Methods for Collecting, Measuring, and Modeling Activity Spaces

Scott T. Yabiku
Jennifer E. Glick
Elizabeth Wentz
Qunshan Zhao

Arizona State University
Methods for Collecting, Measuring, and Modeling Activity Spaces

Abstract

Individual actions are both constrained and facilitated by the social context in which individuals are embedded. But research to test specific hypotheses about the role of space on human behaviors and well-being is constrained by the difficulty of collecting accurate and personally relevant social context data. We report on a project in Chitwan, Nepal, that directly addresses challenges to collecting accurate activity space data. Our project has two aims. First, we have a methodological aim to test if a computer assisted interviewing (CAI), tablet-based approach to collecting activity space data was superior to a paper map-based approach. Our second aim was to test substantive hypotheses linking respondents’ demographic characteristics to their activity space patterns. Models from our second aim show that despite three different ways of measuring activity spaces, women have consistently smaller activity spaces than men, and people with more education have more expansive activity spaces.
Methods for Collecting, Measuring, and Modeling Activity Spaces

Introduction, Background, and Hypotheses

Individual actions are both constrained and facilitated by the social context in which individuals are embedded. But research to test specific hypotheses about the role of space on human behaviors and well-being is constrained by the difficulty of collecting accurate and personally relevant social context data (Entwisle, 2007; Sampson et al., 2002). Until recently, prior research has focused mainly on neighborhood context as the primary social context that is most relevant to behavior. Researchers, however, are increasingly looking beyond the residential neighborhood for important markers of social context. Studies from developed settings find that individuals spend relatively small portions of their days in their residential neighborhood. This may explain why some studies have found weak or no influence of neighborhood context on individual health and well-being, even when these associations have been strongly predicted by theory (Sastry & Pebley, 2010; Crowder & South, 2011; Wodtke et al., 2011). Measuring the social context at individuals’ various activity spaces, which are not bounded by the residential neighborhood, may more realistically capture the social context to which individuals are exposed. For example, Calder et al. (2011) used the detailed location data in the LAFANS to create a spatial map of individuals’ activity spaces. This allows a broader conceptualization of social context, because it is not necessarily tied to the respondent’s neighborhood but reflects the space in which individuals live their day-to-day experiences (Kwan et al., 2008). Inagami, Cohen, and Finch (2007) argued that ignoring exposures to social context outside an individual’s neighborhood may bias estimates of associations between social context and health outcomes.

Measuring activity spaces requires detailed information on where each individual participates in various domains of activity, such as consumption, education, production, and recreation. Collecting highly accurate activity space data can be difficult. For example, the LAFANS project in Los Angeles is a
leading study for examining links between place and health, socioeconomics, and family well-being. LAFANS asked respondents for precise location data (numeric addresses or cross-streets) on the physical location of places respondents frequented, such as grocery stores, churches, and prior residences. Even though activity space data were requested of places just within Southern California, only about three-fourths of activity locations were successfully geocoded (Pebley & Sastry, 2004, p. 39).

Collection of activity space data is even more challenging in areas that do not use numeric address systems, such as developing settings. Portable, wearable GPS data loggers and mobile phones offer the possibility of continual monitoring of activity spaces and highly accurate spatial measurement, but these technologies have drawbacks. The battery life of most of these devices is not long enough to be deployed without recharging by the respondent during data collection (Vazquez-Prokopec et al., 2009). Ideally, activity spaces would be recorded for at least a week, given the differences between weekday and weekend behavior. No currently available wearable GPS data loggers can operate continually without a charge for an entire week. Not only does recharging require participation from the respondent but it may not even be possible for data collection in populations without reliable electricity sources. In addition, population subgroups with limited familiarity to mobile technology, such as older individuals, may not be comfortable recharging the devices on their own. Using the technology of respondents’ mobile phones is a promising alternative for research with some subgroups in some locations (Matthews, 2011). But mobile phone penetration is far from universal, and respondents may not agree to participate in a project that they perceive as violating their privacy (even the project passes ethical and IRB review). Finally, a respondent’s personal mobile device cannot assist the collection of retrospective data (unless the researcher can convince the respondent and/or data carrier to release administrative or automatically collected data).

