

Potentiale geographischer Daten für die Sozialberichterstattung im Rahmen der Familienforschung

Daniela Gesell, Arne Bethmann, Christian Alt

1 Allgemeine Einleitung

Eine moderne Sozialberichterstattung muss den Anspruch haben, diejenigen Faktoren zum Gegenstand einer nachhaltigen Forschung zu machen, die die Lebensbedingungen und die Lebensverhältnisse der Individuen beeinflussen. Bislang blieb bei der Berichterstattung über familiales Leben vielfach außer Acht, dass sich die Veränderung wie z.B. der ökonomischen Struktur, der demografischen Zusammensetzung und der sozialräumlichen Entwicklung vor allem vor Ort, d.h. im Nahumfeld der Familien vollziehen. Diese Entwicklungen auf lokaler Ebene setzen jene Rahmenbedingungen, die familiales Leben beeinflussen. Dies zeigt sich unter anderem an den verschiedenen Teilhabechancen in den Bereichen Erwerbsarbeit, Bildung oder Betreuung. Wenn Sozialberichterstattung dazu einen Beitrag liefern will, muss sie einen genaueren Blick auf räumlich unterschiedliche Lebensbedingungen werfen, um die Basis für die Gestaltung einer passgenauen Familien- und Sozialpolitik zu schaffen. Dies verdeutlichen auch die Bemühungen des „Audit zu familiengerechten Kommunen“ (vgl. *Strohmeier et al. 2011*).

Eine besonders interessante Fragestellung im Kontext einer so gearteten Sozialberichterstattung ist die Inanspruchnahme von öffentlich geförderten Betreuungsangeboten und die damit, nach der Vereinbarkeit von Familie und Beruf, einhergehenden Bedingungen für die Erwerbstätigkeit der Mütter. Dabei zeigt sich nicht nur in Deutschland, dass die Vereinbarkeitsproblematik stark bestimmt ist von räumlichen Rahmenbedingungen und zum Teil erheblich differiert (vgl. *Berghammer/Verwiebe 2015; Keller/Haustein 2014; Knittel et al. 2014; Rüling/Kassner 2007; Wengler et al. 2008*). Bislang wurde in der Wissenschaft weitestgehend darauf verzichtet, die Analysen auf kleinräumigen Maßstäben, z.B. Gemeinden durchzuführen, stattdessen hat man sich mit Indikatoren begnügt, die sich auf hohem Aggregationsniveau bewegen (Bundesland oder Kreis). Mittlerweile liegen aber eine Vielzahl von Informationen und Daten vor, die hier Abhilfe schaffen können. So ergeben sich, durch die Kombination von individuellen Befragungsdaten, amtlich erhobenen Daten und POIs – „Points of interest“ (z.B. Schulen, Krankenhäuser oder Kindertageseinrichtungen) neue Chancen und Möglichkeiten der Berichterstattung, die den Anspruch an eine moderne Sozialberichterstattung erfüllen können. Im Folgenden soll gezeigt werden, um welche Daten es sich dabei handelt und wie man diese einsetzen kann, um mit der Kombination aus Befragungsdaten und amtlichen Daten zu neuen Er-

kenntnissen zu gelangen. Die Befragungsdaten stammen aus der KiföG-Länderstudie 2015, die sich mit den von Eltern geäußerten Bedarfen an öffentlich geförderten Betreuungsangeboten für ihre Kinder beschäftigt. Der Fokus der Studie lag auf der Untersuchung der elterlichen Bedarfe für Kinder unter drei Jahren. Insgesamt wurden 12.661 Eltern in allen Bundesländern befragt (vgl. Aust et al. 2015).

2 Amtliche Datenquellen und ihre Nutzung

Eine erste Datenquelle bietet das *Bundesinstitut für Bau, Stadt- und Raumforschung* (BBSR), welches bundesweite Analysen zur Raumentwicklung durchführt und Prognosen zur künftigen Raumentwicklung erstellt. Dabei werden die regionalen Lebensbedingungen in Deutschland auf der Ebene von Gemeinden, Kreisen und Raumordnungsregionen analysiert und bewertet. Die aufbereiteten und mittels Metadaten beschriebenen Basisdaten sind die Grundlage für die Berechnung von Indikatoren. Die Datenbank (INKAR) enthält fast 600 Indikatoren regionalstatistischer Daten zu fast allen gesellschaftlich bedeutsamen Themenfeldern. Die räumliche Bezugsebene in Deutschland kann nach administrativen (Länder, Kreise, Gemeindeverbände) und nicht-administrativen (Raumordnungsregionen, IHK-Bezirke, Siedlungsstrukturtypen) Raumbezügen ausgewertet werden.

