Incorporating Spatial Analyses into Early Care and Education Research

Van-Kim Lin and Rebecca Madill

OPRE

Child TRENDS

OPRE Research Brief #2019-88 | September 2019

Introduction

This resource focuses on how spatial analysis can be used to understand early care and education (ECE) patterns and trends. **Spatial analysis** is an analytic method that uses location-based variables or maps to understand how places, the characteristics of places, and the people and things in places are arranged in space, as well as the reasons for these arrangements. Spatial analysis is growing in popularity in research and its uses encompass more than what is presented in this resource (see Additional Resources for other information about spatial analysis).

Spatial analysis is particularly useful for ECE research because of the localized nature of many ECE-related research topics. For example, parents can only use ECE providers located within a reasonable distance from their home or workplace. Spatial analysis allows researchers to integrate location-based information into their analyses. Additionally, findings of research using spatial analysis can be displayed visually, which is appealing to a variety of audiences. Maps that present information based on spatial analysis can be static, such as a map depicting the number of child care centers or family child care homes in a state at one point in time. Maps also can be interactive, allowing users to manipulate the type of information they want to see, such as by adding a layer of information about the quality of the child care providers or the density of poverty surrounding the programs. Further, many of the data elements necessary for spatial analysis in ECE research can be found in administrative data. For example, program locations may be found in child care resource and referral data, and census data can yield information about neighborhood demographics.

This resource was developed as part of the Child Care Administrative Data Analysis Center (CCADAC) through the Child Care and Early Education Policy and Research Analysis contract at Child Trends. The work is funded by the Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, with funds set aside for research in the Child Care and Development Block Grant Act. CCADAC works to strengthen the ability of state/territory child care administrators and their research partners to utilize administrative data to address policy-relevant early care and education research questions.

Spatial analysis encompasses multiple techniques that can be carried out with a variety of software programs. Spatial analysis techniques for ECE research include defining an area based on geographical characteristics (e.g., areas within 10 miles of a child care provider) or quantifying the relationship between two locations (e.g., the distance between a child's home and a child care provider). Researchers can use geographic information system (GIS) software programs[1] to conduct spatial analysis; these programs allow researchers to gather information from pre-existing maps, or create new variables based on spatial information and analyze spatial patterns. They can also use statistical programs[2] that have specific capabilities or commands for analyzing location-based information.

This resource first highlights three uses of spatial analysis that are common in ECE research: (1) categorizing geographical areas, (2) creating variables using spatial information, and (3) analyzing spatial patterns and associations. This is not an exhaustive list of all that can be done with spatial analysis; rather, it is a summary of common approaches in ECE research that are most suitable for ECE researchers new to spatial analysis.

The resource then presents challenges to using spatial analytic techniques for ECE research and offers tips to address those challenges. This information may be most helpful for researchers who are new to integrating spatial analysis into their research and want to understand how spatial analysis may help answer critical questions about ECE.

Common Uses of Spatial Analysis in ECE Research

This section provides an overview of three ways spatial analyses can be used in ECE research and provides examples of ECE research questions each approach can answer. For each approach, we also share an example of an ECE study.

We recommend that researchers work closely with any partners to select the techniques that will best answer jointly developed research questions. This resource focuses on spatial analytic research using state ECE administrative data, which would likely require a partnership or data-sharing agreement with one or more state agencies. Spatial analytic research could also be conducted in partnership with other local, regional, or national organizations.

Use #1: Categorizing geographical areas

Researchers and state administrators may be interested in categorizing geographical areas based on a particular characteristic or set of characteristics relevant to ECE. By categorizing geographical areas, researchers and their partners can answer questions such as:

- Are there fewer ECE providers in areas with high concentrations of poverty than in areas with low concentrations of poverty?
- Do counties with a greater uptake of child care subsidy vouchers among eligible children have a greater proportion of children from low-income families enrolled in ECE settings with higher quality ratings?
- Which counties have the largest number of license-exempt providers, as a proportion of all providers? Which characteristics of counties explain variations in the number of license-exempt providers?
- Which neighborhoods have a high percentage of young children at-risk for school difficulties but low capacity to serve these children in existing high-quality ECE programs?[3]

[1] Researchers can use a range of specialized software programs, such as ArcGIS, GeoDa, and Texas A&M Geocoder to conduct spatial analysis in ECE research.
[2] Examples of statistical programs with some geospatial capabilities include Stata, R, SPSS, and SAS
[3] This question was adapted from a report of analysis of integrated data to inform pre-K expansion in Philadelphia. Available at: https://repository.upenn.edu/pennchild_briefs/1/.

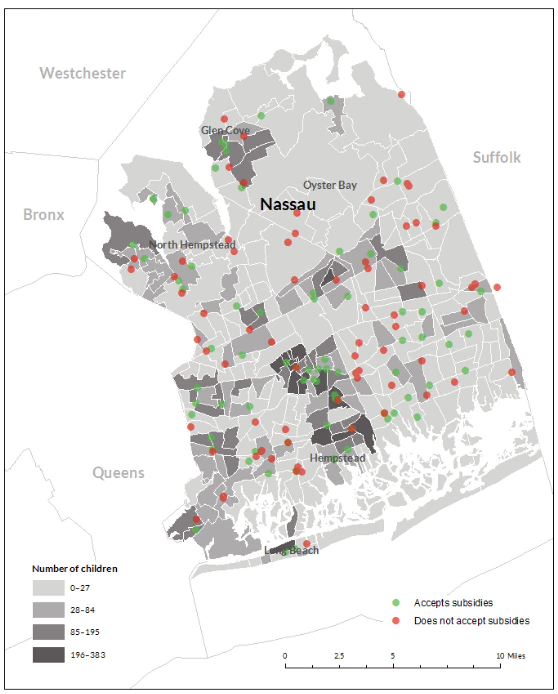
Researchers can categorize geographical areas based on a variety of characteristics, such as access to ECE,[4] density of children with an immigrant parent, density of households with limited English proficiency, urbanicity, child care subsidy reimbursement rates, uptake of subsidy vouchers, or average length of subsidy spells. Researchers can then visually examine maps generated from these categories or use this information in statistical analyses. For example, information about geographical areas can be used as predictors in spatial regressions (see Use #3).

