Columbia Population Research Center
Advancing Research in Population, Health, and Society

CPRC Working Paper No. 10-02

Typologizing Temporality: Time-Aggregated and Time-Patterned Approaches to Conceptualizing Homelessness

William McAllister
Institute for Social and Economic Research and Policy
Columbia University

Li Kuang
Department of Sociomedical Sciences
Columbia University

Mary Clare Lennon
Department of Sociology
The Graduate Center, City University of New York

April, 2010

1255 Amsterdam Avenue ▪ New York, New York ▪ 10027
cupop.columbia.edu

TYPOLOGIZING TEMPORALITY: TIME-AGGREGATED AND TIME-PATTERNED APPROACHES TO CONCEPTUALIZING HOMELESSNESS

William McAllister/*Columbia University*

Li Kuang/*Columbia University*

Mary Clare Lennon/*The Graduate Center, City University of New York*

MARCH 2010

William McAllister is Senior Research Fellow at the Institute for Social & Economic Research & Policy, Columbia University. Direct all correspondence to McAllister. Email address: wm134@columbia.edu. Mail address: 420 West 118th St, MC 3355, New York, NY 10027.

Li Kuang is a Ph.D. candidate in Sociomedical Sciences at the Mailman School of Public Health, Columbia University.

Mary Clare Lennon is Professor of Sociology at The Graduate Center, City University of New York.

Typologizing Temporality: Time-aggregated and Time-patterned Approaches To Conceptualizing Homelessness

Abstract

This paper shows how we can use a relatively new method to construct time-based homeless typologies that expand our ability to theorize and to make policy. Our key argument is that commonly created "time-aggregated" typologies lose potentially useful information by summing, averaging or otherwise summarizing events that occur over time. We suggest instead a "time-patterned" approach that captures events as they unfurl over time by measuring their timing, duration and sequencing. Comparing time-aggregated and time-patterned analyses of Kuhn and Culhane's prominent three-category typology, we find the time-patterned approach performs marginally better. We argue, however, that both analyses reveal problematic heterogeneity in the three groups and that the initial theorizing is not robust. These deficiencies suggest the utility of further analysis. Using the time-patterned results, we identify a four-pattern/ten-group typology that technically and substantively contrasts strongly with the prevailing three-category typology. We then imagine how structural factors and individual traits can combine to generate these observed homeless patterns, and conclude that either approach and either typology may be appropriate, depending on theorizing and the uses to which the findings are to be put.

Keywords: homelessness; typology; optimal matching; time-patterned; time-aggregated

Typologizing Temporality: Time-aggregated and Time-patterned Approaches To Conceptualizing Homelessness

Theorizing and policymaking are improved when we have better concepts. Forming typologies helps us conceptualize. This paper shows how we can use a relatively new method to construct time-based homeless typologies that expand our ability to theorize and to make policy. Our key argument is that commonly created "time-aggregated" typologies lose potentially useful information by summing, averaging or otherwise summarizing events that occur over time. We suggest instead a "time-patterned" approach that captures events as they unfurl over time by measuring their timing, duration and sequencing.

We first describe the prevailing typology in homeless research and policymaking, explain some problems with how temporal information is employed in that typology and describe a time-patterned approach that allows us to utilize more of that information. Using theorizing that describes the prevailing typology, we then carry out a time-patterned analysis and compare it to a time-aggregated analysis. We find the time-patterned approach performs marginally better. More important, however, the extent of within-group variability and the weakness of theorized relationships suggest this commonly used typology can be improved. Further analyzing time-patterned results, we identify a four-pattern/ten-group typology that technically and substantively contrasts strongly with the prevailing three-group typology. We then go on to suggest how structural factors and individual traits can combine to generate observed homelessness histories, and conclude by arguing that either approach and either finding may be appropriate, depending on theorizing and the uses to which an analysis is being put.

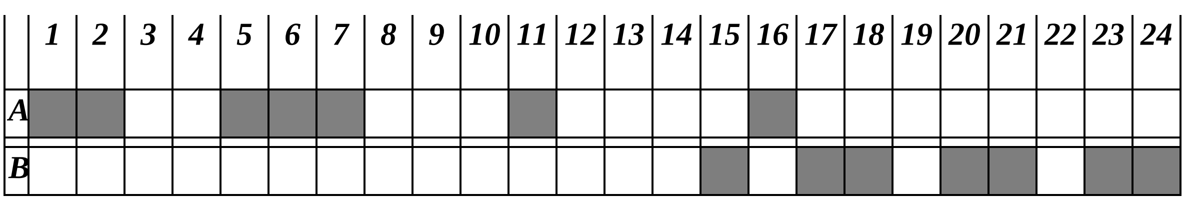
Homeless typologizing

Typologies combine different values of relevant phenomena to identify concepts, e.g., the chemical table or Weber's typology of organizations (Stinchcombe 1968; Bailey 1994; Doty & Glick 1994). The major theorizing on and evidence for a typology based on the temporal experience of homelessness of adults unaccompanied by mates or children are found in Kuhn and Culhane (1998).[1] Based on extant homelessness research, they characterize three kinds of homeless shelter experience.[2] *Transitionally* homeless people "are forced to spend a short time in a homeless shelter before making a transition into a more stable housing arrangement, and in most cases they do not return to homelessness." *Episodically* homeless people "frequently shuttle in and out of homelessness. . . . [They] often find their way back to the shelters." And *chronically* homeless people "are likely to be entrenched in the shelter system, . . . for whom shelters are more like long-term housing than an emergency arrangement." (211). The transitional category is expected to be very large compared to the other two, which are expected to be about the same size. Last, Kuhn and Culhane also theorize relationships between kinds of homelessness and individuals' features, e.g., age, race, mental health, physical health and substance abuse problems.

To empirically analyze this theorizing, Kuhn and Culhane conceptualize two temporal features to typologize, the frequency and duration of people's homelessness. They utilize shelter administrative data to measure these dimensions for first-time sheltered people by counting the number of shelter episodes (frequency) and the total number of days sheltered (duration) over three years. Those transitionally homeless are expected to enter shelters once or twice for a very short time; episodically homeless to enter shelters many times but for short periods; and chronically homeless to enter shelters once or very few times but spend most of their time homeless. Kuhn and Culhane find the expected three group typology with their expected sizes, and they show traits hypothesized to be associated with each group generally were associated and in the right direction. (See Kertesz et al 2005 for further support regarding hypothesized traits.)

Time-aggregated and time-patterned approaches

In Kuhn and Culhane's approach, time is aggregated by treating each occurrence of an event as a cross-sectional variable to be summed over the time period, i.e., each person's total number of homeless days or episodes. Research in this approach may also calculate rates, averages, proportions or otherwise aggregate events over time, across or within individuals. Such time-aggregated measures, however, cannot represent peoples' patterns of homelessness or housing situations as they occur over time in their lives. That is, it may be more faithful to how people experience their lives as well as more analytically useful to typologize based on patterns of when, in what sequence and for how long episodes of homelessness occur as people's lives unfold. The following diagram illustrates what we mean. It shows two hypothetical persons housing histories over 24 time periods. A white cell means the person was housed the entire time period, and a gray cell means he or she was homeless the entire period:



Taking a time-aggregated approach, the homeless histories of these two people appear identical, e.g., their number of days homeless and of homeless episodes are the same. Taking a time-patterned approach that analyzes the timing, duration and sequencing of homeless and non-homeless episodes suggests they have very different histories. This approach indicates the first person has perhaps fitfully freed herself from homelessness while the second has increasingly tumbled into it. We might then theorize differently about these peoples' lives, as we would be explaining different histories or perhaps suggesting these different histories have different effects.

A time-patterned approach allows us to construct typologies that identify more refined concepts regarding the temporal character of homelessness. Relative to time-aggregation, time-patterning expands the number of dimensions observed (timing and sequencing of homeless and non-homeless episodes) and, by incorporating this temporality, effectively translates frequency and duration into more subtle measures by specifying when each episode occurs and how long each lasts. The more analytically useful dimensions employed and the more those dimensions are appropriately measured, the richer typology-based concepts can be. In both these ways, incorporating how events occur over time seems appropriate to refining the temporal-based conceptualizing of Kuhn and Culhane. It allows us, for example, both to see the ambiguity in a “relatively short” episode and length of time between episodes (transitional category) and to construct a typology that does not require such characterization but rather just incorporates whatever the time interval is. Similar ambiguities for the other categories can be better seen and empirically addressed using time-patterning, as we shall show.[3]

In the rest of the paper, we support these arguments by analyzing homeless shelter data using both approaches. We first compare results from the two analyses for the theorized three-group typology, and find the time-patterned analysis produces marginally better results than the time-aggregated analysis. However, within-group heterogeneity remains great in both analyses, and the time-patterned results are not technically “optimal”. Further time-patterned analysis produces a four-pattern/ten-group solution that both technically improves on the three-group solutions and substantively identifies a very different typology from that produced by Kuhn and Culhane.

