the cost report period
SNF Total Days	Total Skilled Nursing Days
SNF Direct Care Cost	Reported nursing, social services, and activities personnel costs associated with the NF-B. Costs exclude the audit adjustment factor and include temporary agency personnel expenditures.
Direct Care Per Diem ("Per Diem")	SNF Direct Care Costs / SNF Total Days

Each county represented in the analysis population was also identified as being either "Urban" or "Rural" in order to statistically test the hypothesis that direct care costs are influenced by urban or rural status. Classification denoting urban or rural status by county is identified in the following Table 2:

PEER GROUPING CLUSTER ANALYSIS

Table 2

Urban and Rural Classification by County

Classification	*County Number*	*County Name*	*Classification*	*County Number*	*County Name*
Rural	3	Amador	Urban	1	Alameda
Rural	4	Butte	Urban	7	Contra Costa
Rural	5	Calaveras	Urban	10	Fresno
Rural	6	Colusa	Urban	19	Los Angeles
Rural	8	Del Norte	Urban	21	Marin
Rural	9	El Dorado	Urban	27	Monterey
Rural	11	Glenn	Urban	28	Napa
Rural	12	Humboldt	Urban	30	Orange
Rural	13	Imperial	Urban	33	Riverside
Rural	14	Inyo	Urban	34	Sacramento
Rural	15	Kern	Urban	36	San Bernardino
Rural	16	Kings	Urban	37	San Diego
Rural	17	Lake	Urban	38	San Francisco
Rural	18	Lassen	Urban	41	San Mateo
Rural	20	Madera	Urban	42	Santa Barbara
Rural	23	Mendocino	Urban	43	Santa Clara
Rural	24	Merced	Urban	44	Santa Cruz
Rural	29	Nevada	Urban	48	Solano
Rural	31	Placer			
Rural	32	Plumas			
Rural	39	San Joaquin			
Rural	40	San Luis Obispo			
Rural	45	Shasta			
Rural	47	Siskiyou			
Rural	49	Sonoma			
Rural	50	Stanislaus			
Rural	51	Sutter			
Rural	52	Tehama			
Rural	54	Tulare			
Rural	55	Tuolumne			
Rural	56	Ventura			
Rural	57	Yolo			
Rural	58	Yuba			

PEER GROUPING CLUSTER ANALYSIS

B. Summary Statistics – Key Observations

The following summary statistics were calculated for each county on the per diem amount: median direct care cost per day, average direct care cost per day, standard deviation, and a frequency count, representing the number of NF-Bs in each county. The following Table 3 summarizes this information:

Table 3

Urban	*County*	*Median Cost*	*Average Cost*	*Std Dev.*	*Frequency*
TRUE	Los Angeles	$ 57.95	$ 59.98	14.63	337
TRUE	San Bernardino	58.63	61.54	10.51	45
TRUE	Riverside	61.03	64.16	10.52	43
TRUE	Fresno	63.34	63.07	8.60	33
TRUE	Solano	67.15	72.76	10.68	9
TRUE	San Diego	69.75	68.73	14.35	67
TRUE	Orange	70.20	72.46	13.71	60
TRUE	Santa Cruz	72.71	69.79	11.85	10
TRUE	Sacramento	74.83	76.79	12.04	34
TRUE	San Francisco	77.65	84.94	14.48	11
TRUE	Santa Barbara	80.32	83.86	18.15	11
TRUE	Monterey	81.42	80.48	11.33	11
TRUE	San Mateo	81.67	82.68	9.09	12
TRUE	Alameda	82.49	81.51	14.35	59
TRUE	Santa Clara	83.28	86.82	16.38	48
TRUE	Marin	83.84	94.74	22.14	10
TRUE	Contra Costa	84.63	83.20	12.29	26
TRUE	Napa	86.40	86.63	10.43	8
FALSE	Imperial	$ 49.92	$ 51.74	8.25	3
FALSE	Del Norte	52.44	52.44		1
FALSE	Lake	53.99	59.22	10.90	3
FALSE	Yuba	55.94	55.94		1
FALSE	Tulare	56.99	57.42	2.82	13
FALSE	Lassen	58.23	58.23		1
FALSE	Kings	58.65	60.97	4.34	3