Because the activity spaces cannot be referenced by an address in many developing settings, researchers in these settings have continued to rely primarily on measuring social context that is tied to
the residential neighborhood. Surveys in developing settings often ask if there is or is not a certain type of organization or infrastructure in the respondent’s community, such as a school, hospital, or road, but the boundary of the community is often artificial. Alternatively, some data collections ask how far away the nearest organization is. This approach is taken by the Chitwan Valley Family Study, Mexican Family Life Survey, Indonesian Family Life Survey, and the Latin American Migration Project. In rapidly developing settings, however, there are often many schools, markets, and health clinics in and around a given community. Knowing only the distance to the nearest organization provides a truncated measurement of context devoid of activity usage information.

In this paper, we report on a project in Chitwan, Nepal, that directly addresses these challenges to collecting accurate neighborhood context and activity space data in developing or rural settings where comprehensive number address systems are not used. Our project has two aims.

First, we have a methodological aim to test if a computer assisted interviewing (CAI), tablet-based approach to collecting activity space data was superior to a paper map-based approach. Our data collection plan randomly assigned CAI or paper map mode to respondents. We hypothesized that the CAI approach would yield more accurate data.

Our second aim was to test substantive hypotheses linking respondents’ demographic characteristics to their activity space patterns. Activity space research is a relatively new domain, and analytic methods are being developed to best conceptualize and analyze these spatial data. Given our knowledge of the Chitwan, Nepal, context, we constructed several preliminary hypotheses.

1) Because poor health restricts mobility, people of poorer health will have more limited activity spaces than people in better health.

2) Because Nepal is a highly gender-segregated society, women’s activities outside the home are more limited relative to men. We expect that women will have more limited activity spaces than men.
3) Although Chitwan, Nepal, is primarily agricultural, there are growing opportunities for non-farm labor, such as businesses, markets, shops, government, and industries. These non-farm opportunities usually require more education. We expect that individuals with more education will have more expansive activity spaces because they will have the opportunity to engage with more institutions and organizations outside the home. In other words, education is associated with non-farm modes of production and greater integration with services and organizations beyond the home.

Above, our hypotheses have not specified what we mean by “more limited” or “more expansive” activity spaces. In our methods section, we describe several operationalizations we use to convert our activity space spatial data into variables for our analyses.

The data have recently been collected, and in this brief abstract we overview the data collected, present some descriptive statistics, and discuss the planned multivariate analyses for the complete paper.

**Setting, Data, and Methods**

Since 1996, the Chitwan Valley Family Study (CVFS) has extensively measured social change and family behaviors in the Chitwan Valley of Nepal. The Chitwan Valley is 450 feet above sea level, about 100 miles south-west of Kathmandu, the capital city of Nepal. Chitwan is located in the Terai, a region of low-lying plains along the southern borders of the country. This is an ideal location for our proposed project. First, the study site is typical of many developing areas and does not use numeric address systems, which makes an excellent location for testing our instrument. Second, there is significant variability across Chitwan, ranging from urban to very rural areas, which allows us to test the instrument in areas of dense and sparse contextual features and activity spaces.
In Summer/Fall 2015, we fielded a household, face-to-face survey of approximately 1400 individuals. The survey consisted of both a standard, interview-led structured questionnaire and a spatial activity space component. The activity space component asked respondents about the following activities of everyday life: going to school, shopping, visiting a health provider, restaurants, place of employment, worship, recreation, visiting friends/relatives, banking, clubs/groups, and visiting local government offices. For each activity respondents were asked if they did the activity in the past week, how many times in the past week they did that activity, and where the activity was located. If a respondent did not do an activity in the past week (for example a respondent may have a usual doctor, but did not visit the doctor recently), the respondent estimated when he/she last did the activity, and indicated the location of the activity. How the location of each activity was measured was randomly assigned to respondents: tablet or paper.

Computer assisted interviewing (CAI) tablet. Approximately 2/3 of respondents recorded the activity location using a CAI approach. The tablet program was custom designed and programmed by our team in C#, and it was administered on Windows 8 tablets. The CAI interface screen was mostly satellite imagery and allowed respondents to easily locate themselves on the map. The interface was fluid, pannable, and zoomed in and out easily, much like any modern tablet app. Although the interviewer was always present, the CAI interface was designed to be approachable and easy to use, even for a non-expert. The goal was to increase respondent engagement and make the interview collaborative. Figure 1 shows the CAI interface.

