Implications of Geographic Information Systems (GIS) for targeted recruitment of older adults with dementia and their caregivers in the community: A retrospective analysis

Danny L. Scerpella, Atif Adam, Katherine Marx, Laura N. Gitlin

A B S T R A C T

5.5 million Americans are living with Alzheimer’s dementia (AD) or related dementias. Developing evidence-based interventions for these people and their caregivers (dyads) is a public health priority, and is highly dependent on recruiting representatives from the community. Precision recruitment methodologies are needed to improve the efficiency of this process. Geographic Information Systems (GIS) offer the potential to determine location trends of an older adult population of people living with dementia in the community and their caregivers.

American Community Survey (ACS) 2015 5-year estimates were analyzed at the census tract level in ESRI ArcMap v. 10.5.1. Datasets included summarized estimates of age, gender, income, and education in Maryland. Using a two-step process, geographic regions were identified in ArcMap that contained various combinations of available data variables. These areas were compared to participant locations from a previously completed traditional recruitment effort to determine overlap (Dementia Behavior Study - R01AGO41781).

The largest number of existing participants were identified in derived regions defined by combining age, education, gender, and income variables; predicting 184 (79%) of 234 participants regardless of the population density within census tracts. 208 (89%) were identified when matching this variable combination to the highest density census tracts (city/urban), and 66 (28%) in regions with the lowest population density (rural).

This study successfully defined specific geographic regions in the state of Maryland that overlapped with a large number of known dementia dyad locations obtained via traditional recruitment efforts. Implications for these findings allow for more targeted recruitment efforts of difficult to recruit populations, and less utilization of resources for doing so.

1. Introduction

An estimated 5.5 million Americans are living with Alzheimer’s dementia (AD) or related dementias and the prevalence of dementia is expected to nearly double by 2050 [1]. Thirty percent of these older adults rely on three or more unpaid caregivers [1]. A public health priority is to develop an evidence base for care and services that are needed by persons living with dementia and their family caregivers.

Designing and testing care interventions in clinical trials is highly dependent on recruiting eligible caregivers that are representative of the desired research population [2]. Most people with dementia and their family caregivers (dyads) live at home, and successfully recruiting them to clinical trials is crucial. Recruitment efforts to involve these dyads are varied, including clinical referrals and mailing lists. Setbacks in recruitment reduce study efficiency, increase associated costs, and can lead to the abandonment of trials altogether [3,4]. Evidence exists that postal correspondence is the most effective way to reach potential participants in the community [5,6].

The ability to analyze and identify geographical trends in the aging population could allow for more precise recruitment efforts. One strategy that may be efficient and productive in this endeavor is the use of Geographical Information Systems (GIS). GIS allows for the integration of population demographics with locational data, and provides visual representations of multiple, complex data points, allowing for the application of spatial analytical processes to locate potential study participants with pre-specified characteristics [7,8].

Applications of GIS for gerontological populations have steadily grown in the last two decades with the rise of technological advances in GIS and ease of access to the programs. GIS has also been used for various population-level analyses including those identifying specific geographic regions of potentially vulnerable older adults [9–12]. No current literature has been identified at the time of this writing that describes the usage of GIS methodologies to assess improved recruitment strategies of individuals with dementia or their caregivers with an...
emphasis on potential cost savings.

The purpose of this paper is to demonstrate the ability of GIS methodologies to leverage geographic trends in recruitment of an older adult population of people living with dementia in the community and their caregivers. It is expected that areas containing a greater density of older adults and a high prevalence of risk factors associated with the probability of having dementia or being the caregiver of a person with dementia will have a strong association with the locations of a large percentage of participants currently enrolled in a clinical trial involving dementia and dementia care. Through retrospective analysis, we illustrate the potential utility of using GIS as part of a targeted recruitment strategy for reaching populations that are challenging to include in dementia care clinical trials.