PEER GROUPING CLUSTER ANALYSIS

Urban	*County*	*Median Cost*	*Average Cost*	*Std Dev.*	*Frequency*
FALSE	Kern	58.95	60.38	9.09	15
FALSE	Colusa	60.14	60.14		1
FALSE	Tehama	61.08	67.88	12.29	3
FALSE	Mendocino	63.44	65.81	4.42	5
FALSE	Butte	63.94	65.96	9.95	12
FALSE	Madera	65.21	64.24	2.80	4
FALSE	Yolo	65.78	68.46	14.09	7
FALSE	Inyo	65.94	65.94		1
FALSE	San Luis Obispo	66.16	64.79	5.82	7
FALSE	Merced	66.55	64.76	4.80	8
FALSE	Humboldt	67.49	69.19	7.55	5
FALSE	Ventura	68.91	72.53	12.30	17
FALSE	Stanislaus	69.73	69.88	6.63	16
FALSE	Shasta	69.90	70.95	6.01	6
FALSE	Calaveras	69.93	69.93		1
FALSE	San Joaquin	70.53	69.91	6.78	24
FALSE	Glenn	70.91	70.91		1
FALSE	Plumas	71.12	71.12		1
FALSE	Sutter	72.07	72.15	4.62	4
FALSE	Siskiyou	74.20	74.20	22.29	2
FALSE	El Dorado	75.54	76.59	3.89	3
FALSE	Placer	77.13	81.02	17.20	9
FALSE	Nevada	77.90	80.85	8.13	4
FALSE	Sonoma	77.94	79.82	10.57	16
FALSE	Amador	78.97	78.97		1
FALSE	Tuolumne	79.25	79.25		1

Some of the key observations regarding these summary statistics are as follows:

1. Los Angeles County contains the largest number of facilities providing skilled nursing services to Medi-Cal beneficiaries (337 facilities or 32% of total NF-Bs in the analysis population). Despite its size, Los Angeles has the lowest average direct care per diem cost of all counties designated as urban. The next largest urban county, San Diego, is only one-fifth the size of Los Angeles County (67 facilities or 6.5% of all NF-Bs in the analysis population).

PEER GROUPING CLUSTER ANALYSIS

Using the criteria specified in the prior table, Los Angeles County has been identified as an outlier in terms of "size" of the county, measured by the number of NF-Bs providing services to Medi-Cal beneficiaries. In simple terms, an outlier is defined as a value that is located far from other values in a set of data. There are different criteria often used for determining which observations are outliers, including: (1) observations that are more three standard deviations away from the mean, and/or (2) the relative distance from the observation to the next closest observation.

2. On average, urban counties have more NF-Bs than their rural counterparts. Even when excluding Los Angeles, the urban counties have approximately 29 NF-Bs per county, compared to an average of six in the rural counties. In many cases, the rural counties have only one facility serving Medi-Cal beneficiaries. Note that the urban/rural classification was assigned to each county without regard to the number of facilities in a county.

3. In order to test the observation in point #2, we compared the distribution of the urban and rural facilities in relation to the median size of counties statewide. The median size of a county was determined to be eight facilities. By selecting the median as the measurement of central tendency, 25 counties had more than eight facilities (considered to be "large" counties) and 26 counties had less than or equal to eight facilities (considered to be "small" counties). If no relationship exists between the urban/rural classification assigned to each county and the county size, we would expect to see the proportion of urban and rural counties that are "small" and "large" to be similar. However, 94% of the urban counties (17 of 18 counties) were found to be a "large" county. Only one urban county (Napa) was identified as a "small" county, represented by eight facilities. The following Table 4 indicates the strong relationship between urban/rural classification and the size of the county:

PEER GROUPING CLUSTER ANALYSIS

Table 4

Frequency Table

	Small	*Large*	*Total*	
Rural	25	8	33	*>> Rural: 76% are "Small" (25 / 33)*
Urban	1	17	18	*>> Urban: 94% are "Large" (17 / 18)*
Total	26	25	51	

4. On average, urban counties represent higher direct care per diem costs than their rural counterparts. This indicates that NF-Bs in urban counties tend to be more expensive facilities, with regard to direct care staffing costs. The average per diem cost per facility in the urban counties (excluding Los Angeles) is $74.532, compared to $68.648 in the rural counties. If Los Angeles County is included in the calculation, the overall average per diem cost for the urban counties drops to $68.654. Including Los Angeles County with the other urban counties automatically skews all calculations due to the tremendously large number of facilities located in Los Angeles County. This alone suggests that Los Angeles County should be its own NF-B peer group.

5. The distribution of direct care per diem costs within individual counties is skewed. If the population is not skewed, the median and the mean (average) would be approximately equal. In a skewed distribution, the median may be a better measure of central tendency than the mean. For example, a median home price is most often used as the indicator of the affordability of homes in a region. Since the direct care per diem costs are skewed, the median cost per diem per county was selected as the appropriate input to use in the clustering algorithm.