Methods

Data. To make comparisons between the two approaches more evident, we use a recent version of the dataset used by Kuhn and Culhane: the Single Client Information Management System (SCIMS) of the City of New York. For all single adults ($\geq$ 18 year old) unaccompanied by children or mates entering New York City supported shelters, SCIMS records dates of each shelter entrance and exit and collects self-reports of demographic characteristics, substance abuse and mental and physical health. Crucial for the time-patterned approach, SCIMS tracks very well entrance and exit dates of every shelter stay for each client.

For our analysis, we select individuals in the manner of Kuhn and Culhane. We choose people first entering a shelter between January 1, 2000 and December 31, 2003. (Kuhn and Culhane selected people first entering a shelter between January 1, 1988 and September 30, 1992.)[4] Per Kuhn and Culhane, we allow each person to remain in or to leave and re-enter a shelter for three years from the date of his or her initial entrance, yielding three years observation of shelter use. This results in 40,169 cases.[5] To make our analysis more analytically tractable, we randomly sampled one in eight of these cases, for a sample size of 5,000. This is both large enough to discern groups with sufficient numbers of cases yet small enough to be computationally tractable for the software used in the analysis. The sample strongly represents the population from which it was drawn. Comparing the two on demographic characteristics, shelter use measures and theorized covariates, we found the largest difference in proportions is 0.007. No differences are statistically significant.

Measurement. The time-patterned and time-aggregated analyses use the same temporal information but capture it differently. Following Kuhn and Culhane, our time-aggregated analysis measures shelter use frequency and duration by counting each person's shelter stays and his or her total amount of time sheltered over the three-year observation period. The first is identified by the number of shelter stays separated by at least 30 continuous days out of shelter and then summing the number of these episodes; the second is calculated by summing the number of days sheltered across all stays. Both calculations replicate Kuhn and Culhane.

The time-patterned analysis breaks up the three-year time period into segments of 30 days — a "month". Each month, a person can be sheltered for 0 to 30 days. This number is the input for the time-patterned analysis. We use 30 day time periods because anything less (e.g., weeks) creates too many time periods, making the analysis less tractable, and anything greater (e.g., quarters) is analytically unnecessary and increases measurement error. Also, this is consistent with New York City policy that people have to be absent from shelters for more than 30 continuous days for the City to conclude a shelter stay has ended. Thirty day periods are also consistent with Kuhn and Culhane's analysis which used New York City's 30-day rule to define a stay. In addition, we use the exact number of days each time period because this approximates Kuhn and Culhane's using the total number of days sheltered to measure duration.

In sum, for the time-patterned analysis, we capture frequency and duration as did Kuhn and Culhane but measure them by how they are distributed — and not summed — over time. This occurs because we capture the timing and sequencing of events, as Kuhn and Culhane did not.[6] Combined, our measures allow us to identify temporal patterns of homelessness in each person's life. Additionally, we measure theorized co-variates exactly as did Kuhn and Culhane and exactly the same way in both analyses.[7]

Analytic techniques. To carry out the time-patterned analysis, we use optimal matching analysis (OM. See Appendix A for a fuller description of what we did.[8]) In brief, OM first calculates a value expressing how different each individual's history is from every other person's in the dataset. This is arrived at by calculating the smallest ("optimal") sum of the number of weighted changes needed to transform the homeless sequence of one person into the exact same sequence of another. Doing this for all pairs of individuals in the dataset produces an n by n dissimilarity matrix, where cell values reflect how different each person's sequence is from that of every other person. For people with relatively similar histories, this cell value will be small compared to the value of people with relatively different histories.[9] To derive the dissimilarity matrix, we used the software package TDA (Rohwer and Pötter 1999).

This matrix is then cluster or otherwise analyzed to group together people with similar matrix values, i.e., with similar timing, duration and sequencing of shelter and non-shelter histories. We use cluster analysis, specifically Ward's method. This identifies clusters by finding solutions at each stage that exhibit the smallest change in the within-group sum-of-squared deviations from the mean of each cluster summed across all clusters, i.e., the total within-group sum-of squares (Everitt et al 2001).[10] For example, having found a ten-group solution, Ward's then finds the nine-group solution that has, among possible nine-group solutions, the smallest change in the total within-group sum-of-squares from the ten-group solution. We use Ward's because its sum-of-squares criteria is convenient for discussing within-group homogeneity, and we and others have found it usefully clusters individuals using the optimal matching dissimilarity matrix, e.g., Stovel and Bolan (2003); Stark and Vedres (2006). We use the Ward's algorithm in ClustanGraphics 8.0 (Wishart 2004. For more technical descriptions of OM, see Sankoff and

Kruskal 1983; Abbott and Hrycak 1990; for housing-relevant examples of its use, see Stovel and Bolan 2003; Clark et al 2003; for sociological critiques of OM, see Levine 2001 and Wu 2001).

To carry out the time-aggregated analysis, we reproduce Kuhn and Culhane. We cluster analyze each person's number of days sheltered and number of shelter episodes to identify groups of those most similar to one another along both dimensions. To make the time-patterned and time-aggregated analyses more comparable, we use Ward's method, rather than the nearest centroid sorting technique used by Kuhn and Culhane.[11] It is important to note that because we follow Kuhn and Culhane, we are constructing typologies of homeless shelter use and not of the broader experience of homelessness. In the next section we examine the three-category typology using both approaches.

Three-Group Analyses

Time-patterned analysis. Table 1 shows the three groups produced by our time-patterned analysis. To illustrate the kinds of cases in each group, the table reports ten cases selected at random from each group. Numbers in cells are the number of days people were sheltered for each 30 day period; blank cells indicate zero days in shelter for a particular 30 day period.

The table suggests Group A resembles the theorized transitional category — 74.5% of all cases, and people are sheltered relatively briefly and have relatively little shelter experience after leaving, and Group C resembles the theorized chronic category — 6.1% of all cases, and people have relatively very long, continuous shelter stays and few non-shelter breaks. But Group C also contains people with strong in-and-out of shelter histories. This leaves group B, then, to support the theorized episodic category. This group is appropriately small relative to Group A (19.6% of all cases), and does contain people with histories of moving in-and-out of shelter. But durations for these people are longer than might be expected, and the group is dominated by people sheltered for a relatively long, maximal (29-30 days of a time period), continuous (12-18 months) time — and then not at all. This evidence is weak support, at best, for the theorized episodic category.[12]

Another way to see how these results relate to theorized categorizations is to identify people in each group with shelter histories akin to what is expected for that group, and then calculate for each group how different these histories are from all other histories in that group. For any group, the theoretically expected histories should more closely match all histories than the histories not theoretically expected for that group. For example, histories in Group A that typify the transitional type should be less different from all histories in Group A than histories that typify Groups B and C. Table 2, Panel A shows the logic of this reasoning: typical histories are all exactly the same, are found only in their theorized category and are the only cases in that category (0.0 dissimilarity on the diagonal), and the typical histories from the other groups differ from these histories (> 0.0 dissimilarity in the off-diagonal).

To carry out this analysis, we identify in each group the cases with histories consistent with the theorized expectations for that group, randomly sample each group's population of such illustrative cases, and construct dissimilarity matrices using the dissimilarities initially produced by optimal matching. (The bold, italicized cases in Table 1 are examples of the population of illustrative cases in each group.[13]) We then calculate how different on average each sample of illustrative cases is from all other cases in each group.[14]

Table 2, Panel B shows the average dissimilarity in Group A is smallest for the transitional illustrative cases (7.3 vs 51.7 and 120.8); in Group B for the episodic illustrative cases (45.4 vs 50.2 and 79.0); and in Group C for the chronic illustrative cases (25.6 vs 70.8 and 105.2). These results support our previous characterizations of each group, though, as before, support for Group B is much weaker than for the other groups. Group B has the lowest variation among dissimilarities across illustrative cases, indicating its illustrative cases are almost as similar to all other cases in Group B as are the chronic and transitional illustrative cases. In addition, all transitional and chronic illustrative cases are found in Groups A and C, respectively, but episodic illustrative cases are found in all three groups.

These analyses suggest some, but not strong, support for the three-group hypothesis. But how well does this time-patterned approach perform relative to a time-aggregated approach? And how well does it generate relatively homogeneous groups, a crucial criteria for typology formation?

Time-aggregated comparison. To compare these results to a time-aggregated analysis, we re-produce Kuhn and Culhane's three-group study using our dataset. Table 3 reports these results for their frequency and duration measures. The results are strongly consistent with expectations for the transitional, episodic and chronic categories and with the group characteristics Kuhn and Culhane initially found.[15] We compare these results to the time-patterned analysis for within-group homogeneity and for co-variate associations. Typological concepts are more analytically useful the more homogeneous their empirical referent and the more they are demonstrated to be associated with measures of other concepts.

To assess within-group homogeneity, we calculate total within-group sum-of-squared deviations from the group average (cluster centroid). This measure is based on the principle used in optimization cluster analysis to assess cluster homogeneity (Everitt et al 2001). Also, the Ward's method clustering algorithm we use groups people based on minimizing changes in sum-of-squares. Table 4 shows the raw within-group sum-of-squares for each group for each approach and, to standardize for differences in group size, the average within-group sum-of-squares.