2. Methods

2.1. Data source

Study sample: This study included data from 250 family caregivers and persons living with dementia who were enrolled in a National Institute on Aging supported trial, referred to as the Dementia Behavior Study (NIH R01AG041781 NCT01892579). This trial tested the efficacy of a novel tailored activity intervention to reduce common neuropsychiatric symptoms for community-dwelling persons with dementia [13]. Participants were recruited from a large region including in Maryland and Washington DC through multiple means including radio advertisement, paper mailing, community outreach, restaurant placemat advertising, newspaper advertisements, and referrals from partner studies. For the present study, participants from Washington DC were excluded from the geographic analysis to simplify the analysis solely to the state of Maryland in which the majority of participants were recruited. A total of 234 participants were included in the present investigation and are described in Table 1.

Both persons with dementia and their caregivers needed to meet eligibility criteria to be included in the study. Persons with dementia were eligible if: 1) they spoke English, 2) had a physician diagnosis of dementia (mild, moderate, severe), 3) were able to participate in at least two activities of daily living (e.g., bathing, dressing, grooming, toileting), and 4) demonstrated at least one agitated and/or aggressive behavior as measured by the caregiver-reported NPI-C.

Caregivers were eligible who: 1) spoke English, 2) were a family member (broadly including neighbors and fictive kin), 3) lived with the person with dementia or within 5 miles or 15 min travel time, 4) were accessible by telephone to schedule interviews and intervention sessions, and 5) were planning to live in the same area for at least six months to reduce loss to follow-up.

Geographic data: The dataset used for this investigation consists of georeferenced digital layers at the census tract spatial scale retrieved from the publicly available American Community Survey (ACS) 2015 5-year estimates from the U.S. Census Bureau. Analyses were performed at the census tract level due to this being the finest geographic scale available from this dataset. The study area consisted of all areas within the state of Maryland which make up 1395 census tracts with population size varying between 22 and 14,953 per census tract.

3. Analysis

Census data were imported into ESRI ArcGIS v. 10.5.1. Analyses were conducted in 3 steps: (1) Census data preparation, (2) identifying key population characteristics that inform recruitment efforts, and (3) analysis of population catchment areas.

Census data for the state of Maryland were extracted at the census tract geographica level. These datasets included summarized estimates of population distribution by age, gender, median household income, and education levels for each census tract. This data was then matched to census tracts in the ArcGIS software.

Previous studies have identified age, gender, income, and education levels of adults as variables of risk for either having dementia (age, education, gender) or being a caregiver of someone with dementia (income, gender) [1,14–17]. These variables were therefore targeted for our recruitment effort. We used a grouping analytical approach in ArcGIS to derive optimal numbers of high and low prevalence groups for each inclusion criteria variable (age, gender, income, and education level) and the combination thereof. Based on a k-mean clustering algorithm, the Grouping Analysis tool stratifies each variable within a respective census tract into a number of groups so that all the variables within each group are as similar as possible (e.g., high prevalence in one group, another with only low), while all the resulting groups themselves are as different as possible. The number of groups generated by this analysis were user-defined yet significant, verified by referring to a pseudo F-statistic, which ensures a statistically meaningful number of output groups for each stratified variable [18]. Groups with the desired prevalence of each variable (representing the global upper quartile for age, gender, and education data sets and global lower quartile for the income data set) created by the Grouping Analysis tool were selected, and the corresponding census tracts were used for further analysis.

We then compared geocoded participant address data from the Dementia Behavior Study with the identified census tracts resulting from the Grouping Analysis, thus allowing for detection of relevant catchment areas. We define catchment areas as regions wherein participant addresses overlap with selected census tracts with population characteristics stratified and selected via Grouping Analysis.