PEER GROUPING CLUSTER ANALYSIS

II. Cluster Analysis

A. What is Cluster Analysis?

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its objective is to sort people, things, events, etc. (in our case, counties into peer groups, or clusters) so that the degree of association is strong between members of the same cluster and weak between members of different clusters. The degree of association is measured in terms of "distance". The clustering algorithm seeks to minimize the "distance" between members of the same cluster while, at the same time, maximizing the "distance" to the members of different clusters. A cluster analysis uses variables identified by the user, including the number of desired clusters to be produced and the specific inputs that the software will utilize to group observations in a way that minimizes the "distance" within clusters and maximizes the "distance" between clusters. Cluster analysis can be particularly useful in grouping data that may otherwise lack structure or certain explanatory power in and of itself. SAS, a widely used and accepted statistical software package, was used to perform the cluster analysis.

B. Cluster Analysis – NF-B Results

Based on certain key observations noted above in Section 1.B., the following variables were initially selected as input variables for the cluster analysis: median/average direct care per diem, urban/rural status, and frequency (number of NF-Bs in a county). This initial analysis confirmed that there is a distinct difference between facilities operating in urban versus rural counties. This difference was demonstrated not only in terms of the average direct care per diem cost, but also in terms of the average number of facilities, the standard deviation (dispersion), and the "size" of the county, as measured by facility count. As noted, size of the county was a key characteristic that segregated Los Angeles County from all other counties statewide. In the initial cluster analysis, the clustering algorithm assigned Los Angeles County into its own cluster, logically explained by the fact that it is

five times larger than the next largest county and it has the lowest median and average per diem cost of all urban counties.

After examining both the median and mean as a potential clustering variable, median was chosen, as direct care per diem costs are skewed within the counties. As noted, the median was determined to be the better measure of central tendency. The initial clustering and statistical analyses indicated that urban and rural counties were distinctly different from one another. Accordingly, the cluster analysis was refined so as to allow the clustering algorithm to consider other variables besides urban/rural designation in its calculations. Since the urban and rural counties were found to be distinct, urban and rural counties were analyzed separately. Also, Los Angeles County was found to be an outlier and to skew urban costs significantly, supporting a separate cluster to account for the relatively low costs identified in this large urban county. These refinements to the cluster analysis resulted in the use of the median cost per county as the only input variable in the clustering algorithm. The final cluster analysis was therefore based on the median direct care cost per diem cost per county, and was run separately for urban versus rural counties (with Los Angeles County classified as a separate cluster).

The clustering algorithm requires the user to input the desired number of clusters to be output in the result set. Initially, three urban clusters (including Los Angeles County) and three rural clusters were created, totaling six potential peer groups. After reviewing the three rural clusters, however, it was determined that the size of the peer groups and the standard deviation of the peer groups could be improved by creating an additional fourth rural cluster. After refinement, the clustering algorithm created results based on four rural NF-B peer groups and three urban NF-B peer groups.

The characteristics of the clusters (peer groups) were analyzed to determine the appropriateness of the clustering results; in other words, were the clustering results optimal? First, each peer group within the urban or rural designation was found to be relatively similar in size (in terms of number of NF-Bs). To further test the resulting clusters, the level of dispersion of costs (measured as standard

deviation) was analyzed. It was found that increasing the number of clusters does not reduce the dispersion (standard deviation) of each of the clusters. In fact, in certain instances, increasing the number of clusters actually increases standard deviation. For example, increasing the number of clusters from the three urban/four rural clustering previously discussed to a five urban/five rural clustering results in increases to the standard deviation in some of the clusters, as illustrated in the following Table 5:

Table 5
Dispersion of Costs for Seven versus Ten Clusters

Urban	*Std Dev*	*Count*	*Urban*	*Std Dev*	*Count*
TRUE	14.63	337	TRUE	14.63	337
TRUE	12.76	267	TRUE	10.33	130
TRUE	14.72	230	TRUE	13.78	171
			TRUE	13.83	104
			TRUE	15.73	92
FALSE	7.60	44	FALSE	8.81	7
			FALSE	7.24	37
FALSE	8.04	49	FALSE	8.04	49
FALSE	8.05	70	FALSE	8.05	70
FALSE	11.81	36	FALSE	11.81	36

In analyzing the resulting clusters, it was determined that four rural clusters, and three urban clusters (including Los Angeles as a separate cluster) resulted in peer groups that are: (1) relatively even in terms of peer group "size" (measured by the number of facilities within urban versus rural groups), (2) relatively tighter standard deviation ranges between urban and between rural clusters, (3) reduced standard deviations within clusters, and (4) distinctly different median direct care per diem costs between clusters. These important combinations of goals and results are illustrated in the preceding Table 5 and following Table 6:

NAVIGANT CONSULTING

PEER GROUPING CLUSTER ANALYSIS

Table 6

Urban	*Cluster*	*Avg.*	*Median*	*Std Dev*	*Count*
TRUE	A (LA)	$ 59.98	$ 57.95	14.63	337
TRUE	B	67.10	65.50	12.76	267
TRUE	C	83.15	81.62	14.72	230
FALSE	A	59.05	58.26	7.60	44
FALSE	B	66.13	65.62	8.04	49
FALSE	C	70.79	70.21	8.05	70
FALSE	D	79.62	77.81	11.81	36

The final cluster designation for each individual county is included in the following Table 7:

Table 7

Urban	*Cluster*	*County*	*Median*	*Average*	*Std. Dev.*	*NF-B Count*
TRUE	A	Los Angeles	$ 57.95	$ 59.98	14.63	337
TRUE	B	San Bernardino	58.63	61.54	10.51	45
TRUE	B	Riverside	61.03	64.16	10.52	43
TRUE	B	Fresno	63.34	63.07	8.60	33
TRUE	B	Solano	67.15	72.76	10.68	9
TRUE	B	San Diego	69.75	68.73	14.35	67
TRUE	B	Orange	70.20	72.46	13.71	60
TRUE	B	Santa Cruz	72.71	69.79	11.85	10
TRUE	C	Sacramento	74.83	76.79	12.04	34
TRUE	C	San Francisco	77.65	84.94	14.48	11
TRUE	C	Santa Barbara	80.32	83.86	18.15	11
TRUE	C	Monterey	81.42	80.48	11.33	11
TRUE	C	San Mateo	81.67	82.68	9.09	12
TRUE	C	Alameda	82.49	81.51	14.35	59
TRUE	C	Santa Clara	83.28	86.82	16.38	48
TRUE	C	Marin	83.84	94.74	22.14	10
TRUE	C	Contra Costa	84.63	83.20	12.29	26
TRUE	C	Napa	86.40	86.63	10.43	8

PEER GROUPING CLUSTER ANALYSIS

Table 7 *continued*

Urban	*Cluster*	*County*	*Median*	*Average*	*Std. Dev.*	*NF-B Count*
FALSE	A	Imperial	$ 49.92	$ 51.74	8.25	3
FALSE	A	Del Norte	52.44	52.44		1
FALSE	A	Lake	53.99	59.22	10.90	3
FALSE	A	Yuba	55.94	55.94		1
FALSE	A	Tulare	56.99	57.42	2.82	13
FALSE	A	Lassen	58.23	58.23		1
FALSE	A	Kings	58.65	60.97	4.34	3
FALSE	A	Kern	58.95	60.38	9.09	15
FALSE	A	Colusa	60.14	60.14		1
FALSE	A	Tehama	61.08	67.88	12.29	3
FALSE	B	Mendocino	63.44	65.81	4.42	5
FALSE	B	Butte	63.94	65.96	9.95	12
FALSE	B	Madera	65.21	64.24	2.80	4
FALSE	B	Yolo	65.78	68.46	14.09	7
FALSE	B	Inyo	65.94	65.94		1
FALSE	B	San Luis Obispo	66.16	64.79	5.82	7
FALSE	B	Merced	66.55	64.76	4.80	8
FALSE	B	Humboldt	67.49	69.19	7.55	5
FALSE	C	Ventura	68.91	72.53	12.30	17
FALSE	C	Stanislaus	69.73	69.88	6.63	16
FALSE	C	Shasta	69.90	70.95	6.01	6
FALSE	C	Calaveras	69.93	69.93		1
FALSE	C	San Joaquin	70.53	69.91	6.78	24
FALSE	C	Glenn	70.91	70.91		1
FALSE	C	Plumas	71.12	71.12		1
FALSE	C	Sutter	72.07	72.15	4.62	4
FALSE	D	Siskiyou	74.20	74.20	22.29	2
FALSE	D	El Dorado	75.54	76.59	3.89	3
FALSE	D	Placer	77.13	81.02	17.20	9
FALSE	D	Nevada	77.90	80.85	8.13	4
FALSE	D	Sonoma	77.94	79.82	10.57	16
FALSE	D	Amador	78.97	78.97		1
FALSE	D	Tuolumne	79.25	79.25		1

PEER GROUPING CLUSTER ANALYSIS

C. Additional Information – Sub-acute Care Units

The AB1629 legislation relates to both NF-B facilities and sub-acute care units of freestanding NF-B facilities. As illustrated in the preceding Table 7, the NF-B median direct care cost per diem ranges from a low of $49.92 in Imperial County to a high of $86.40 in Napa County. The median direct care cost per diem for sub-acute units ranges from approximately $174.00 in Fresno County to $267.00 in Santa Clara County. Given the small number of sub-acute care units and their significantly higher direct care per diem costs, it is logical to maintain these facilities as their own single peer group in the new AB1629 reimbursement methodology.

Chart 1