For the time-patterned results, Panel A reports average sum-of-squares are 10.4, 31.6 and 28.8 for the transitional, episodic and chronic groups, respectively, and 31.3 for the entire dataset.[16] As the ratios of these measures show (third column in table), the heterogeneity in the episodic and chronic groups is about as great as in the dataset as a whole (i.e., treating the dataset as one cluster). The transitional group, however, is markedly more homogeneous. These results are consistent with the degree of variability we found in the initial time-patterned analysis. Additionally, the time-patterned solution explains about 50% of total dataset variability, while the time-aggregated solution explains about 62%, both leaving a good deal to be explained. Thus, this analysis suggests neither approach is generating homogeneous chronic and especially episodic groups.[17]

To assess the relationship of these groups to measures of other concepts, we examine their associations with Kuhn and Culhane's theorized covariates for which we have measures: age and medical, mental health and substance abuse problems. Following Kuhn and Culhane's analysis, we trichotomized age at 30 and 50 years; for the other covariates, SCIMS measured only whether or not people had a particular kind of problem. Table 5 shows direction of theorized relationship, odds ratios and associated 99% confidence intervals, and whether result is statistically significant.[18] For example, taking the first row in Panel A, the table reports the transitionally homeless are expected to have more people under 30 than the chronically homeless, and the odds ratio is estimated to be a statistically significant 2.52, supporting the theorizing.

In general, for the time-patterned analysis, the transitional/chronic and transitional/episodic comparisons show strong support for theorized claims. All relationships with all covariates are in the right direction and are statistically significant[19], (save substance abuse problems, which is significant with a 95% confidence interval.) However, except for age, odds ratios for chronic/episodic relationships are not statistically significant and are in the wrong direction for mental health and substance abuse problems. In general, these odds ratios are stronger and more of the associations are statistically significant at .01 than for the time-aggregated results, shown in Panel B. Only the statistical significance for mental health problems in the transitional/chronic comparison and the estimate of the odds ratio for substance abuse problems in the transitional/episodic comparison are less robust for the time-patterned groups. In both analyses, the chronic/episodic relationships are, for the most part, not statistically significant. We note, however, that the confidence intervals overlap for all statistically significant odds ratios for the same relationship in the two analyses, making it impossible to say the odds ratios are statistically distinguishable. All we can say is that the odds ratios in the time-patterned analysis are, in general, more likely to be stronger, in the right direction and statistically significant than the comparable odds ratios in the time-aggregated analysis.

Conclusions for three-group analyses. This analysis suggests several things. First is that the time-patterned and time-aggregated approaches identify similar three-category typologies and these typologies support the theorized categories. Second, the within-group homogeneity produced by each analysis has the same character (comparing across groups within each analysis). Third, the time-patterned results are more strongly and significantly related to theorized covariates. Fourth, and most important, the three-group solutions identified by both approaches create groups that retain a good deal of heterogeneity, evidenced by the proportion of remaining dataset variation and by the absence of support for theorized relationships. The transitional group in both analyses performs well — relatively homogeneous and strong, statistically significant relationships with covariates. But the chronic and especially the episodic groups perform badly on both criteria. This suggests further analysis to gain a stronger understanding of the temporal-based structure of shelter use.

Further analysis is also supported by the weakness of the theorizing cited by Kuhn and Culhane.[20] It effectively states that the transitional and chronic categories have the same frequency (low) but vary on duration. Anything else is thus episodic. Or, more colloquially, homelessness consists of a short-lived emergency or a long-term condition or something else. This theorizing perhaps oversimplifies short-term and long-term homelessness, and leaves episodic a residual category. The analysis presented here supports this characterization of the episodic category, and points to the oversimplification of long-term homelessness.

This theorizing and the empirical research based on it have been an important starting point for understanding the temporal structure of homelessness. In the next section, we build on this start by identifying a time-patterned solution that addresses the issues just identified. This allows us to suggest an alternative to the three-category typology that has greater homogeneity and a different characterization of time-based homelessness.

Time-patterned analysis

To identify potential solutions, we employ the analysis already described and use two guides. Figure 1 reports fusion values (left axis; these are changes between solutions in total within-group sum-of-squares) and t-values (right axis; this evaluates the statistical significance of the fusion values). Commonly, solutions are favored when the change in statistically significant fusion values is slight, i.e., when solutions with more groups have little impact on reducing the total within-group sum-of-squares (Wishart 2005). By this criterion, the figure suggests one possibly useful set is the nine- to twelve-group solutions.

Our second guide is to rely on theorizing temporally-based kinds of homelessness where it is relatively strong. One such kind is the temporary category, a version of which is articulated in Kuhn and Culhane's transitional category and by other researchers, e.g., Rossi et al 1986; Culhane and Metraux 1999; Burt et al 2001. That is, due to one-off kinds of events (a fire that razes a house, a breakup in an intimate relationship, a unique financial moment, a transition from institutionalized care to a home of one's own, and so forth), there is thought to exist a relatively large population of people who enter shelters once, for a very short time and never return. Thus, from among the nine- to twelve-group solutions, we choose one in which people were sheltered only once and for no more than 30 days, and is most parsimonious (i.e., has the fewest groups). The ten-group solution meets these criteria.[21]

Results. Table 6 reports exemplars for the ten-group solution. These are cases in each group who have the smallest within-group average dissimilarity to all other cases in the group (Wishart 2004) and which visually appear representative of the dominant temporal structure of cases in the group. The ten groups can be understood as expressions of four basic patterns resulting from the timing, duration and sequence of shelter stays: temporary, structured-continuous, structured-intermittent and unstructured-intermittent. These patterns describe the categories of the typology. (Weisburd et al 2004 use a similar strategy to describe trajectories found by using group-based modeling techniques.)

The first pattern is the theorized *temporary* category: one very brief shelter stay in the initial 30-day period and no shelter re-entrance. In some sense, this finding is not surprising, since we anchored our selection of a solution by choosing from available solutions those best representing this temporary type. Also, based on theory and empirical evidence from New York City shelters, we expected to find such a pattern. Nevertheless, we had no assurance, using the time-patterned approach and optimal matching, of finding such a pattern in any solution. Having done so encourages our confidence in this particular solution, though there are also other solutions with such a temporary category.

The *structured-continuous* pattern is one continuous shelter stay for the maximum time each 30-day period with little re-entrance. The major variation in this pattern is the duration of the stay. We note that people in Group G are still sheltered at the end of the observation period, but this group shares the fundamental maximal and continuous time-in-shelter traits of this category. A longer observation period would likely have shown additional groups with the same pattern leaving at some point after 36 months. Not shown in the exemplar is that the groups contain people who do return, but generally for less than 30 days and for one 30-day period.

The *structured-intermittent* pattern is sequences of shelter and out-of-shelter stays, maximal shelter stays each 30-day period, occurring at different points in the observation period. The two groups in this category show variations in the timing and duration of this pattern. Both have a short shelter sequence after people enter, but Group H has another sequence almost 18 months later, lasting for a year, while Group I's has a second sequence six months later, lasting for six months. These histories suggest Group H may be more likely to return to shelter after the observation period. We note that both groups contain variability not displayed in the exemplars; in particular, re-entrance for brief periods of time with less than maximum time.

The *unstructured-intermittent* pattern is individually-varying sequences of shelter and out-of-shelter stays that, relative to the other patterns, are more intermittent, more unstructured and highly variable in the amount of time sheltered each 30-day period. Because histories are so variable across individuals, identifying an exemplar would not be as informative as for the other groups. To indicate this pattern graphically, the table in Appendix B shows five randomly drawn cases.

Evaluating results. We undertook this analysis to see if we could improve the homogeneity and reduce the substantive ambiguity of the time-patterned, three-group solution. Further demonstrating the utility of the time-patterned approach itself, we expect this analysis to produce different substantive results from the time-aggregated approach. We thus compare the four-pattern/ten-group solution to our earlier time-patterned analysis and then to comparable solutions generated by the time-aggregated approach.

To measure within-group homogeneity, we calculate weighted means for each solution for the within-group sum-of-squares for the groups comprising each solution. This average sum-of-squares is 9.1 for the ten-group solution compared to 15.6 for the three-group solution, a 42% reduction in within-group sum-of-square. To measure total dataset homogeneity, we calculate average total sum-of-squares and estimate how much of this variation each solution explains. The ten-group solution explains 71.0%, compared to 50.1% for the three-group solution, also an improvement of 42%.

The patterns and groups of the ten-group solution are also substantively different from the three-group solution. Perhaps one similar group is found in both analyses — the transitional or temporary. But a segment of the chronic homeless are now seen to be one group among many others comprising what we call structured-continuous, and the episodic and chronic are now seen to comprise two very different kinds of patterns, structured and unstructured intermittent. Thus, the ten-group solution has produced typological categories very different from those produced by the three-group solution.

These categories are also different from those produced by time-aggregated “optimal” and ten-group solutions. In the time-aggregated analysis, we identified a six-group solution that is optimal using the same criteria employed to identify our optimal time-patterned, ten-group solution, i.e., the point at which additional solutions had little impact on fusion values. To compare across similar numbers of groups, we also identify a time-aggregated, ten-group solution. Both solutions are only more refined versions of the three-group typology, i.e., sub-groups within each category that are slightly different variations on the themes of transitional, episodic and chronic.[22] (Results not shown but available on request.) Thus, the time-patterned approach has generated a typology that could not have been identified using a time-aggregated approach.