(Figure 1)

Paper map. Approximately 1/3 of respondents recorded activity location on a paper map. This was a large rolled-up map of the Chitwan study area; the map was approximately 4 feet wide by 8 feet long. This allowed respondents to see the general area of Chitwan and identify major intersections of roads, but not enough to identify individual buildings.
**Methods for assessing Hypothesis 1: Data accuracy of CAI versus paper approach.** To assess the accuracy of the location data collected with the tablet versus paper modes, we need some reference point of the actual locations respondents intended to mark. When we collected the locations, in both tablet and paper mode, we also asked for the complete name and address information for a subset of activities. In Fall 2015, we are currently “ground truthing” the activity space locations for a subset of respondents (we plan to ground truth the activity locations of 400 tablet respondents and 400 paper respondents). Interviewers will visit the activity mentioned in the data (e.g., a school or restaurant or bank) and use a GPS unit to mark the location of the activity. This will serve as the “gold standard” reference point. Distance between this reference point and the respondent’s location (marked via tablet or paper) will serve as the dependent variable for assessing accuracy in assessing the accuracy of paper or tablet mode. The ground truth data is expected to be completed this fall.

**Methods for assessing Hypothesis 2: linking respondents’ demographic characteristics to their activity space patterns.** Hypothesis 2 involves substantive hypotheses about demographic background factors and activity space patterns. We predicted that some background factors would be associated with “more limited” or “more expansive” activity spaces. There are multiple ways to conceptualize these activity spaces, but here we present three different ways of thinking about how to create variables from the spatial activity space data.

In our initial analyses, we have restricted the activities to only those that happened to the respondent in the last week. In addition, we have begun using only the data from the tablet mode, as these (N=864) have been cleaned and available for our use. Consider a respondent whose activity spaces in the past week appear as in Figure 2. Note: we have removed the base layer satellite and road imagery from this Figure in order to protect respondent confidentiality. To give a sense of the scale, the longest distance between two points here is about 3.5 km.
There are different ways to approach transforming these spatial points into an assessment of the extent of this respondent’s activity space. We have thus far explored three approaches:

1) Number of different activity locations. This approach is the simplest and it ignores the spatial information in the data. If one conceptualizes activity space as the diversity of different locations a respondent occupies, then a simple count of locations provides variation in activity spaces. As shown in Figure 2, Panel A, the respondent indicated 15 different locations in the past week.

2) Convex hull. The simple number of different activity locations ignores the spatial distribution of activities. Two respondent may have the same number of activities, but if one respondent’s activities all took place within the same ½ square kilometer area, then that respondent’s activity space would be thought of as more limited than a respondent whose activities were spread out across 10 square kilometers. A convex hull is the minimum convex polygon that contains a set of features, such as points. In short, it takes a collection of points and creates a measure of area. In Figure 2, Panel N, the area of the convex hull is 4.7 square kilometers.

3) Average nearest neighbors (ANN). ANN calculates the average distance between each point and all neighboring points. This results in a vector of average distances, the length of which is the number of total points. We then take the average of this as a way to summarize how tightly clustered the points are. The mean of the average nearest neighbors for the points in Figure 2 is about 300 meters.

Each of these operationalizations have drawbacks. The simple count ignores spatial information. The convex hull is very sensitive to outliers, and cannot be performed on sets of activities with only 1 or two points. The average nearest neighbors is slightly less sensitive to outliers, but still can be affected by a stray location that is far from the main cluster. Given the relatively newness of activity space data, the literature does not yet offer clear guidance on which approaches are ideal.
The demographic background factors we use in our preliminary models are self-rated health (1 to 5 scale), gender, and years of education. We also control for age, employment status (non-farm employment outside the home), marital status (currently married or not).

Results

In the current paper, we present a few descriptive statistics and simple models. By the time of PAA, we will have the ground truth data as well as a more comprehensive set of multivariate models. (Table 1)

In Table 1, we present the descriptive statistics for different conceptualizations of activity spaces. On average, respondents visited 2.7 unique locations in the past week. Of course, our survey measured only a limited set of locations (school, employers, clinics, etc.), and it is likely that respondents engaged in activities that we did not assess. The convex hull bounding these activities averaged 2.0 square kilometers and the average of the average nearest neighbors between all activities of a respondent was 1.2 kilometers.