Neben dem BBSR stellt auch die amtliche Statistik Zahlen zur gesamten Bevölkerung unter anderem auf Basis repräsentativer Erhebungen, wie dem Mikrozensus oder Zensus 2011, zur Verfügung. Die Daten aus diesen Quellen reichen in der Regel nur bis auf die räumliche Bezugsebene der Kreise und kreisfreien Städte herunter.

Einen deutlich höheren Informationsgehalt, mit Blick auf die Nahumwelten der Familien, bieten „Points of Interest“ (POI). Dabei handelt es sich um die genauen Adressen oder geographischen Koordinaten von thematischen Orten (z.B. Krankenhäuser oder Restaurants). POIs können auch für sozialwissenschaftliche Zwecke Verwendung finden, betrachtet man die räumliche Verteilung von Objekten wie Kindertageseinrichtungen, Schulen oder Jugendzentren. Grundlage für die Untersuchung von räumlichen Disparitäten und Verteilungen sind zunächst administrative Regionaleinheiten. Kreise und kreisfreien Städte sind dabei häufig die kleinste verwendete Ebene, auf die sich auch ein Großteil des amtlichen Datenangebots bezieht.

Die gewonnenen Informationen zu den verschiedenen Themenbereichen können auf Grundlage verschiedener räumlicher Aggregationsniveaus in einem Geoinformationssystem (GIS) weiter bearbeitet werden. Sie können in Form von verschiedenen einzelnen Kartenebenen (Layer) neu zusammen gestellt werden, um die zu bearbeitende Fragestellung visuell anhand von diversen Indikatoren zu erklären. Indikatoren zur Beschreibung von Regionen sind zum Beispiel das Kinderbetreuungsangebot, die vorherrschende Altersstruktur, die Erreichbarkeit von wichtigen Infrastrukturen, die Ausbildungschancen von Jugendlichen oder auch die Erwerbssituation in den betrachteten Räumen. Um Veränderungen der Raumstrukturen beschreiben zu können, werden auch Indikatoren wie die Bevölkerungsentwicklung, Wanderungssalden oder Zuwanderungsraten verwendet. Dabei kann zusätzlich auf thematische Karten, die als Basisdaten oder Sachdaten standardmäßig vorliegen, zurückgegriffen werden (vgl. Lange 2013). Üblicherweise bilden diese Layer die Informationen auf Bundesebene oder auf Kreisebene ab. Ein solches Verfahren soll

Unterschiede zwischen den Räumen hinsichtlich ihrer Entwicklung, den Rahmenbedingungen und den vorherrschenden Strukturen kenntlich machen.

3 Zusammenführung der Daten

Die zuvor beschriebenen Daten über Räume und deren verschiedene Indikatoren sollen mit den Daten der Befragten kombiniert werden. So ist es möglich, regionale objektive Indikatoren mit den subjektiven Informationen der befragten Eltern gemeinsam auszuwerten. Die Individualdaten der KiföG-Länderstudie 2015 lassen sich dabei auf der Kreisebene, anhand der Kreiskennziffer, mit den amtlichen Daten und den dort vorliegenden Indikatoren verknüpfen. Eine kleinräumigere Verknüpfung auf der Gemeindeebene, ist technisch möglich. Diese kann aber aufgrund von datenschutzrechtlichen Bestimmungen nur vom Erhebungsinstitut infas durchgeführt werden.

Die Fragestellung beinhaltet die Darstellung des tatsächlichen räumlichen Betreuungsangebots in Deutschland, im Vergleich zur räumlichen Verteilung der Bevölkerung. Ziel der regionalen Analyse war zunächst die Identifizierung von Gebieten, die eine räumliche Unterversorgung an Kindertagesstätten aufweisen. Zunächst wurden die Daten der KiföG-Länderstudie 2015 im räumlichen Kontext betrachtet. Dafür wurden die einzelnen, als Samplepoints der Befragung gezogenen Gemeinden auf Kreisebene aggregiert. Die Basis bildet ein mehrstufiges Stichprobenverfahren aus den Einwohnermelderegistern. In jedem Bundesland sollte insgesamt aus 30 Points gezogen werden (vgl. Aust et al. 2015). Um den Datenschutz bei dieser Untersuchung zu gewährleisten, wurden nur Kreise in die Analyse mit einbezogen, in denen mehr als zehn Befragte leben. Durch die Kombination dieser Einschränkung und der Art der Stichprobenziehung ergibt sich ein eher lückenhaftes räumliches Bild, da zu vielen Kreisen keine Aussage getroffen werden kann¹.