Discussion

Sosin’s (2003) discussion of theorizing the causes of homelessness points out that a complete explanation for homelessness has to account for why homeless people move between different housing circumstances and has to include the impact on those transitions of factors beyond individual characteristics. Using a time-patterned approach to form typologies creates opportunities to generate and test such explanations. These temporal patterns identify changes in people’s lives that time-aggregated analyses cannot, i.e., transitions from one status to another at particular moments in time and given a particular history. Explaining these requires incorporating factors different from (but in addition to) individual traits, since traits are relatively immutable and so cannot by themselves generate changes in sequences. Structural conditions, on the other hand, provide such explanations, in two ways. One is that conditions change, e.g., new policies governing shelters go into effect; how mental health services are offered is altered; and so forth. Another is that the operation of conditions themselves generates temporal patterning. Homeless people with mental health problems are subject to the procedures, practices and organization of the mental health system and, perhaps as well, of the criminal justice and health care systems. When such people initially use shelters, how long they stay, when they return and so forth are likely determined by their responding to features of these systems, given their individuality. To explain the structured-continuous pattern, for example, it is not enough to say its groups are linearly related to age (as they are), but we should then go on to specify mechanisms (structural conditions) that translate age into the observed groups.[23]

In this section, we illustrate this argument by imagining theoretically suggestive answers to how the four patterns occur, i.e., why people remain in, leave and re-enter shelters.[24] Our answers are not the only ones possible; we intend them only to suggest the utility for theorizing of the time-patterned approach.

Temporary. People in this category leave within 30 days, never to return. The theoretically relevant questions are why do they remain sheltered so briefly and why do they never return (at least during our observation period). Not returning is key, since, in our typology, returning would place a person in a different category.

One set of possible answers has to do with how the rules and regulations, the physical condition and social environment of shelters repel people who, for instance, have low tolerance for such rules or are fearful of living among strangers.[25] Such structural conditions select against certain kinds of people remaining. Further, since relevant individual traits do not change and if the organization, physical conditional or social environment of the shelter does not change over the observational period, the person is unlikely to re-enter. Of course, people’s not re-entering also depends on their ability to negotiate structural conditions outside the shelter.

These suggestions are consistent with the little data available: relative to other categories, the temporary homeless are substantially more white with relatively low levels of mental or physical ill health and substance abuse problems. These are qualities that make it more possible to negotiate living conditions outside shelters and, so, eschew the problematic conditions of shelter life.[26]

Structured-continuous. People in this category stay continuously sheltered each month until they leave, with groups determined by successively later time points of leaving — a strongly monotonic, concave curve. This suggests shelter conditions are such that individuals can or want to stay fully sheltered while using them, but are more or less done with shelters by the time they leave. To explain this, we might imagine that, as people have experience with institutionalized living (jail/prison, hospitals as well as shelter), they are more able to deal with shelter conditions, and as people have more problems, they have greater use for services shelters may provide. We might also imagine why people leave so discretely, e.g., in-shelter time-eligibility for permanent-housing. Also, shelter conditions may make people's problems more difficult to deal with, even as problems make shelter living more necessary. People can deal with this tension for variable lengths of time.

Using available data, we can only make a small, very tentative examination of this argument. Our suppositions suggest a monotonic positive relationship between the successively greater durations of the six groups comprising this category and age (as a measure of experience) and mental, physical and substance abuse problems. This is what we find.

Structured-intermittent. People in this category enter and leave shelter in clumps of durations marked by maximal monthly stays. One explanation may be that they have more acute involvement with non-shelter governmental and non-profit agencies that also house people than do people in other categories. That is, compared to these latter, they may live more completely the "institutional circuit" pattern described by Hopper et al (1997)[27], moving among shelters, jails/prisons, half-way houses, hospitals, mental health facilities and so forth because they are poor, have weak or exhausted social support and perhaps mental illness, and so are more likely to come into contact with the criminal justice system, are less likely to take medicine, are more likely to continue to abuse drugs and alcohol and so forth. If this is the case, the groups in this category may be structured by the conditions of non-shelter institutionalized living, e.g., how long they can stay in hospitals or half-way houses; the physical and social conditions and daily living rules of these places; the arrest and trial process and sentence for those meeting up with the criminal justice system; and so forth.

Unstructured-intermittent. People in this category do not have discernibly shared patterns. It is characterized by relatively less maximal monthly stays, lasting one or a few continuous months, but without sufficient commonality to the timing, duration and sequencing to produce one or a few exemplary patterns. They may re-enter, remain and leave shelter as they do because it is but one place in a highly fluid mix of less institutionalized places to sleep, such as the homes of friends or relatives or on the street. In contrast to the structured-intermittent category, this category may be more determined by the structural conditions of the lives of friends and relatives rather than the structural conditions of institutional settings. (Both descriptions, however, may fit the kind of semi-permanent state of homelessness described by Sosin et al 1990.) People may spend time in the kinds of institutions described above, but the personal settings may be more dominant. These conditions may be more likely to select out for people who are relatively young, since young people are more willing and able to stay for short periods of time with friends and relatives or on the street, will tend to be unwilling to remain institutionalized in mental health or substance abuse facilities for long periods and are less likely to be convicted of more serious criminal offenses or receive longer sentences (due to a shorter history of arrests and convictions). And we do find people in this pattern are younger than those in any other pattern.

We re-iterate that these are some explanations that can be imagined for each category. We cannot present rigorous evidence supporting them. Our point has been to roughly suggest show how time-patterned analysis constructs the lives of homeless people in a way that fosters considering how structural conditions interact with individual traits. A more definitive typology will be one that identifies distinct temporal conceptualizations of homelessness that have distinct sets of causes and effects.

Conclusion

We end by highlighting some conclusions about how to construct typologies, the appropriate number of typological categories, and structural theorizing in homelessness.

Constructing time-based typologies. The major point of our analysis has been to suggest the utility of forming temporally-based homeless typologies by incorporating more temporal information. The dominant typology in homeless research and policymaking uses much less temporal information than is available. Our evidence concerning group homogeneity and the substantive nature of the categories suggests we can construct more homogeneous groups that reveal a different picture of the temporal nature of shelter-use in people's lives. We are not arguing, however, that a time-patterned approach is the only "right" way to typologize homelessness. The approach taken should be determined by theory and the uses to which the typology is put. If theory necessitates that time be aggregated, e.g., the theory only argues that the amount of time homeless matters, then, in a test of that theory, going to the trouble of unpacking that aggregation does not do us much good. But if a theory argues that, say, the sequencing of housing events explains the condition of their mental health or that structural conditions explain people's histories of homelessness, then unpacking how events occur over time will be useful. Further, as we have shown here, temporally unpacking events can help develop concepts not theoretically anticipated.

Number of typological groups.[28] There can be more than one "correct" number of typological categories. With either approach, it is possible to identify typologies having different numbers of technically valid groups. (Kuhn and Culhane, for example, do not discount that more, or fewer, than three groups could be found.) The issue, basically, is how to trade off parsimony and accuracy; specifically, in our case, how to trade off a smaller number of more temporally heterogeneous groups for a larger number of more temporally homogeneous groups. As we argued regarding choice of analytic approaches, smaller or larger numbers of technically valid groups can be "correct" even with the same dataset, depending on theorizing and the uses to which the analysis is being put. Theorizing three categories and finding support for it may be sufficient for explaining some phenomenon, for example, or for improving policymaking. But this does not mean findings cannot be technically improved and conceptually advanced, as we have shown. Our argument here is true for different solutions that are technically valid, but can also be true, as Gelman and Rubin (1996) show, even when a possible solution is not technically valid.

Structural Theorizing. Last, we have emphasized the utility of a time-patterned approach for structural theorizing. By identifying previously unknown shared housing durations and transitions for groups of people, we can better theorize the role of specific structural factors in generating these sequences: how, for example, the organization of and access to services, shelter and other homeless policies, the operation of the criminal justice and health care systems — and changes to all these — generate transitions and sustain the new status. Structural theorizing is, of course, possible with a time-aggregated approach, if perhaps more difficult. The opportunity to explain change created by a time-patterned approach enhances its utility. Shared histories of movements between housing/homeless states as well as the amount of time in each state can be analyzed by considering structural conditions that change over those histories and by considering how individual traits interact with these conditions to produce observed histories. This better allows us to generate the more complete explanation for homelessness and of its effects that homelessness researchers seek.

Appendix: Our Optimal Matching Analysis

Optimal matching has two steps: calculating an n by n dissimilarity matrix expressing how different each sequence is from every other sequence, and analyzing the dissimilarity matrix to group similar sequences. Critical to the first step is setting weights to value the different transformations used to turn one sequence into another. This transformation happens by *substituting, inserting or deleting* values into a sequence. Substitutions replace the value in one sequence with the same value from the sequence to which it is being compared; insertions and deletions (indels) insert or delete values from a sequence relative to the values in the sequence to which it is being compared. Different substitutions and indels can be differently weighted. All possible transformations may not be equally important. In our case, each person's history consists of counts of the number of days homeless each consecutive 30 day period, i.e., from 0 to 30. (Over the observation period, there are 36 such time periods.) Thus, we need to evaluate the "cost" of substituting, inserting or deleting these values for one another, e.g., the cost of substituting five days sheltered in one sequence for ten days sheltered in another.