Research spotlight: Examining local access to ECE in Illinois and New York[5]

Sandstrom et al. (2018) were interested in understanding local child care markets in target areas of Illinois and New York. They wanted to understand how child care programs were distributed and the relationship between supply and demand. Across four study counties in Illinois and New York, the researchers used geographic information to define areas with low access to ECE for subsidy-eligible families, which they labeled as child care deserts. They defined geographical areas by census tract.

The research team used administrative data from child care resource and referral (CCR&R) agencies to obtain providers' location and capacity, and linked it with additional administrative data (e.g., records of subsidy payment to providers) to describe the supply of child care in terms of provider characteristics (e.g., accepts subsidies and contracts, number of full-time slots, participated in quality improvement, obtained accreditation). First, the team used each provider's street address to obtain a latitude and longitude value, a process called "geocoding." The team then placed providers on a map using the software ArcGIS.[6] This allowed them to see the distribution of providers across census tracts. To draw attention to communities where families may have fewer care options, they color-coded the points representing providers according to different characteristics. For example, Figure 1 shows the locations of providers that accepted subsidies and those that did not in Nassau County, New York.

Figure 1. Full-time child care centers accepting subsidies, Nassau County Census Tracts



Source: Sandstrom, H., Claessens, A., Stoll, M., Greenberg, E., Alexander, D., Runes, C., Henly, J.R. (2018). *Mapping Child Care Demand and the Supply of Care for Subsidized Families.* Washington DC: Urban Institute. Retrieved at https://www.urban.org/sites/default/files/publication/97286/mapping_child_care_demand_and_the_supply_of_care_for_subsidized_families.pdf.

[4] Access to ECE can be measured in multiple ways, as exemplified by the different approaches taken by the researchers cited in this resource. For more information about measuring access to ECE, see Friese, Lin, Forry, & Tout (2017). *Defining and Measuring Access to High-Quality Early Care and Education: A Guidebook for Policymakers and Researchers*. OPRE Report #2017-08. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

[5] Sandstrom, H., Claessens, A., Stoll, M., Greenberg, E., Alexander, D., Runes, C., Henly, J.R. (2018). *Mapping Child Care Demand and the Supply of Care for Subsidized Families*. Washington DC: Urban Institute. Retrieved at https://www.urban.org/sites/default/files/publication/97286/mapping_child_care_demand_and_the_supply_of_care_for_subsidized_families.pdf.

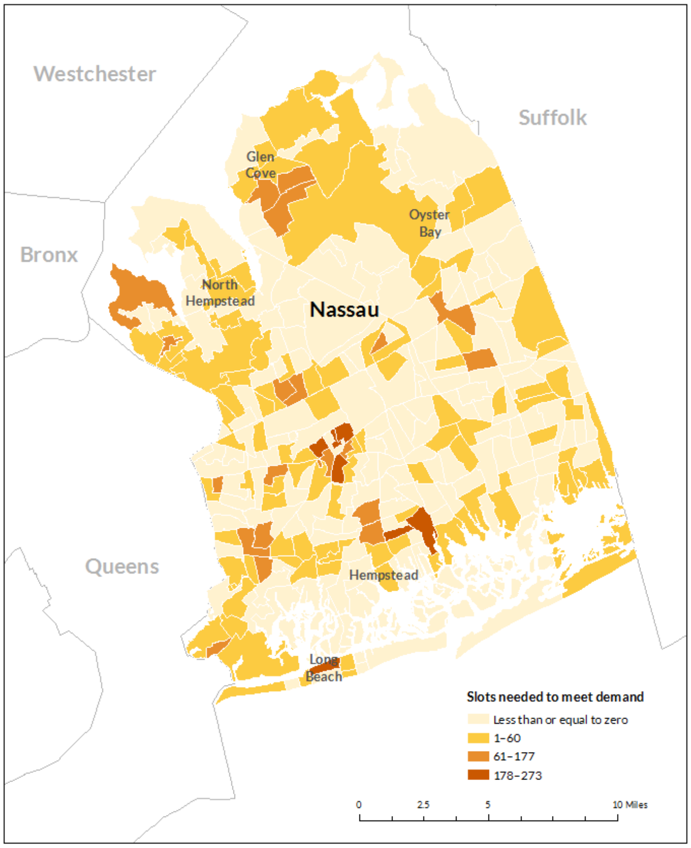
[6] The research spotlights include the software used by the researchers who conducted the analysis. Inclusion in this report does not imply endorsement or recommendation by the authors or OPRE.

Next, the team used 2010-2014 American Community Survey[7] data to estimate the number of children under age 6 from low-income families with all parents working, as a way of measuring demand for subsidized child care in each census tract. Finally, they defined high or low access as the difference between the supply (number of licensed care slots with providers accepting subsidy) and demand (number of subsidy-eligible children) for each census tract and identified child care deserts as areas with low access. With this information, the team created maps that visually displayed the spectrum of need across the county (Figure 2).

Additional maps considered parents' work hours and highlighted areas with greater demand for early morning care, evening care, and overnight care. The maps revealed that a large share of subsidy-eligible children lived in areas with few child care programs. Furthermore, even in areas that had greater access, many providers were not of high quality, licensed for infant care, or open during nonstandard hours.