Um die räumliche Verteilung aller Kindertageseinrichtungen darzustellen, wurden die von einem Geomarketingunternehmen gesammelten Adressen der Einrichtungen geokodiert und einem GIS zugespielt. Zur Klärung der Fragestellung, wie sich die räumliche Ausstattung an Kindertageseinrichtungen in der Fläche verteilt, wurde auf Kreisebene eine Dichteverteilung (Anzahl der Einrichtungen pro km²) berechnet. Unberücksichtigt bleiben dabei die Anzahl der angebotenen Plätze pro Einrichtung sowie die tatsächliche Zahl der betreuten Kinder. Zusätzlich wurde anschließend die fußläufige Dauer der Wegstrecke der befragten Eltern von unter dreijährigen Kindern zur nächstgelegenen Kindertageseinrichtung berechnet. Dabei konnte aus erhebungstechnischen Gründen leider nicht berücksichtigt werden, ob diese Einrichtung auch tatsächlich von dem Kind besucht wird. Der Fokus lag somit lediglich darauf, wie weit die Befragten von der nächsten Einrichtung entfernt leben. Um den Datenschutz der Befragten zu gewährleisten, musste die Zuweisung der Distanzen vom Erhebungsinstitut infas durchgeführt werden. Die visuelle Darstellung in Kartenform erfolgte wieder auf aggregierter Kreisebene. Exemplarisch kann dadurch gezeigt werden, dass in den ostdeutschen Bundesländern die Betreuungsquote und der Anteil der Kinder, die einen langen Gehweg zur nächsten Kindertageseinrichtung haben (über 20 Minuten), höher ist als im restlichen Bundesgebiet. Der mittlere zurückzulegende Weg über die gesamte Stichprobe beträgt weniger als 20 Minuten. Die Analyse zeigt, dass es in vielen Teilen Deutschlands ein, räumlich gesehen, geringes An-

gebot an Betreuungseinrichtungen gibt, allerdings kann noch keine Aussage über die tatsächlich angebotenen Plätze gemacht werden.

Insgesamt geben die Auswertungen bereits einen Einblick in räumliche Verteilungsmuster zum Thema Kinderbetreuung. Jedoch finden die Analysen nach wie vor auf Kreisebene statt. Ziel ist es, die Auswertungen ohne administrative Grenzen durchzuführen, da sich räumliche und individuelle Muster nicht zwangsläufig an vordefinierten räumlichen Einheiten orientieren müssen.

4 Möglichkeiten für kleinräumige Analysen

Die Verfügbarkeit von geokodierten Daten eröffnet neue Möglichkeiten der kleinräumigen Analyse. Mussten bisher zwangsläufig Aggregate gebildet werden um bspw. die Daten der amtlichen Statistik räumlich auszuwerten, bietet die exakte räumliche Verortung über Geokoordinaten die Möglichkeit genaue Distanzen zwischen interessierenden Analyseeinheiten zu bilden. Daraus ergeben sich zwei erhebliche Vorteile gegenüber bisherigen räumlichen Analysestrategien.

Zum Einen kann die räumliche Analyseebene prinzipiell bis auf den Einzelfall und das individuelle, direkte räumliche Umfeld herunter gebrochen werden um die tatsächlich handlungsrelevanten Strukturen, z.B. die familienbezogenen Angebotsstrukturen, in den Blick zu nehmen. Gerade Hypothesen mit handlungstheoretischem Bezug können so deutlich valider getestet werden. Zum Anderen wird auch das Analysepotential für regionale Aggregatinformationen verbessert indem die Notwendigkeit entfällt sich an administrativen oder anderen nicht auf die spezifische Fragestellung bezogenen Regionalabgrenzungen zu orientieren. Dadurch entfallen einerseits Unschärfen, die entstehen können wenn etwa die handlungsrelevanten, lokalen Strukturen faktisch Kreisgrenzen überschreiten (z.B. Nutzung von Kinderbetreuungseinrichtungen im benachbarten Kreis). Dies hat, insbesondere hinsichtlich der zum Teil stark variierenden Flächenunterschiede Problempotential. Andererseits lassen sich so auch Verzerrungen vermeiden die entstehen können, wenn regionale Abgrenzungen, wie im Falle administrativer Raumeinheiten, mit politischen Prozessen zusammenhängen die wiederum die lokalen, familienrelevanten Strukturen mitbestimmen. Auch wenn die politischen Prozesse regelmäßig von großem Interesse sind, sollten sie nicht über die räumlichen Abgrenzungen in die Analysen eingehen, da dann nicht mehr zwischen politischen Einflussfaktoren und dem tatsächlichen Einfluss kleinräumiger Strukturen unterschieden werden kann.