Substitution costs: 1-30 shelter days. We utilize the metric in the continuous measure of number of days sheltered each time period to identify substitution values by simple subtraction, i.e., the weight of substituting 10 days sheltered for 7 days shelter is 3. Evidence from our data suggests, however, that the probability of leaving shelter in any one month declines quickly after the first few days and then stays more or less constant, with perhaps a slight uptick near the end of the month. That is, it more difficult for individuals to stay sheltered a second or third day than, say, a 19^{th} or 20^{th} day. We want substitution costs to reflect this behavior, so we take the natural log of the difference between days to generate these costs for states 1-30.

Substitution costs: out-to-in shelter. Evidence from our data suggest that (re-)entering a shelter is much more difficult than remaining sheltered. To remain sheltered each night is to return to a place whose rules, social character and physical qualities are known, if not familiar. Such a person has more or less successfully lived under these conditions. For people newly entering and even, to some extent, for those re-entering shelter, the features are more likely to be a concern. Thus, we want the cost of substituting any number of days sheltered for out-of-shelter (i.e., zero days sheltered) to be "meaningfully" larger than the cost of substituting the log of any number of days sheltered for one day sheltered. For example, substituting one day for out-of-shelter should be meaningfully greater than substituting two days sheltered for one day sheltered. To accomplish this, we arbitrarily assign out-of-shelter the value of .025, since subtracting the natural log of this number from any number of days in-shelter provides a sufficiently large cost. Doing so set the cost of substituting one day for out-of-shelter to just over five times greater than substituting two days sheltered for one day sheltered.

Insertion and deletion (indel) costs. These costs are typically set in relation to substitution costs, depending on how much indels are valued for forming the dissimilarity matrix. Employing indels emphasizes the importance of strings of similar values in a sequence (duration) and de-emphasizes the importance of when those values occur (timing). Technically, the greater the value of indels relative to substitution costs, the less likely is the algorithm to use them. (For discussions of the logic and meaning of indels, see Abbott and Tsay 2000; Lesnard 2006; and Lesnard n.d.).

Indels can be particularly useful when sequence lengths are unequal; deleting and inserting values is equivalent to inserting and deleting time periods, thus equalizing sequence lengths. In our case, because all lengths are equal, we can think of indels as more refined substitution costs.

That is, when sequence lengths are equal, every insertion requires a subsequent deletion (and vice versa) for each pair of sequence lengths to remain equal, effectively a substitution.

Our sequences are dominated by out-of-shelter time periods, suggesting the importance of setting indels relative to the costs of substituting being out-of-shelter for being sheltered any number of days. We found indels = 1.8 generated the most useful results relative to theorizing. (This is .50 of the out-of-shelter to one day sheltered substitution cost and .26 of the maximum substitution cost.) With this value, indels can play a role when equating a time period by exchanging out-of-shelter for any number of days sheltered (or vice versa); it plays no role when equating a time period by exchanging any number of days sheltered. In this latter instance, only the substitution matrix is used.

References

Abbott, Andrew. 2001. "Conceptions of Time and Events in Social Science Methods." 161-82 in *Time Matters,* edited by Andrew Abbott. Chicago, IL: University of Chicago Press.

Abbott, Andrew and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Analysis: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96 (1): 144-85.

Abbott, Andrew and Angela Tsay. 2000. "Sequence Methods and Optimal Matching Analysis in Sociology." *Sociological Methods and Research* 29 (1): 3-33.

Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, CA: Sage Publications, Inc.

Bassuk, Ellen L. 2007. "Comment on Dennis P. Culhane et al.'s "Testing a Typology of Family Homelessness Based on Patterns of Public Shelter Utilization in Four U.S. Jurisdictions: Implication for Policy and Program Planning." *Housing Policy Debate* 18 (1): 29-41.

Bauer, Daniel J. and Patrick J. Curran. 2003. “Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes.” *Psychological Methods* 8 (3): 338-363.

Burt, Martha, Laudan Y. Aron, Edgar Lee, and Jesse Valente. 2001. *Helping America's Homeless: Emergency Shelter or Affordable Housing?* Washington, DC: The Urban Institute Press.

Caton, Carol L. M., Boanerges Dominguez, Bella Schanzer, Deborah S. Hasin, Patrick E. Shrout, Alan Felix, Hunter McQuistion, Lewis A. Opler, and Eustace Hsu. 2005. "Risk Factors for Long-Term Homelessness: Findings from a Longitudinal Study of First-Time Homeless Single Adults." *American Journal of Public Health* 95 (10): 1753-59.

Clark, William A. V., Marinus C. Deurloo, and Frans M. Dieleman. 2003. "Housing Careers in the United States: 1968-93: Modeling the Sequencing of Housing States." *Urban Studies* 40 (1): 143-60.

Culhane, Dennis P. and Stephen Metraux. 1999. "Assessing Relative Risk for Homeless Shelter Usage in New York City and Philadelphia." *Population Research and Policy Review* 18 (3): 219-36.

Culhane, Dennis P., Stephen Metraux, Jung Min Park, Maryanne Schretzman, and Jesse Valente. 2007. "Testing a Typology of Family Homelessness Based on Patterns of Public Shelter Utilization in Four U.S. Jurisdictions: Implications for Policy and Program Planning." *Housing Policy Debate* 18 (1): 1-28.

Cunningham, Mary. 2009. "Preventing and Ending Homelessness - Next Steps." *Metropolitan Housing and Community Center Newsletter*. Washington, DC: The Urban Institute.

Danesco, Evangeline R. and E. Wayne Holden. 1998. "Are There Different Types of Homeless Families? A Typology of Homeless Families Based on Cluster Analysis." *Family Relations* 47 (2): 159-65.

Doty, D. Harold and William H. Glick. 1994. "Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling." *Academy of Management Review* 19 (2): 230-51.

Eggleston, Elaine P., John H. Laub, and Robert J. Sampson. 2004. “Methodological Sensitivities to Latent Class Analysis of Long-Term Criminal Trajectories.” *Journal of Quantitative Criminology* 20 (1): 1-26.

Everitt, Brian S., Sabine Landau, and Morven Leese. 2001. *Cluster Analysis, Fourth Edition*. London: Arnold.

Gelman, Andrew and Donald Rubin. 1995. "Avoiding Model Selection in Bayesian Social Research." *Sociological Methodology* 25: 165-73.

Goering, Paula , George Tolomiczenko, Sheldon Sheldon, Katherine Boydell, and Donald Wasylenki. 2002. "Characteristics of Persons Who Are Homeless for the First Time." *Psychiatric Services* 53 (11): 1472-74.

Hopper, Kim, John Jost, Terri Hay, Susan Welber, and Gary Haugland. 1997. "Homelessness, Severe Mental Illness, and the Institutional Circuit." *Psychiatric Services* 48 (5): 659-64.

Humphreys, Keith and Robert Rosenheck. 1995. "Sequential Validation of Cluster Analytic Subtypes of Homeless Veterans." *American Journal of Community Psychology* 23 (1): 75-98.

Jahiel, Rene and Thomas F. Babor. 2007. "Toward a Typology of Homeless Families: Conceptual and Methodological Issues." in *Characteristics and Dynamics of Homeless Families with Children: Final Report*, edited by Debra J. Rog, C. Scott Holupka, and Lisa C. Patton. http://aspe.hhs.gov/hsp/homelessness/improving-data08/apb.htm

Kertesz, Stefan G., Mary Jo Larson, Nicholas J. Horton, Michael Winter, Richard Saitz, and Jeffrey H. Samet. 2005. "Homeless Chronicity and Health Related Quality of Life Trajectories among Adults with Addictions." *Medical Care* 43 (6): 574-85.

Kuhn, Randall and Dennis P. Culhane. 1998. "Applying Cluster Analysis To Test a Typology of Homelessness by Pattern of Shelter Utilization: Results from the Analysis of Administrative Data." *American Journal of Community Psychology* 26 (2): 207-32.

Lennon, Mary Clare, William McAllister, Li Kuang, and Daniel B. Herman. 2005. "Capturing Intervention Effects over Time: Re-analysis of a Critical Time Intervention For Homeless Mentally Ill Men." *American Journal of Public Health* 95 (10): 1760-66.

Levine, Joel H. 2000. “But What Have You Done For Us Lately?: Commentary on Abbott and Tsay.” *Sociological Methods and Research* 29 (1): 34-40.

Lesnard, Laurent. 2006. "Optimal Matching and Social Sciences." Centre de Recherche en Economie et Statistique, Institut National de la Statistique et Des Etudes Economiques, Working Paper 2006-01.

—. n.d. "Describing Social Rhythms with Optimal Matching." http://laurent.lesnard.free.fr/IMG/pdf/Describing_social_rhythms_with_optimal_matching-2.pdf.

Mayer, Karl Ulrich. 2009. "New Trends in Life Course Research." *Annual Review of Sociology* 35: 493-514.

McAllister, William, Li Kuang, and Mary Clare Lennon. 2009. "Re-thinking Research on Typologizing Homelessness." New York: Institute for Social & Economic Research & Policy, Columbia University.