The team's maps can be used to inform decisions about program development and policy interventions. For example, the information in Figure 2 suggests that access might be improved by incentivizing providers in high-need areas to accept subsidies, especially family child care providers who can serve infants and families during nonstandard hours. High-quality centers that accept subsidies might also be incentivized to expand into areas of high need. In general, maps like Figure 2 can help researchers and policymakers identify patterns in ECE-related topics that may not be as easy to detect in numerical output.

Figure 2. Additional full-time subsidized child care slots needed to meet demand, by census tract, Nassau County, New York



Source: Sandstrom, H., Claessens, A., Stoll, M., Greenberg, E., Alexander, D., Runes, C., Henly, J.R. (2018). *Mapping Child Care Demand and the Supply of Care for Subsidized Families.* Washington DC: Urban Institute. Retrieved at https://www.urban.org/sites/default/files/publication/97286/mapping_child_care_demand_and_the_supply_of_care_for_subsidized_families.pdf.

[7] For more information about the American Community Survey, see https://www.census.gov/programs-surveys/acs/.

Use #2: Creating variables using spatial information

Researchers can use spatial information to develop location-based variables. Location-based variables may include driving distance between two points, proximity to topographical features like rivers and mountains, or neighborhood features as seen on satellite images of physical locations. By creating a new variable that considers location, researchers can answer questions with information that cannot be easily gathered from administrative data, other currently available data sources, or primary data collection. When researchers develop location-based variables, they can answer questions such as:

- What type of care is available along public transportation routes?
- What infrastructure characteristics (e.g., physical decay of buildings and sidewalks, availability of green space, litter on the street) can be used to categorize a child's neighborhood using aerial maps and street-level photographs?
- What is the availability of licensed family child care homes within two miles of businesses that require employees to work during night hours (or second shift)?

To create location-based variables that describe the distance between places, researchers first need to obtain the latitude and longitude for each place of interest (e.g., child care centers and public transit stops). Researchers might get this information from a dataset that includes physical addresses, which can then be geocoded (e.g., addresses of all public transit stops). They can also use GIS programs that have latitude and longitude for certain features, such as public transit stops, embedded in their maps.

Using the geocoded location variables, researchers can then use GIS programs or statistical software to create the distance variables. For instance, a researcher who is interested in households' access to ECE programs close to public transit can use a software program such as Stata or R to calculate the walking distance between each child care center and its closest public transit stop.

It is often desirable to summarize multiple location-based variables into a single indicator, and to combine distance information with other characteristics of the ECE provider. For instance, researchers can combine their geocoded location data with information about providers' quality rating and improvement system (QRIS) rating to calculate the average distance between high-quality ECE providers and the closest public transit stop.

Research spotlight: Creating a family-centered measure of ECE access in Minnesota[8]

Davis, Lee, and Sojourner (2018) were interested in developing a new measure of access that was centered on the family's home location and measured the distance to nearby providers. As a measure of demand for ECE, they simulated addresses of families with young children in Minnesota using combined data from the American Community Survey (ACS) 2011-2015 5-year estimates and the 2010 decennial census data. From the ACS, the researchers obtained recent estimates of the number of children in each block group.[9] From the 2010 census, the researchers obtained more detailed information about the *distribution* of households with young children within each block group, so they could more accurately simulate the locations of children.

As a measure of supply for ECE, they acquired addresses of Minnesota ECE providers in 2015 from the CCR&R database managed by Child Care Aware of Minnesota. To create a measure of families' access to ECE, the researchers geocoded Minnesota's ECE programs (licensed center and family child care providers, public preschool, and Head Start providers) and the simulated locations of households. Next, they calculated the driving times between the simulated locations of households and all ECE programs in Minnesota using a command in the statistical software Stata that incorporates road network data.[10] Finally, they calculated the number of child care slots within a 20-minute drive for each

[8] Davis, E. E., Lee. W. F., & Sojourner, A. (2018) Family-Centered Measures of Access to Early Care and Education. *Early Childhood Research Quarterly*. doi: 10.1016/j.ecresq.2018.08.001. An earlier version of this paper was developed as a discussion paper and can be accessed at https://www.iza.org/publications/dp/11396/family-centered-measures-of-access-to-early-care-and-education.

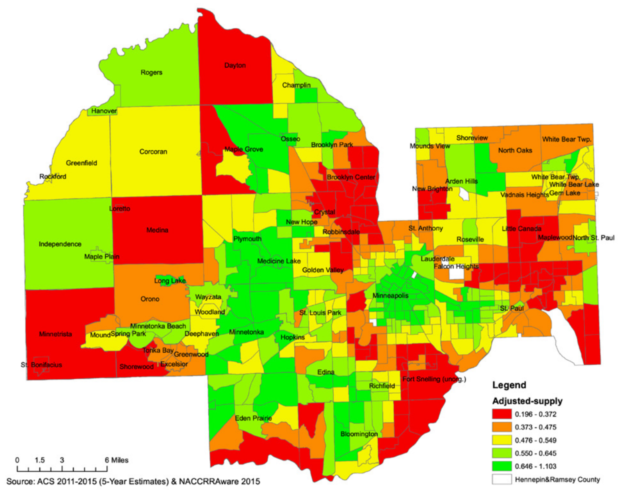
[9] Census block groups are geographical areas typically containing 600 to 3,000 people. https://www.census.gov/geo/reference/gtc/gtc_bg.html.

[10] Huber, S., & Rust, C. (2016). Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM). *Stata Journal, 16*(2), 416–423.

simulated household, giving more weight to providers closer to the family. This count was then used to create an adjusted measure of access by calculating the number of slots reasonably available for each child in Minnesota, after accounting for competition for slots from the number of children living near the provider.