We used a two-step design to analyze catchment areas. Step 1, by isolating census tracts identified using the Grouping Analysis, we identified Dementia Behavior Study participants residing within these census tracts using an overlay selection tool. This feature allows for a count of participants living within these regions. Next, we applied a ‘spatial buffer’ of 0.5 miles around each census tract polygon identified in our Grouping Analysis to account for participants living in close proximity to identified areas, but just outside of the boundaries of the tract. Spatial buffers increase the representative area of a polygon (census tract), in this case encompassing all area 0.5 miles in Euclidian distance in all directions from the boundary of each selected census tract in ArcMap. Given the reduced mobility of older adults in the US, the size of these buffers is reasonable, as supported by Ref. [19] who report that age is inversely associated with physical activity outside of a 0.5 mile residential buffer. This is further supported by Ref. [20] in a review of built environment and healthy aging in which it was found that in GIS-based walking studies, buffers around participant’s homes range from 100 m to 1,000 m scales suggesting an assumption that older

Table 1

| Background characteristics of study sample (N = 234). | Caregiver | Person with Dementia |
|----------------------------------------------------|-----------|----------------------|
| Gender (%)                                         | 43 (18.4%)| 83 (35.5%)           |
| Male                                               | 191 (81.6%)| 151 (64.5%)          |
| Female                                             |           |                      |
| Age in years Mean (SD, range)                      | 65.3 (12.8, 28-93) | 81.5 (7.9, 56-99) |
| > High School                                      | 10 (4.3%) | 47 (20.2%)           |
| High School                                        | 35 (15.0%)| 67 (28.8%)           |
| Some college/Associates                            | 85 (36.5%)| 42 (18.0%)           |
| College Degree                                     | 40 (17.2%)| 33 (14.1%)           |
| Post-graduate                                      | 63 (27.0%)| 44 (18.9%)           |
| Race % **                                          |           |                      |
| White, non-Hispanic                                | 135 (59.0%)| 139 (60.2%)          |
| Black, non-Hispanic                                | 86 (37.6%)| 87 (37.7%)           |
| Native American                                    | 1 (0.4%) | 1 (0.4%)             |
| Asian                                              | 2 (0.9%) | 0                    |
| Native Hawaiian                                    | 2 (0.9%) | 1 (0.4%)             |
| Other                                              | 3 (1.3%) | 3 (1.3%)             |

Note: *N = 233; **N = 229 for caregivers, N = 232 for persons with dementia.
adults are more influenced by their proximal environment. Step 1 was repeated and catchment areas were noted for each risk variable and their combination, as shown in Table 2a. 

Step 2: In order to account for areas of high and low clustering, we used tract-specific population density to look at the effect that geographic distribution of adults over 65 years of age has on the definition of census tracts identified via Grouping Analysis in Step 1. Each census tract was compared with its own specific census tract population density of persons 65 years and older, and those within two standard deviations of the state average (−0.5 to 1.5 SD range) were selected for further analysis of catchment areas using the overlay selection tool. Census tracts containing the appropriate density of adults aged 65 or older were isolated and again compared with the locations of Dementia Behavior Study participants, yielding catchment areas based on both risk variables as well as population density. To account for the potentially overlooked population of people aged 65 and older living in lower density locations such as rural and suburban areas, Step 2 was repeated with a focus on areas containing a density of people aged 65 and over that is less than −0.5 standard deviations from the state average. Each census tract was subsequently analyzed with consideration of population densities that were within two and a half standard deviations (−0.5 to 2.3 SD range) of the less dense population selection (areas with densities less than −0.5 standard deviations of the state average for those 65 and older).

4. Results

Age (65 and older), gender (female), education (high school degree or less), and income (global lower quartile of all incomes) were categorized into statistically significant geographic groups wherein each variable was highly prevalent. Census tracts grouped for age with an applied 0.5 mile buffer accounted for 133 (57%) of 234 participants. Analysis of census tracts grouped for education identified 151 (65%) of the 234, gender identified 111 (47%) of 234, and income identified 147 (35%) of 234 participants (Table 2). Combined groupings for age and education yielded census tracts accounting for 171 (73%) of the 234 participants. Age, education, and gender combined identified 179 (76%) of 234 participants. The largest number of participants were identified in census tracts grouped by a combination of age, education, gender, and income with 184 (79%) of 234 participants (Table 2a) (Fig. 1).