Although these are different metrics, there are some similarities. Namely, activity spaces are highly variable. The maximum values are many times greater than the mean. For the spatially explicit measures (convex hull and average nearest neighbors), the standard deviation is greater than the mean. In the Chitwan context, this high variability is expected. Chitwan is agricultural and many respondents will have modes of production and daily lives that are centered on the house and farm. Although there is bus service and motorbikes are common, cars are a rarity among most families, and walking or bicycles are the main mode of transportation. This setting is very different than, for example, the United States where commuting for work is common and driving for shopping, pleasure, and school is the norm. Thus life in Chitwan is more confined spatially, although there are individuals who do travel greater distances. (Table 2)
Table 2 present the preliminary regression results. Several common themes emerge from the models, even though each dependent variable is a different approach to measuring activity space. The most consistent predictors are gender and education. Women in Chitwan have significantly more limited activity spaces than men: they had activities in .3 less locations in the prior week, the bounded area in which their activities took place were 1.7 square km less than men’s, and the average distance between their activities was 800m less than men’s.

A factor that consistently expanded activity spaces was education. Respondents with more education had more activities in different locations and had activities spread across greater areas. They had greater average distance between activity locations, indicating that their activities were less tightly clustered than people with less education.

Health did not have a consistent association with activity spaces. Self-rated health was not related to the number of activities or the convex hull bounding area, but better health was positively and significantly associated with greater distance between activities (average nearest neighbors).

The remaining control predictors in the model (age, non-farm and non-home employment status, and marital status) also showed somewhat inconsistent relationships. Older people, people with non-farm employment, and married individuals had significantly more activity locations. Being married was associated with significantly larger bounding areas, but age and employment did not predict bounding areas.

Discussion and Next Steps

Our next steps are to analyze the ground truth data and further develop our multivariate models. We are encouraged by the consistent effects of gender and education on diverse conceptualizations of activity spaces. These associations are strongly grounded in theory and also have intuitive appeal.
The results for the other predictors are not as consistent, but as we develop more refined measures of activity spaces from our data we hope to better capture the spatial distribution of human activity.
References


### Table 1: Descriptive Statistics of Activity Space Distributions

<table>
<thead>
<tr>
<th>Respondent's prior week activity space area</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>As measured by number of different locations</td>
<td>2.7</td>
<td>1.8</td>
<td>0.0</td>
<td>15.0</td>
</tr>
<tr>
<td>As measured by convex hull (sq km)</td>
<td>2.0</td>
<td>7.2</td>
<td>0.0</td>
<td>108.7</td>
</tr>
<tr>
<td>As measured by average nearest neighbors (km)</td>
<td>1.2</td>
<td>2.5</td>
<td>0.0</td>
<td>19.6</td>
</tr>
</tbody>
</table>

N=864 respondents
<table>
<thead>
<tr>
<th>Variable</th>
<th>Locations in the Prior Week</th>
<th>Convex Hull Bounding Area</th>
<th>Average Nearest Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Rated Health</td>
<td>-0.047</td>
<td>-0.264</td>
<td>155.101**</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.228)</td>
<td>(77.401)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.348**</td>
<td>-1.688***</td>
<td>-800.296***</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.572)</td>
<td>(194.312)</td>
</tr>
<tr>
<td>Years Education</td>
<td>0.069***</td>
<td>0.231***</td>
<td>77.133***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.065)</td>
<td>(22.132)</td>
</tr>
<tr>
<td>Age</td>
<td>0.015*</td>
<td>0.017</td>
<td>11.650</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.033)</td>
<td>(11.063)</td>
</tr>
<tr>
<td>Non-Farm Employment</td>
<td>0.268*</td>
<td>-0.588</td>
<td>-19.817</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td>(0.665)</td>
<td>(225.993)</td>
</tr>
<tr>
<td>Married</td>
<td>0.485***</td>
<td>1.525**</td>
<td>-170.264</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.700)</td>
<td>(237.926)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.711***</td>
<td>0.607</td>
<td>438.389</td>
</tr>
<tr>
<td></td>
<td>(0.351)</td>
<td>(1.444)</td>
<td>(490.763)</td>
</tr>
</tbody>
</table>

N=864

*p<0.1; **p<0.05; ***p<0.01
Figure 1: Computer Assisted Interview Activity Space Tablet Interface
Figure 2, Panel A: Activity Locations in the Past Week for a respondent

Figure 2, Panel B: Convex Hull of Activity Locations in the Past Week for a respondent