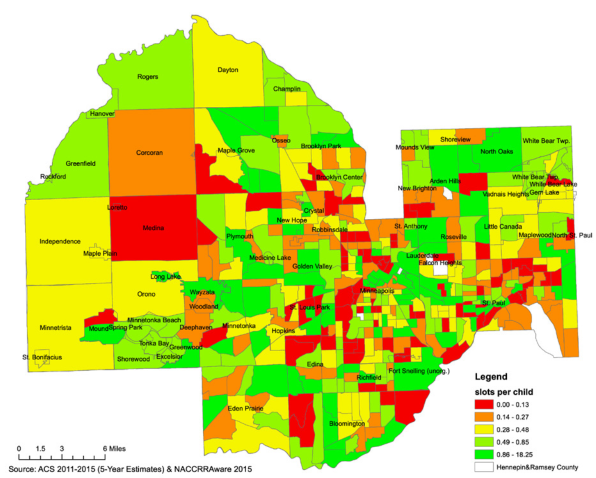
The researchers used this family-centered measure of ECE access to calculate the average number of slots per child in several Minnesota census tracts (Figure 3). Notably, this family-centered measure assumes that a family has access to any provider within a 20-mile drive, even if that provider is in a different census tract. For comparison, the researchers also calculated the average number of slots per child when a more common area-based measure of ECE access was used (Figure 4). With this area-based measure, families were assumed to have access only to providers within their home census tracts. Comparing Figures 3 and 4 reveals very little overlap between the two methods of defining access. This lack of overlap highlights the importance of defining variables used in spatial analysis: Different definitions produce different results.

Figure 3. Family-centered (distance-based) measure of access: Slots per child by census tract in Hennepin and Ramsey counties, 2015



Source: Davis, E. E., Lee. W. F., & Sojourner, A. (2018). Family-Centered Measures of Access to Early Care and Education. *Early Childhood Research Quarterly*. doi: 10.1016/j.ecresq.2018.08.001. An earlier version of this paper was developed as a discussion paper and can be accessed at https://www.iza.org/publications/dp/11396/family-centered-measures-of-access-to-early-care-and-education.

Figure 4. Area-based measure of access: Slots per child by census tract in Hennepin and Ramsey Counties, 2015



Source: Davis, E. E., Lee. W. F., & Sojourner, A. (2018). Family-Centered Measures of Access to Early Care and Education. *Early Childhood Research Quarterly*. doi: 10.1016/j.ecresq.2018.08.001. An earlier version of this paper was developed as a discussion paper and can be accessed at https://www.iza.org/publications/dp/11396/family-centered-measures-of-access-to-early-care-and-education.

The researchers concluded that policymakers and ECE entrepreneurs who are deciding where to invest in ECE programs may find it beneficial to examine maps of travel time-based measures of ECE access. These maps may provide a more accurate representation of areas with the highest need for additional ECE programs because they account for the fact that families often use ECE programs in ZIP codes and census tracts other than their own.

Use #3: Analyzing spatial patterns and associations

Researchers can also determine whether there are non-random patterns in how objects are distributed on a map, or whether location-based information can predict other details about that location. Unlike the visual inspection of maps to identify patterns, these analyses involve significance tests that can tell researchers the likelihood that patterns are not occurring by chance. ECE researchers who are interested in analyzing spatial patterns and associations can answer research questions such as:

- Does the presence of hospitals, casinos, and other businesses that require shift work predict greater availability of nonstandard-hours care options for subsidy-eligible families?
- Are counties with similar subsidy policies clustered together geographically?

Researchers can use a variety of spatial methods to analyze spatial patterns, such as spatial autocorrelation, spatial regression, or spatial interpolation.[11]

- **Spatial autocorrelation** quantifies how similar one object is to nearby objects. For instance, researchers can answer questions about whether counties with similar levels of access to ECE are near each other within a state. Researchers may find that there are clusters of counties that all have high access to ECE, and other clusters of counties that all have low access. ECE researchers may look for spatial autocorrelation when they are interested in spillover effects; for example, when the price of child care increases in one neighborhood, does the price also go up in nearby neighborhoods? Outlier counties can also be identified. For example, a county that has high access to ECE but is surrounded by counties with low access is an outlier geographically.
- **Spatial regressions** look for factors that can explain why certain spatial patterns exist.[12] For example, a researcher may notice that there are neighborhoods in a state that have many more family child care programs per household compared to other neighborhoods. Researchers can use spatial regression to ask whether family child care is more common in neighborhoods with many non-English-speaking families both in the neighborhood proper and also in the surrounding neighborhoods. That is, a spatial regression allows the dependent variable to be a characteristic of an individual neighborhood, region, or other geographic area, while the independent (explanatory) variables include characteristics of the specific geographic area as well as characteristics of nearby areas. Using a spatial regression rather than an ordinary least squares (OLS) regression is essential when there is autocorrelation in one or more of the independent variables. To continue with our example, neighborhoods with many non-English-speaking families may be clustered in just a few areas of the state. The presence of these clusters violates the OLS assumption that observations on the independent variable (i.e., the number of non-English speaking families in each neighborhood) are independent. Spatial regressions account for this autocorrelation by including information about the surrounding neighborhoods in the model.

  Spillover effects. A major advantage of spatial data is that it allows researchers to use information about surrounding areas to make predictions about a given area. In this way, spatial regression—like spatial autocorrelation—can test for spillover effects. For example, researchers found that traditional public school students performed better when charter schools opened nearby. The closer the charter schools were to the students' public school, the stronger the positive effect. These findings suggest a spillover effect from charter schools on the quality of traditional public school.[13]

---

[11] Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis, 38*(1), 5-22.doi: 10.1111/j.0016-7363.2005.00671.x.

[12] For more information about regression analysis basics, see the ESRI tutorial at http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/regression-analysis-basics.htm. Cordes, S. A. (2018). In pursuit of the common good: The spillover effects of charter schools in on public school students in New York City. *Education Finance and Policy, 13*(4), 484-512.