Trends were similar with census tracts defined using population density (−0.5 to 1.5 standard deviations) and a 0.5 mile buffer. Alone, education accounted for the locations of the most participants with 188 (of 234; 80%), followed by age (178 of 234; 76%). The combination of variables in high prevalence census tracts for age, gender, education, and income together accounted for the most participant locations with 208 (of 234; 89%) (Table 2b) (Fig. 2).

When analyzed individually, age identified the most (56 of 234; 24%) participants located in defined census tracts that represent low density areas of people aged 65 and older (−0.5 to 2.3 SD of < -0.5 SD). Age, gender, education, and income combined accounted for 66 (of 234; 28%) of participants in these selections (Table 2c) (Fig. 3).

Table 2a) Number of participants identified by each combination of variables by Census tract, with and without 0.5 mile buffer. 2b) Number of participants identified by each combination factoring in −0.5 to 1.5 SD total population density by Census tract. 2c) Number of participants identified by each combination factoring in −0.5 to 2.3 SD of the smallest portion of total population density by Census tract (comprising less than −0.5 SD).

5. Discussion

Most people with dementia live at home either alone or with family [21], and as the disease progresses diminished cognitive and physical abilities signal the need for more caregiving [22]. This burden often falls to family and informal, unpaid carers [1]. Therefore, it is a critical
public health priority to improve support services for people with dementia and their families in the community. Successfully reaching people with dementia and/or family caregivers in the community to participate in meaningful and contributory research can be more challenging due to several factors such as a general decline in the proportion of dyads willing to participate in clinical trials, as well as a lack of trust in clinical researchers [23]. Recruitment efforts to involve these dyads in clinical trials typically involve mailing lists, telephone surveys, community-oriented events, or referrals from clinical settings. While a combination of varied recruitment strategies is seen as ideal [24], in a nonpharmacologic, community-based study of dementia, direct mailing was the most effective and least costly method of recruiting this population [25]. An obvious limitation of this finding is that targeting geographic regions in which a representative portion of potential participants actually reside is difficult to accomplish, especially when taking into account the shifting of community compositions and demographics over time.

GIS have been leveraged for a number of topics including locating lead and disease exposure “hot spots” [26], racial population disparities in access to healthcare [27], analyzing crime patterns in metropolitan areas [28], and even recruitment of targeted populations to relevant studies [8,11]. Specifically concerning older adults, one prominent area of study is neighborhood-level analyses of both subjective and objective factors (e.g., built environment, mobility patterns, and perceptions thereof) related to health outcomes for residents such as walkability and related physical activity [29–34]. Disparities in access to health care services has also been explored using GIS [35,36], as well as geographic patterning of chronic conditions in older adults [37,38]. There are studies that have used GIS technology to identify and analyze populations of older adults in community settings [9–12]. Studies have also leveraged GIS methodologies to improve representative recruitment and analysis of specific populations of interest to unique research questions [8,39]. This study is a novel usage of GIS methodologies utilized retrospectively to demonstrate the ability to geographically predict locations in which older adults with dementia and their caregivers reside based on simple, publicly available data.
We sought to determine the methodological effectiveness of using GIS to retrospectively demonstrate that participants recruited into a clinical dementia trial were located within identifiable geographic areas defined by analyzing specific risk variables retrieved from publicly available Census databases. This study successfully defined specific geographic regions in the state of Maryland that overlapped with a large number of home addresses of dementia dyads recruited into the Dementia Behavior Study via a traditional recruitment effort. The implications for these findings are numerous. Locating persons with dementia and their caregivers in the community is crucial to expand the representativeness of clinical trials aimed at addressing an increasingly urgent health risk in a rapidly growing older population in the United States. Having the ability to predetermined geographic locations in which high concentrations of these subjects are likely to reside in high population quantities could help focus recruitment efforts in a manner which eliminates extraneous spending in communities that do not contain the desired audience. Additionally, this study also identified regions outside the major metro areas of the state of Maryland wherein the population density of older adults was low, suggesting that these methodologies can be used to locate persons with dementia and their caregivers in potentially underserved, rural locations. This allows for a more representative recruitment sampling by identifying eligible participants in potentially under-recruited rural localities. There are also implications for targeted allocation of public health services outside of traditionally populated, metropolitan areas.