Mowbray, Carol T., Deborah Bybee, and Evan Cohen. 1993. "Describing the Homeless Mentally Ill: Cluster Analysis Results." *American Journal of Community Psychology* 21 (1): 67-93.

Muthén, Bengt O. 2003. “Statistical and Substantive Checking in Growth Mixture Modeling: Comment on Bauer and Curran (2003).” *Psychological Methods* 8 (3): 369-377.

__. 2004. "Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data." 345-68 in *Handbook of Quantitative Methodology for the Social Science,* edited by D. Kaplan. Newbury Park, CA: Sage Publications.

Nagin, Daniel S. 2004. “Response to ‘Methodological Sensitivities to Latent Class Analysis of Long-Term Criminal Trajectories’.” *Journal of Quantitative Criminology* 20 (1): 27-35.

__. 2005. *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press.

National Alliance to End Homelessness. 2007. "Chronic Homelessness." *Factchecker*.

Rohwer, Götz and Ulrich Pötter. 1999. *TDA User's Manual*. Ruhg-Universität Bochum, Germany.

Rossi, Peter H., Gene A. Fisher, and Georgiana Willis. 1986. *The Condition of the Homeless of Chicago*. Amherst: Social and Demographic Research Institute, University of Massachusetts.

Sankoff, David and Joseph B. Kruskal. 1983. *Time Warps, String Edits, and Macro-Molecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley.

Sosin, Michael R. 2003. "Explaining Adult Homelessness in the US by Stratification or Situation." *Journal of Community & Applied Social Psychology* 13 (2): 91-104.

Sosin, Michael R., Irving Piliavin, and Herb Westerfelt. 1990. "Toward a Longitudinal Analysis of Homelessness." *Journal of Social Issues* 46 (4): 157-74.

Stark, David and Balazs Vedres. 2006. "Social Times of Network Spaces: Network Sequences and Foreign Investment in Hungary." *American Journal of Sociology* 111 (5): 1367-411.

Stinchcombe, Arthur L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace and World.

Stinchcombe, Arthur L. and James C. Wendt. 1975. "Theoretical Domains and Measurement in Social Indicator Analysis." 37-73 in *Social Indicator Models*, edited by Kenneth Land and Seymour Spilerman. New York: Russell Sage Foundation.

Stovel, Katherine and Marc Bolan. 2003. "Residential Trajectories: Using Optimal Alignment to Reveal the Structure of Residential Mobility." *Sociological Methods & Research* 32 (4): 559-98.

United States Interagency Council on Homelessness. 2003. *The 10-Year Planning Process to End Chronic Homelessness in Your Community*. Washington DC: US Government Printing Office.

Weisburd, David, Shawn Bushway, Cynthia Lum, and Sue-Ming Yang. 2004. "Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle." *Criminology* 42 (2): 283-322.

Wishart, David. 1969. "An Algorithm for Hierarchical Classifications." *Biometrics* 25 (1): 165-70.

—. 2004. *ClustanGraphics Primer*. Edinburgh, Scotland: Clustan Limited.

—. 2005. "Number of Clusters." 1442-46 in *Encyclopedia of Statistics in Behavioral Science, Vol 3*, edited by Brian C. Everitt and David C. Howell. Chichester, England: John Wiley & Sons.

Wu, Lawrence L. 2000. “Some Comments on “Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospects.” *Sociological Methods and Research* 29 (1): 41-64.

Notes

This study was supported by NIMH Grant R34 MH073651-01 and by the Center for Homelessness Prevention Studies at Columbia University. We appreciate their support. We are also grateful to the Department of Homeless Services of the City of New York (DHS) for providing the data to conduct this study. The analyses and interpretations expressed herein represent the opinions and conclusions of the authors and do not necessarily reflect the opinions of DHS or its staff. We thank George Gavrilis, Tali Schaefer, Glyn Vincent, DHS staff and participants in the Grand Rounds of the Center for Homeless Prevention Studies at Columbia University for their thoughtful comments.

[1] See Culhane et al (2007) for similar theorizing and techniques applied to homeless families. We focus on Kuhn and Culhane because other researchers have used their approach or cited their research (e.g., Goering et al 2002; Caton et al 2005; Kertesz et al (2005), among many others), and their typology is common-place in policymaking discussions (e.g., National Alliance to End Homelessness 2007; Cunningham 2009; United States Interagency Council on Homelessness 2003, among many others).

Non-temporal dimensions have also been used to characterize homelessness (e.g., Mowbray et al 1993; Humphreys and Rosenheck 1995). Kuhn and Culhane argue that because homelessness is commonly understood as the experience of actually living on the streets, in abandoned building or in shelters, non-temporal features (substance abuse, psychological functioning and so forth) should be considered either a cause or a result of that experience and not a constituent dimension of a typology. Rather, we would argue, this depends on what one's theory looks like. Non-temporal features may be typologically useful because those features are expected to be causally important. Danesco and Holden (1998) rightly construct a typology based on features of family ecology because ecology is theorized to explain outcomes for children in homeless families. For a critique of Kuhn and Culhane that argues non-temporal features should be included in homeless typologies, see Bassuk 2007. For a review of homeless typologies, see Jahiel and Babor 2007.

[2] Strictly speaking, the homelessness researchers cited by Kuhn and Culhane were theorizing all homelessness over people's lives, not shelter experience over a few years. Relatedly and importantly, in this paper and following Kuhn and Culhane, we construct typologies based only on homeless shelter use, and not people's broader experience of homelessness.

[3] A time-patterned approach is similar to the trajectories approach found in delinquency (Nagin 2005), education (Muthén 2004), intervention (Lennon et al 2005), life-course (Mayer 2009) and other research areas. One conceptual difference may lie in the latter's expectation of "development" over time, e.g., stopping delinquency, achieving intellectual development or succeeding programmatically; we have no such expectation regarding shelter use.

[4] Since this paper is not interested in how homelessness has or has not changed over time, the difference in time periods is irrelevant. We use a later time period because we are substantively interested in analyzing more current homelessness.

[5] Among these are 1,864 right-censored cases who remained sheltered past the end of the three year observation period. This censoring is not relevant for our analysis because, following Kuhn and Culhane, we are only interested in groups that can be formed from data over the three year period.

[6] Because we break-up the observation period into 30-day months for the time-patterned analysis, this analysis records fewer stays than the time-aggregated analysis, which measures stays as do Kuhn and Culhane. That is, a stay of, say, five days in two consecutive months appears as one stay in a time-patterned analysis, but could appear as two stays in a time-aggregated analysis if the two stays were separated by more than 30 days. In this way, the time-aggregated analysis perhaps better captures frequency. The difference is very slight, however. For the analyses reported here, based on Kuhn and Culhane's measurement strategy, the average number of stays is 2.49 in the time-aggregated analysis and 2.42 in the time-patterned analysis.

[7] These measures are likely much less reliable than the temporal measures, as they are mostly self-reports at entrance and are taken by administrative personnel untrained in data collection.

[8] Other methods/techniques are used to identify time-patterned typologies. Group-based modeling (Nagin 2005) and latent class growth analysis (Muthén 2004) are very similar methods that have been developed and used to analyze developmental theorizing about criminal behavior (Nagin) and educational outcomes (Muthén), among other areas. The primary difference between these approaches and optimal matching is that they are semi-parametric (Nagin) or parametric (Muthén) while OM is non-parametric. (See Abbott 2001 for a discussion of parametric and

non-parametric approaches. Reporting other differences goes beyond the scope of this footnote.) We mainly use OM because the large number of cases and time points in our dataset creates the opportunity to identify acutely varied patterns, and, from our experience using all these methods, OM better identifies such variety. (For critiques and discussion of group-based modeling, see Eggleston et al (2004) and response by Nagin (2004), among other such critiques and discussions; of latent class growth analysis, see Bauer and Curran (2003) and response by Muthén (2003), among other responses therein.)

[9] Typologies are usually generated by combining values of variables. OM generates typologies by combining cases. We can cross-walk between these two approaches by considering OM from the variable perspective. In this approach, we group together individuals who uniquely share sequences of values of the relevant phenomenon (e.g., homeless status). Thus, the unique combination typologies seek is, in OM, not found in combinations of particular values, but in strings of values that are commonly shared simultaneously across cases at one point in time and across each case over the relevant time period. (Of course, unique combinations are more approximated empirically than attained. For example, Bailey (1994) argues unique combinations are only attained in conceptual, not empirical, typologies.)

From the OM perspective, the dimensions being captured may not be obvious. For example, in matching people who stop and start homeless episodes at the same time, OM captures homeless duration — how much homelessness an individual has endured — but it does so not by grouping individuals based on the number of nights they were homeless but by grouping them based on similarities in how long each episode of homelessness and non-homelessness last, when they occur in the sequence and what comes before and what comes after each episode.

[10] These deviations are squared Euclidean distances. The Ward's algorithm in the clustering software we use, ClustanGraphics 8.0, is able to treat the Levenshtein distances produced by optimal matching as, effectively, Euclidean distances. See Wishart 1969.