[13] Cordes, S. A. (2018). In pursuit of the common good: The spillover effects of charter schools in on public school students in New York City. *Education Finance and Policy, 13*(4), 484-512.

Regional variation in an association. In some cases, the association between a predictor and the outcome of interest varies by geographic region. For example, the number of non-English-speaking families may predict the presence of family child care programs in some neighborhoods, but not in others. Instead of using a basic OLS regression, which assumes that the association is the same in every neighborhood, researchers can use a technique called geographically weighted regression to allow the intercept and slope of a regression to vary by neighborhood.[14]

- **Spatial interpolation** allows researchers to estimate unknown information for one location using available information about nearby locations. When predicting unknown values, spatial interpolation assumes that the characteristics of nearby locations are better predictors than are characteristics of more distant locations. For example, ECE researchers might conduct surveys of parents in a subset of ZIP codes within a state to determine how far parents are willing to drive for child care. The researchers could then use interpolation methods to determine how far parents in the non-surveyed ZIP codes are willing to travel for child care.

Research spotlight: High-quality ECE in disadvantaged neighborhoods[15]

Hardy (2011) wanted to understand why some very low-opportunity neighborhoods (i.e. neighborhoods lacking educational, socioeconomic, and health opportunities for children) have high-quality ECE programs, even though high-quality ECE programs are more likely to be located in higher opportunity neighborhoods. Hardy designed an analysis to test two alternative hypotheses: First, very low-opportunity neighborhoods may have high-quality ECE when they are surrounded by high-opportunity neighborhoods; this might be due to the spillover of high-quality care from nearby high-opportunity neighborhoods. Alternatively, very low-opportunity neighborhoods may have high-quality ECE when they are surrounded by other low-opportunity neighborhoods; this could happen if organizations that seek to improve ECE quality are targeting regions with concentrated disadvantage.

To test these hypotheses, Hardy used spatial autocorrelation analyses. The data set consisted of neighborhoods (census tracts) in Massachusetts and Ohio.[16] Variables indicated whether a neighborhood had at least one high-quality ECE center, the level of opportunity in the neighborhood, and the level of opportunity in each surrounding neighborhood. A spatial autocorrelation analysis suggested that the majority of very low-opportunity neighborhoods with at least one high-quality ECE center were surrounded by other low-opportunity neighborhoods. This finding supported the second hypothesis: Neighborhoods located in areas of *concentrated disadvantage* were more likely to have higher quality care, compared to very low-opportunity neighborhoods that were close to high-opportunity neighborhoods. This could be because areas of concentrated disadvantage received more resources from nongovernmental organizations.

Hardy illustrated this finding with a map (Figure 5). In this map, the neighborhoods of interest (i.e., the very low-opportunity neighborhoods with at least one high-quality ECE provider) have green borders. Bright blue clusters represent opportunity “cold spots,” or areas of concentrated disadvantage; these are low-opportunity neighborhoods that are adjacent to another low-opportunity neighborhood. The map reveals that the green-bordered areas tend to be in these opportunity cold spots. It also illustrates that the vast majority of very low-opportunity neighborhoods with a high quality ECE provider are in areas of concentrated disadvantage.

In addition to the bright blue “cold spots,” the map in Figure 5 illustrates other spatial patterns that researchers and policymakers may want to further explore:

- White tracts represent neighborhoods that are not significantly similar to adjacent neighborhoods in terms of opportunity level.
- Bright red clusters represent opportunity “hot spots,” or clusters of neighborhoods where opportunity values are high.

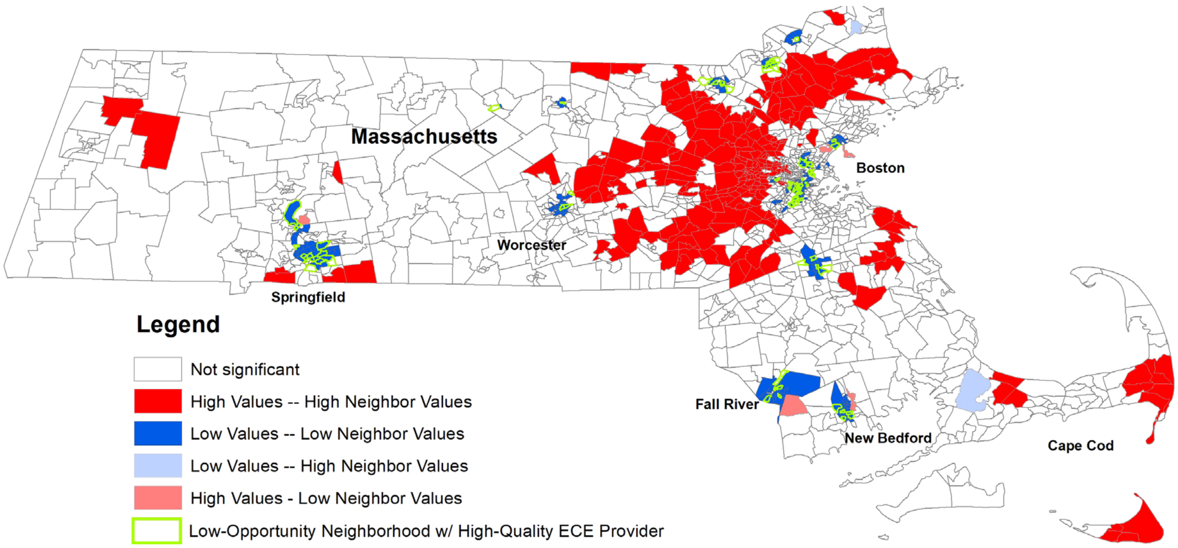
[14] http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/regression-analysis-basics.htm

[15] Hardy, E. H. (2011). "Equity in Early Opportunities to Learn: Examining the Roles of Place, Space, and Race - Exploratory Spatial Analysis of the Density of High-Quality Early Learning Centers, Neighborhood Opportunity and Racial Composition in Ohio and Massachusetts." *Cleveland Federal Reserve Bank Policy Summit*. Cleveland, OH.