Understanding geographic trends in population recruitment efforts can offer insight regarding the efficacy of recruitment strategies across both geography and pre-identified demographic variables. Local and regional governments seeking to allocate and connect individuals with resources would benefit greatly from a more detailed understanding of the populations they serve. The methods in this study potentially allow for the prioritization of recruitment efforts based not only on just the factors analyzed for this project, but variables not considered in this preliminary examination such as quality of the neighborhood built environment, local access to healthcare resources, and other measures of vulnerability in this and other populations. It is important to understand the environmental contexts of people with dementia and their caregivers in the community in order to effectively reach them for intervention and participation in valuable clinical trials.

Utilizing freely available datasets to acquire pertinent variables related to these contexts contributes to the cost effectiveness of this methodology for recruitment strategies.

There are several limitations to the current study. Census data used in this project are based on 5-year estimates extrapolated from the 2010 U.S. Census. Due to population mobility, areas in which older adults reside may have shifted since initial data was collected and would affect the resulting models. Another limitation concerns the retrospective nature of this analysis. While findings reported in this study are promising for aiding geographically targeted recruitment efforts, utilization of this methodology in an actively recruiting study are needed to determine true cost savings and effectiveness. Finally, methodologies in this study could be strengthened with an analysis of individual-level data such as that from electronic health records detailing granular characteristics of study participants beyond simple demographics. Future research seeks to leverage the robust participant data gathered during the Dementia Behavior Study to analyze and identify individual-level trends of participants captured by the current methodology.

Acknowledgment

This work was supported by the National Institute on Aging Grant #R01 AG041781, NCT01892579.

References

[1] Alzheimer’s Association, 2017 Alzheimer’s Disease Facts & Figures, (2017) Retrieved from https://www.alz.org/documents_custom/2016-facts-and-figures.pdf.

[2] S.P. Hinshaw, K. Haagwood, P.S. Jensen, C. Kratchovil, L. Rickman, G. Clarke, et al., AACAP 2001 Research forum: challenges and recommendations regarding recruitment and retention of participants in research investigations, J. Am. Acad. Child Adolesc. Psychiatry 43 (8) (2004) 1037–1045 https://doi.org/10.1097/01.chi.0000129222.89433.3d.

[3] D.B. Hunninghake, C.A. Darby, J.L. Probstfield, Recruitment experience in clinical trials: literature summary and annotated bibliography, Contr. Clin. Trials 8 (4) (1987) 6–30 https://doi.org/10.1016/0197-2456(87)90004-3.

[4] J. Mapstone, D. Elbourne, I.G. Roberts, Strategies to improve recruitment to research studies, Cochrane Database Syst. Rev. 2 (2007) 1–24.

[5] R. Davey, S.M. Edwards, T. Cochrane, Recruitment strategies for a clinical trial of community-based water therapy for osteoarthritis, Br. J. Gen. Pract.: J. R. Coll. Gen. Pract. 53 (489) (2003) 315–317 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12879833.

[6] M. Sinclair, J. O’Toole, M. Malawaraarachchi, K. Leder, Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches, BMC Med. Res. Methodol. 12 (1) (2012) 132 https://doi.org/10.1186/1471-2288-12-132.

[7] C.I.J. Nykielfuk, L.M. Flaman, Geographic Information Systems (GIS) for health promotion and public health: a review, Health Promot. Pract. 12 (1) (2011) 63–73 https://doi.org/10.1177/15240859103854624.

[8] B.W. Strauss, E.M. Valentinier, S. Bhattacharya, M.M. Smerek, A.A. Dunham,