[11] Kuhn and Culhane use nearest centroid sorting because it clusters large datasets like theirs (n = 73,263). Our much smaller dataset removes this concern. Using a different clustering technique is not important: our analysis of the effects of different ways of conceptualizing and measuring time-based concepts is unaffected by clustering technique. Having said this, we do derive a three-group solution using nearest centroid sorting and find no statistically significant differences between the two techniques for group size, days sheltered and average number of shelter episodes.

[12] We validate this solution by drawing another simple random sample of 5,000 people from the full 40,169 person population and carrying out the same analysis. We find a time-patterned three-group solution that was similar to what we report, though we know of no statistical tests for differences.

[13] For instance, case A-1 was sheltered for 27 days the first month, 10 days the second month and not ever sheltered again, all of which is consistent with expectations for transitional homelessness. Cases A-2, A-3 and A-4 are also consistent with these expectations and are included in the group of illustrative cases from which we draw the transitional sample.

[14] More specifically, we disproportionately sample the population of illustrative cases in each group, where the sample size is determined by the smallest number of illustrative cases available for any one group — 84 in the episodic group. The dissimilarity matrix was n by m, where n is all members of a group and m is the sample of illustrative cases. That is, m = 84 for all groups and n = .745*5000 for

Group A, .196*5000 for Group B and .061*5000 for Group C. In addition, since the observed dissimilarities are between relevant pairs of cases, we calculate averages based on the number of pairs of relevant cases in each group, making the denominator n(n-1)/2.

[15] The initial findings for group size, average episodes and average days sheltered are: transitional — 81.0%, 1.4 and 57.8; episodic — 9.1%, 4.8 and 263.8; chronic — 9.8%, 2.3 and 637.8.

[16] Total sum-of-squares is the total within-group sum-of-squared deviations from the cluster centroid (cluster average) for the one-group solution, i.e., when all cases are grouped into one cluster. Our measure = total ss/n. The same logic is used to calculate total dissimilarity. Note that this is not the same total sum-of-squares statistic as appears in Table 2, though they are related.

[17] We do not compare homogeneity across the two approaches. How the time-patterned approach employs time to capture frequency and duration creates greater opportunities for variability than the time-aggregated approach, making comparison untenable.

[18] We choose a 99% confidence interval because, across the covariate analyses, it is the most demanding interval that generates statistically significant results. A 95% interval produces more statistically significant odds ratios in each of the two approaches.

[19] Kuhn and Culhane do not explicitly theorize an age relationship for the transitional/episodic relationship, but we do find the transitional group is younger and the episodic group older. Neither do they explicitly theorize transitional category relationships with having any problem or with having all three problems. Such theorizing, however, seems to follow from the logic of their argument and is empirically explored by them. What these latter relationships should look like for the chronic/episodic relationship is not clear, since chronically homeless people are expected to have more medical problems and episodic homeless people more mental health and substance abuse problems.

[20] To be clear: the extant research suggesting types of temporal-based homeless was theoretically weak, not Kuhn and Culhane's expression of this research. McAllister et al (2009) describes more extensively some theoretical and empirical problems with this typologizing.

[21] An 18-group solution also meets these criteria, but this solution produces the same four substantive patterns as the ten-group solution and only slightly improves homogeneity (e.g., total explained sum-of-squares is 76%, rather than 71%). Valuing parsimony, we report the ten-group solution, although it is possible that for some purposes, e.g., program operation, an 18-group solution may be more useful.

[22] These results support the thrust of Kuhn and Culhane's findings of three basic groups when using a time-aggregated approach.

[23] We emphasize the causal importance of structural conditions since those conditions "select" certain people from a population of individuals, rather than individuals intentionally "selecting" circumstances from a population of structural conditions. This last is not impossible, however, just less likely. Two examples are when people move between towns to take advantage of local conditions they perceive to be more beneficial (e.g., guaranteed shelter) or remain sheltered to gain non-shelter benefits (e.g., guaranteed housing after a certain amount of shelter time).

[24] We do not try to theorize based on why people initially enter shelters. Our empirical analysis — and hence our typology — is based on histories of shelter-use; it does not incorporate people's lives before they first entered shelter.

[25] For any individual, shelter conditions can be a useful explanation relative to what a person imagines his or her non-shelter living conditions to be. We use "shelter conditions" as short-hand, implicitly arguing that, for example, to remain sheltered, these conditions are more "valued" than those outside shelter. Some individuals with substance abuse problems, for instance, may prefer the institutionalized housing of shelters to the institutionalized housing of jail.

[26] Here and in the remainder of this section, we discuss individual traits not to focus on them, but also to draw attention to possible structural conditions and to the relationship between individual qualities and

structural conditions. Since we do not have data on structural conditions, we glimpse them through the combination of the kinds of people they select for and their shelter histories. The language of "selecting for" makes conditions causal; that of "people choosing" makes individuals causal. We believe both are operating, but have no theory or data on how they interact to produce shelter histories

[27] Kuhn and Culhane describe similar people in their episodic category.

[28] The number of categories (or groups) is a short-hand way to talk also about category size and nature. Choosing the number of groups also determines their size and nature, since choosing one number of groups over another means a different size for at least one group and introduces at least one group with a different nature.

Group A

Cases	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
A-1	***27***	***1***																																		
A-2	1																																			
A-3	1																																			
A-4	2																																			
A-5	8																																			
A-6	2																																			
A-7	3	3	3	15																																
A-8	2	9	3	4																																
A-9	28	28	12	11	28	19	28	18																												
A-10	28	27	4						6	27	1																									

Group B

Cases	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
B-1	3	3	3	3	3	3	3	3	3	3	9																									
B-2	6	8	29	3	3	29	3	3	3	3	3	3	3	1																						
B-3	29	29	3	3	28	29	3	28	29	29	27	15	27	6																						
B-4	3	3	3	3	3	3	3	3	3	3	3	3	2																							
B-5	26	3	3	3	3	29	3	3	3	3	3	3	3	3	3	3	3	23																		
B-6	***27***	***28***	***27***	***29***	***3***	***28***	***3***	***3***	***22***	***15***	***28***	***21***	***24***	***12***			***11***	***26***	***25***	***28***	***8***	***3***		***5***												
B-7	3	3	3	3	26	3	25	3	29	3	3	3	3	3	25	3	3	3	18								28	3	5							
B-8	16	28	3	3	3	3	3	29	3	3	3	3	3	3	17																					
B-9	28	24	3	3	3	24	29	17		2	3	28	3	3	28	17	1																			
B-10	3	3	3	3	3	3	3	3	3	3	3	3	29	3	3	3	22																			

Group C

Cases	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
C-1	3	3	3	7						7	27	3	2		1	6		4	22	19			12	3	3	11	15									
C-2	29	26	3	27	28	1						17	3	3	3	3	3	3	3	29	3	3	8	3		17	1			6				7	3	3
C-3	3	3	27	3	29	28	3	29	3	3	27	3	2	28	26	29	3	27	28	27	28	29	28	3	26	7										
C-4	3	3	3	3	3	3	3	29	3	3	27	3	29	27	29	3	3	3	3	3	28	29	29	29	3	15										
C-5	28	3	29	29	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	29	3	17										
C-6	27	25	27	25	28	24	28	28	28	3	29	3	3	3	3	3	3	3	3	2	23	3	3	3	3	3	3	3	27	2						
C-7	3	3	3	29	3	3	29	21	13	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	12								
C-8	3	1	3	3	3	3	6			1	3	3	29	3	16	3	3	3	3	3	3	3	29	3	3	14	23	3	3	3	3	3	27	29	22	
C-9	29	3	3	3	3	3	3	3	3	3	3	3	3	2							17	3	29	28	3	3	3	29	29	3	3	29	3	3	3	3
C-10	***22***	***29***	***3***	***3***	***3***	***3***	***26***	***3***	***19***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***3***	***29***	***3***	***18***	***17***	***21***	***26***	***18***	***19***	***3***	***3***	***3***	***3***

Note: Blank spaces represent zero values, i.e., no days in shelter.

Table 2. Average Total Within-group Distances for Theorized Illustrative Group Types

A. Theoretical Expectations

	Theorized Groups		
Idealized Group Types	**Transitional**	**Episodic**	**Chronic**
Transitional	0.0	> 0.0	> 0.0
Episodic	> 0.0	0.0	> 0.0
Chronic	> 0.0	> 0.0	0.0

B. Empirical Findings

	Empirical Groups		
Illustrative Group Types	**Group A**	**Group B**	**Group C**
Transitional	*7.3*	50.2	105.2
Episodic	51.7	*45.4*	70.8
Chronic	120.8	79.0	*25.6*

Table 3. Time-aggregated Three-group Solution

	Groups		
Measures	**A (Transitional)**	**B (Episodic)**	**C (Chronic)**
Percent N	73.4%	4.6%	22.0%
Avg # episodes	1.28	4.67	1.52
Avg # days sheltered	44.7	243.5	509.1

Table 4. Homogeneity of Three-group Typology

A. Time-patterned Analysis
Within-group Sum of Squares (SS)

Groups	Raw Total SS	Average SS	Ratio of Group Avg to Total Avg
Transitional	38,576	10.4	0.33
Episodic	30,756	31.6	1.01
Chronic	8,792	28.8	0.92
Total within three-group SS	78,123		
Total dataset (one-group) SS	156,502	31.3	
Proportion explained of total SS	0.50		

B. Time-aggregated Analysis
Within-group Sum of Squares (SS)

Groups	Raw Total SS	Average SS	Ratio of Group Avg to Total Avg
Transitional	751	0.2	0.20
Episodic	231	1.0	1.00
Chronic	897	0.8	0.82
Total within three-group SS	1,879		
Total dataset (one-group) SS	4,999	1.0	
Proportion explained of total SS	0.62		

Note: The difference in raw total sum-of-squares is due to the greater opportunity for variation in the time-patterned approach. It measures duration and frequency over 36 time periods, while the time-aggregated approach collapses these measures into, effectively, one moment. Thus, there is more opportunity for people to differ on these measures, and measures and in this dataset they do.