[16] Neighborhoods (census tracts) are categorized as very low-opportunity neighborhoods using the Kirwan Institute Opportunity Index.

- Light blue areas represent low-opportunity neighborhoods that are surrounded by high-opportunity neighborhoods.
- Light red areas represent high-opportunity neighborhoods that are surrounded by low-opportunity neighborhoods.

Figure 5. Spatial patterns in neighborhood opportunity and presence of high-quality ECE providers in Massachusetts.



Source: Map 2 of Hardy, E. (2011). "Equity in Early Opportunities to Learn: Examining the Roles of Place, Space, and Race - Exploratory Spatial Analysis of the Density of High-Quality Early Learning Centers, Neighborhood Opportunity and Racial Composition in Ohio and Massachusetts." *Cleveland Federal Reserve Bank Policy Summit.* Cleveland, OH.

Challenges When Using Geography in ECE Research

Spatial analysis offers an exciting suite of tools for answering policy-relevant questions in ECE research; however, researchers may experience challenges when they begin to apply spatial analytic techniques. This section highlights some potential research challenges and strategies to mitigate them.

- **Availability and accuracy of data.** The data necessary to answer spatially based research questions may be difficult to access. Data may be housed in different agencies, the data may be incomplete (e.g., home addresses of children receiving subsidies are available, but not home addresses for low-income children not receiving subsidies), or the data may be inaccurate (e.g., the listed address for an ECE program may be its headquarters instead of the address where children are attending). Researchers can work closely with data owners, such as state agency staff, to understand the challenges of using specific data elements.[17]
- **Privacy of data.** Researchers must consider how to use and report personally identifiable information (PII). Because they may use information about the location of children to understand demand for ECE, researchers must be sensitive to whether an individual or family can be identified when displayed on maps or when geographically based data are shared. Many options can be used to maintain the privacy of families.

[17] For more information about how to access and use ECE administrative data, see Lin, V., Maxwell, K., & Forry, N. (2017). Determining the Feasibility of Using State Early Care and Education Administrative Data. OPRE Research Brief # 2017-17. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- Researchers could display aggregate data rather than individualized data (e.g., color-coded counties where families have low access to care rather than pinpoints for household addresses that have low access to care).
- They may need to suppress information when the number of individuals or households falls below a certain threshold determined by the researchers and state agencies.
- Researchers could also slightly modify, or "geographically mask," families' addresses rather than show actual addresses on a map. This strategy retains the important information, such as the number of families in a given area, while protecting families' privacy. A popular strategy for masking locations is to systematically add random perturbation to each address. For example, researchers can draw a circle around each address and randomly place a dot within that circle. That dot becomes the masked location that can be placed on maps for the public to view. To ensure that masked data prevent re-identification of households, the circle should be smaller in densely populated areas and larger in less dense areas.[18]

- **Interpretation of data.** When ECE researchers use spatially based variables, they should be thoughtful about how geographic boundaries are defined and cautious when interpreting the findings. For example, existing definitions of geographic areas, such as census tracts or zip codes, may not capture programs that are within a reasonable distance for a family. Many parents cross county, census block, and census block group boundaries to reach the nearest ECE provider.[19] It is important to account for these decision-making differences when interpreting the data.

Tips for Successful Use of Spatial Analysis in ECE Research

ECE researchers offered tips, based on their experiences, for how to successfully use spatial analysis in research.

- **Define the appropriate geographic area.** There are many options for defining a geographic area for spatial analysis, so it is important for researchers to choose the one that best meets the study needs. Researchers may consider a variety of factors when defining the appropriate geographical area, such as:
  - **Policy levers:** Geographic areas may be defined in order to match policies of interest. For example, a state agency may know that counties have different child care subsidy policies and may be interested in examining data by county.
  - **Research question:** Researchers may define the geographic area based on what can best answer the research question. For instance, researchers who are interested in families' access to ECE in terms of *reasonable distance* to providers likely will not want to use zip code as the geographical area. Instead, they could draw a boundary around each household that includes all points that can be reached within a 20-minute drive; all ECE programs within that boundary could be considered as potentially accessible programs. Researchers should not default to using readily available geographic definitions to define an area—such as using a county as a proxy for a neighborhood—without carefully considering the pros and cons of the approach.
  - **Accessibility:** Although a researcher-created boundary may best fit the research questions, researchers may still want to summarize data using defined areas, such as city limits or counties, so that findings can be understood by target audiences. For example, findings by legislative districts may be meaningful to legislators or policymakers, while findings by counties may be meaningful to the public.

[18] Zandbergen, P. A. (2014). Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in Medicine, 2014*, 1-14. doi: 10.1155/2014/567049.
[19] Davis, E. E., Lee. W. F., & Sojourner, A. (2018) Family-Centered Measures of Access to Early Care and Education. *Early Childhood Research Quarterly*. doi: 10.1016/j.ecresq.2018.08.001. An earlier version of this paper was developed as a discussion paper and can be accessed at https://www.iza.org/publications/dp/11396/family-centered-measures-of-access-to-early-care-and-education.

- **Use the appropriate data sets.** Researchers will need to seek out data sources that are compatible with spatial analyses. For example, data sets with information about ECE programs should include program addresses or another geographic locator, such as zip code. Data sets with information about children must have at least one geographic locator, such as children's addresses, zip codes, or school districts. Research teams must understand the pros and cons of each potential data source and develop a clear analysis plan prior to conducting analyses.[20]
    - **Level of data.** Additionally, when identifying the appropriate dataset, it may be useful for researchers to disaggregate the data at the lowest possible level; this affords the maximum flexibility when conducting spatial analysis. For example, a data set with children's addresses would permit a researcher to calculate demand for child care based on any desired boundary (e.g., school district, 20-mile radius around households). Often, however, child-level data are not available due to privacy concerns; this may be the case with some child care subsidy data. Researchers may then need to request counts of subsidized children at the zip code or county level.
    - **Reliability of data.** Another point to consider is that in population estimates from samples of populations, such as the ACS, uncertainty increases as the population of a geographic area gets smaller. For example, a researcher may wish to use counts of children at the census block group level, but this estimate will have more uncertainty in it than counts of children at a broader geographic level, such as the census tract or county.[21] To obtain population estimates for smaller geographic areas, researchers may need to use the ACS 5-year estimates data set, which combines data from five years of surveys.[22]
- **Consider how maps might be interpreted by individuals who are less familiar with the data.** Researchers should carefully consider how to share findings so that the map accurately communicates information to the reader. For example, a researcher who wants to show a county's availability of child care could indicate each ECE provider's location with a dot on a map. Some of these providers, however, may have just a few child care slots (e.g., family child care homes), whereas others may serve hundreds of children (e.g., center-based programs). The researcher could also include information about the *capacity* of providers on the maps to ensure that viewers can identify areas with limited child care.
- **Make findings accessible.** The visual aspect of spatial analysis makes it well suited to various dissemination approaches (e.g., through reports, online interactive maps, or social media posts). Decisions about the type of visuals to produce should be made in conjunction with agency staff who are involved in the project. When making these decisions, the team should consider the following:
    - **Target audience.** Different audiences may prefer a certain display over another. For example, some researchers may prefer data tables with numerical results from analyses that originated from spatial analyses. Some audiences may prefer a single image of a map, whereas others may prefer an online tool that lets them interact with the map and select which data to display.
    - **Costs of making findings accessible.** Research costs and plans will vary depending on these decisions. For example, developing an online tool, which must be hosted on a website and thus requires additional costs and skills, will be more expensive than creating a single image of a map.

[20] For more information about potential data sources and how to access these data, see Shaw, S.H., Lin, V., & Maxwell, K. L. (2018). Guidelines for Developing Data Sharing Agreements to Use State Administrative Data for Early Care and Education Research. OPRE Research Brief #2018-67. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.; Lin, V., Maxwell, K., & Forry, N. (2017). Considerations in Preparing to Analyze Administrative Data to Address Child Care and Early Education Research Questions. OPRE Research Brief # 2017-18. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
[21] Spielman, S. E., Folch, D., & Nagle, N. (2014). Patterns and causes of uncertainty in the American Community Survey. *Applied Geography*, 46, 147-157. Retrieved at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4232960/.
[22] Fuller, S. (2018). Using American Community Survey Estimates and Margins of Error. *U.S. Census Bureau*. Retrieved at https://www.census.gov/content/dam/Census/programs-surveys/acs/guidance/training-presentations/20180418_MOE.pdf.

- **Build a knowledgeable team.** ECE researchers who are new to using spatial analysis in research may benefit from including other experts on their team.
    - **Agency staff.** It is important for researchers to engage state agency staff to help shape the study, select the appropriate data, and interpret the findings.
    - **Geospatial experts.** ECE researchers who used location-based information for the first time acknowledged that having external expertise in spatial analysis was critical. A geographer could provide advice about efficient ways to calculate geographic variables, offer suggestions for maps that would best answer research questions, and run analyses and troubleshoot technical difficulties. While ECE researchers can interpret findings in the context of the ECE field, geographers or spatial analysts are able to guide researchers in analyzing spatial data appropriately. Researchers may find these experts at universities, at private geospatial companies, or they may work as independent contractors.
    - **Communications experts.** Experts, such as website developers or data visualization consultants, may be able to help researchers to maximize the impact and accessibility of maps. Because spatial data are often presented visually, the research team can consult with communications experts who can suggest color schemes that will be easily understandable in map form. Website developers can offer insight into how maps can be displayed in an online format. Data visualization consultants can offer suggestions for color schemes or patterns that may be best displayed on a map. For example, map users who have red-green colorblindness will have trouble reading maps that use red and green.

Ways to present spatial information visually

- Static maps are picture images of maps that remain unchanged. These maps are useful when the researcher wants to use a specific image to convey specific findings. A time lapse map is an example for engaging audiences with static maps. These maps show several static maps displayed together, each representing a different point in time to show change over time.
- Interactive online maps allow users to select one or more characteristics to show up on a map at any given time. These maps are useful when there are multiple characteristics that may be of interest to users, and the researcher wants to offer options to users. This is a space-efficient way to allow users to see each permutation of the map.
- Story maps are a hybrid between static and interactive maps. The user may be able to manipulate some options, but many aspects of the map remain static. This provides an interactive experience with some researcher control.
- Data tables may be used in lieu of a visual map to share numerical results that originated from spatial analyses. For example, a state agency may find a list of the 10 counties with the highest need for resources more helpful than a state map with the same information.