Table 5. Theorized Covariate Relationships for Time-patterned and Time-aggregated Approaches

A. Time-patterned Analysis

	Transitional/Chronic				*Transitional/Episodic*				*Chronic/Episodic*			
Covariates	**Theory**	**OR**	**99% CI**		**Theory**	**OR**	**99% CI**		**Theory**	**OR**	**99% CI**	
Age < 30	> 1.0	2.52	1.48	4.28	?	1.38	1.07	1.79	< 1.0	0.55	0.31	0.97
Age > 50	< 1.0	0.38	0.28	0.51	?	0.59	0.49	0.73	> 1.0	1.58	1.12	2.23
Mental health problems	< 1.0	0.71	0.47	1.07 ns	< 1.0	0.74	0.58	0.95	< 1.0	1.04	0.67	1.62 ns
Medical problems	< 1.0	0.50	0.34	0.73	< 1.0	0.67	0.53	0.85	> 1.0	1.34	0.88	2.03 ns
Subst abuse problems	< 1.0	0.65	0.43	0.99	< 1.0	0.69	0.54	0.88	< 1.0	1.06	0.67	1.68 ns
Any problem	< 1.0	0.44	0.28	0.69	< 1.0	0.58	0.45	0.74	?	1.32	0.81	2.18 ns
All three problems	< 1.0	0.19	0.08	0.46	< 1.0	0.39	0.21	0.74	?	2.03	0.79	5.25 ns

B. Time-aggregated Analysis

	Transitional/Chronic				*Transitional/Episodic*				*Chronic/Episodic*			
Covariates	**Theory**	**OR**	**99% CI**		**Theory**	**OR**	**99% CI**		**Theory**	**OR**	**99% CI**	
Age < 30	> 1.0	1.69	1.31	2.18	?	1.34	0.83	2.17 ns	< 1.0	0.80	0.47	1.34 ns
Age > 50	< 1.0	0.46	0.38	0.56	?	1.13	0.74	1.74 ns	> 1.0	2.44	1.56	3.82
Mental health problems	< 1.0	0.72	0.56	0.91	< 1.0	0.64	0.41	1.00 ns	< 1.0	0.90	0.56	1.44 ns
Medical problems	< 1.0	0.54	0.43	0.67	< 1.0	1.06	0.66	1.70 ns	> 1.0	1.98	1.21	3.24
Subst abuse problems	< 1.0	0.68	0.53	0.86	< 1.0	0.61	0.39	0.94	< 1.0	0.90	0.56	1.44 ns
Any problem	< 1.0	0.49	0.38	0.62	< 1.0	0.70	0.44	1.10 ns	?	1.44	0.88	2.36 ns
All three problems	< 1.0	0.26	0.15	0.47	< 1.0	0.41	0.14	1.24 ns	?	1.57	0.50	4.87 ns

Note: "Theory" indicates theorized direction of odds ratio.
"ns" indicates odds ratio is not statistically significant at $p < .01$.
"?" indicates no or ambiguous theorizing.

Figure 1. Total Within-group Sum-of-squares (Fusion Values) and Associated T-values for Cluster Solutions 30 Through 1

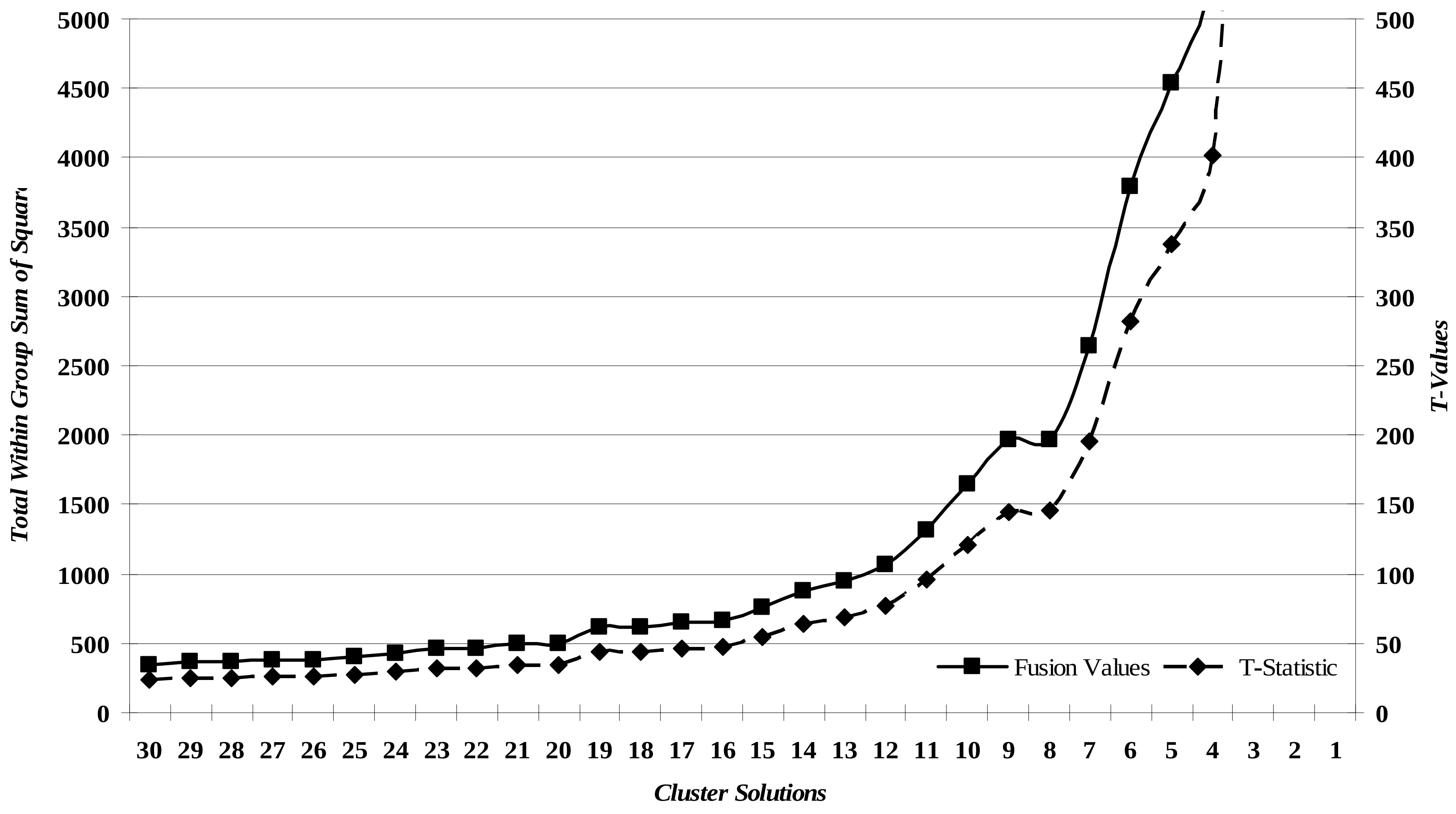


Table 6. Exemplars for Four-pattern/Ten-group Solution

Temporary

Group	*% N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
A	**37.2%**	3																																			

Structured-continuous

Group	*% N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
B	**18.8%**	28	8																																		
C	**12.1%**	29	30	30	30	15																															
D	**6.6%**	30	30	30	30	30	30	30	30	30	30	15																									
E	**6.5%**	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	13																		
F	**2.8%**	30	29	30	30	30	30	29	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	18										
G	**3.3%**	29	30	30	30	29	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
Total %	**50.2%**																																				

Structured-intermittent

Group	*% N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
H	**2.3%**	29	29	29	30	6																17	23	29	30	30	30	30	30	30	30	30	29	8			
I	**4.1%**	26	30	30	7					20	30	30	29	30	30	12																					
Total %	**6.3%**																																				

Unstructured-intermittent

Group	*% N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
J	**6.3%**	No Exemplar																																			

Note: There are too many different kinds of the unstructured-intermittent pattern to identify a meaningful exemplar for Group J. Also, blank spaces represent zero values, i.e., no days in shelter.

Appendix B: Random Selection of Cases from Unstructured-intermittent Pattern

Thirty-day Time Periods

Cases	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
J-1	15	4	1			5	13																											23	6	
J-2	2																1	28	28																	
J-3	9											3	1				12	3	15																	
J-4	18																												4	17	12	1				
J-5	5							1					11	3	7				16	3	3	3	3	2												

Note: Blank spaces represent zero values, i.e., no days in shelter.