- **Delegate roles and responsibilities.** In developing the research plan, the team should identify the appropriate person to conduct the spatial analyses. ECE researchers have the content knowledge, but geographers and spatial analysts have the technical knowledge. The team must decide whether to build the capacity of research staff who have not conducted spatial analyses by seeking expert consultation and training, or by adding a specialist in spatial analysis to the team. It is important to note that building the capacity of researchers may take additional time.
- **Involve stakeholders.** Using spatial data in ECE research has implications for the individuals within the geographic locations of interest, so it is important to consider how stakeholders may be involved and how they can inform the research process. It may be useful for a few individuals in the locations of interest to advise the research team throughout the study. Researchers could also present findings to different stakeholder groups to understand their perspectives before

drawing conclusions. For instance, a map may indicate that families can potentially access child care opportunities by traveling across a bridge; however, conversations with stakeholders may reveal that the cost to use the bridge is a large deterrent for accessing care on the other side. Additionally, researchers may consider using community mapping, or public participatory geographic information systems (PPGIS), in which community members are invited to help collect data that can be used in analyses.[23] This approach builds stakeholder engagement and buy-in and increases the amount of data that can be used to answer policy-relevant questions. Researchers can also incorporate a mixed-methods approach by collecting qualitative data from key stakeholders to inform their quantitative analyses. Overall, maps are a powerful tool for engaging stakeholders, given their intuitive and visual appeal. Researchers will likely find the process of involving stakeholders to be extremely rewarding.

Conclusion

The field of spatial analysis is new and rapidly growing. This resource is a first step toward helping ECE researchers learn about the possibilities and potential for incorporating spatial analyses into their research. Additional work is needed to more fully understand the opportunities and limitations of spatial analysis and the capacity of this methodological approach to address policy-relevant questions in ECE.

Additional Resources

ECE-Related Resources

These resources and citations can be used as starting points for ECE researchers interested in learning more about how to integrate spatial analyses into ECE research.

- Hardy, E., Joshi, P., Ha, Y., & Schneider, K. G. (2018). Subsidized Child Care in Massachusetts: Exploring geography, access and equity. Waltham, MA: Institute for Child, Youth, and Family Policy, Brandeis University, Massachusetts Child Care Research Partnership and diversitydatakids.org. Retrieved at http://www.diversitydatakids.org/files/Library/Policy/GeoofSubsidizedCareFullReport.pdf.

  This report includes examples of spatial analyses in ECE and is an excellent resource for ECE researchers interested using spatial data to understand issues related to ECE access and supply.

- Cordes, S. A. (2018). In pursuit of the common good: The spillover effects of charter schools on public school students in New York City. *Education Finance and Policy, 13*(4), 484-512.

- Davis, E. E., Lee. W. F., & Sojourner, A. (2018). Family-Centered Measures of Access to Early Care and Education. *Early Childhood Research Quarterly*. doi: 10.1016/j.ecresq.2018.08.001. An earlier version of this paper was developed as a discussion paper and can be accessed at https://www.iza.org/publications/dp/11396/family-centered-measures-of-access-to-early-care-and-education.

- Hardy, E. H. (2011). "Equity in Early Opportunities to Learn: Examining the Roles of Place, Space, and Race - Exploratory Spatial Analysis of the Density of High-Quality Early Learning Centers, Neighborhood Opportunity and Racial Composition in Ohio and Massachusetts." *Cleveland Federal Reserve Bank Policy Summit.* Cleveland, OH.

[23] For more information about community mapping, see the National Community Mapping Institute at http://communitymappingforhealthequity.org/.

- Sandstrom, H., Claessens, A., Stoll, M., Greenberg, E., Alexander, D., Runes, C., Henly, J.R. (2018). *Mapping Child Care Demand and the Supply of Care for Subsidized Families.* Washington DC: Urban Institute. Retrieved at https://www.urban.org/sites/default/files/publication/97286/mapping_child_care_demand_and_the_supply_of_care_for_subsidized_families.pdf.

Other Helpful Resources

These resources offer maps that may be most relevant to ECE researchers or tutorials that may be most helpful to ECE researchers as they conduct spatial analyses for ECE research.

- PolicyMap. This service allows researchers to search for maps based on indicators (e.g., demographic information, quality of life, community factors) or by data source (e.g., Census, Head Start, or the Reinvestment Fund Study of Childcare Access) that can be used in spatial analyses.
- The Center for Applied Research and Engagement Systems (CARES) Engagement Network at the University of Missouri. This website allows you to search and access over 15,000 maps that can be customized for different uses. It also highlights other efforts that have used these maps to inform topics, such as community health, agriculture, or natural resource use.
- Spatial Econometrics: Methods and Models. This technical book addresses the role of spatial effects in econometrics.
- UCSB Center for Spatial Studies: This website provides links to research resources, such as sources for data that can be placed on a map, as well as links to educational resources.
- GIS Best Practices: Social Sciences: This resource describes the increasing use of mapping and spatial analyses in the social sciences. It also provides examples of projects within the social sciences that benefited from the creation of maps and spatial analyses.
- Spatial Analysis Tutorials from the University of Chicago: This website provides links to tutorials for various spatial analysis software. The extensive GeoDa software workbook is useful for anyone interested in learning more about spatial data, exploratory data analyses, map creation, and spatial analyses, regardless of their chosen software.
- Making Sense out of Spatial Data: This is a five-part video tutorial series provided by the Harvard Center for Geographic Analysis. Examples are included.
- ArcGIS Tutorials: These tutorials focus on the ArcGIS software platform but will likely be useful to anyone interested in learning about specific elements of spatial analysis.

Project Officers:

Ivelisse Martinez-Beck, PhD, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Kathleen Dwyer, PhD, Project Officer
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP23320095631WC

Project Director:
Kelly Maxwell
Child Trends
7315 Wisconsin Avenue
Suite 1200 West
Bethesda, MD 20814

Lin, V. & Madill, R. (2019). *Incorporating Spatial Analyses into Early Care and Education Research*. OPRE Research Brief #2019-88. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Acknowledgments:

The authors gratefully acknowledge the guidance and support from Kathleen Dwyer, Jenessa Malin, and Danielle Berman at the Office of Planning, Research, and Evaluation. We appreciate the contributions from researchers Elizabeth Davis, Erin Hardy, and Heather Sandstrom, whose research studies and insights are highlighted. Finally, we appreciate the support of the Child Trends' communications team for their assistance in the preparation of this product.