Konzeptionell sind die Vorteile von Analysen auf Basis geokodierter Daten relativ klar zu benennen. In der praktischen Umsetzung ergeben sich allerdings einige zusätzliche Herausforderungen. Zuerst ist hier sicherlich der Datenzugang zu benennen. Zwar hat die Verfügbarkeit von geokodierten Informationen in den letzten Jahren erheblich zugenommen, allerdings heißt das natürlich nicht, dass auch die z.B. POIs verfügbar sind, die für familienbezogene Fragestellungen relevant sind. Eine offizielle Quelle ist das BKG, dessen Angebot sich zu familienrelevanten Themenbereichen zurzeit allerdings noch im Aufbau befindet. Alternativ bieten sich im Internet noch zahlreiche andere Quellen. Neben themenspezifischen Portalen, etwa zur Kinderbetreuung lässt sich natürlich auch auf sehr allgemeine Quellen wie OpenStreetMap zurückgreifen. Darüber hinaus gibt es inzwischen diverse kommerzielle Anbieter, die POIs zu Zwecken des Geomarketings ent-

geltlich zu Verfügung stellen. Bislang ist das Angebot noch verhältnismäßig unübersichtlich, so dass die Datenakquise für konkrete Fragestellungen mit einigem Rechercheaufwand verbunden ist.

Eine weitere Herausforderung ist die Sicherstellung einer hinreichenden Datenqualität. Auf Seiten der Befragungsdaten kann für die Einschätzung der Qualität auf etablierte Durchführungs- und Dokumentationsstandards zurückgegriffen werden, insofern diese im Erhebungsprozess eingehalten wurden. Hingegen existieren übergreifende Standards auf Seiten der sehr heterogenen Datenquellen für POIs bisher nicht. Eine Qualitätseinschätzung gestaltet sich daher oft sehr schwierig, besonders weil die Daten meist nicht explizit zu Forschungszwecken erhoben wurden und über den Prozess der Datengenese oft nur sehr wenige Informationen vorliegen. Bisweilen fällt es bereits schwer den Zeitbezug der Daten festzustellen, an dem sie also erhoben wurden, um einzuschätzen ob sie für den Analysezeitpunkt gültig sind.

In jedem Fall stellt sich weiterhin die Frage des Datenschutzes. Neben der Berücksichtigung der üblichen Vorgaben im Bezug auf die Befragungsdaten, gilt es auch zu prüfen inwiefern die POIs unter Datenschutzgesichtspunkten problemlos für Forschungszwecke verwendet werden können. Im Gegensatz zur Verwendung einzelner Datenquellen wird diese Frage dann komplizierter, wenn Befragungsdaten und Geoinformationen miteinander verknüpft werden sollen. Geschieht das auf der individuellen Ebene, z.B. über die Berechnung von Distanzen zwischen den Wohnkoordinaten der befragten Person und den POIs kann unter Umständen schon anhand weniger Distanzen, bei Kenntnis aller POI-Koordinaten, auf den exakten Wohnort des Individuums zurückgeschlossen werden. Dies birgt daher ein großes Deanonymisierungspotential und dürfte regelmäßig mit Datenschutzvorgaben kollidieren.

Ist der Datenzugang geklärt und die Qualität der Daten hinreichend gewährleistet, besteht eine Möglichkeit die potentiellen Vorteile der präzisen Geoinformationen zu nutzen und gleichzeitig das Risiko der Deanonymisierung möglichst gering zu halten darin, eine Flächenrasterung der POIs durchzuführen. Das bedeutet, dass die gesamte relevante Fläche, wie etwa die Bundesrepublik Deutschland, in ein einheitliches Raster eingeteilt wird, z.B. Quadrate von 500 Meter Kantenlänge. In jeder dieser Rasterzellen kann dann die Anzahl von POIs gezählt werden (vgl. *Kirchner et al. 2014*)². Die Zuspiegelung zu den Befragungsdaten geschieht nun nicht mehr auf der Ebene individueller Distanzen, sondern nur noch für Rasterzellen. Je nachdem wie kleinräumig das Raster gewählt wird, können so trotz geringerem Deanonymisierungspotential noch Analysen des lokalen Kontexts durchgeführt werden. Durch die arbiträre Festlegung des Rasters ist außerdem ausgeschlossen, dass andere Einflüsse, wie etwa politische Prozesse über die Definition der Regionalabgrenzungen die Ergebnisse verzerren. Lediglich das Problem der Unschärfe an den Grenzen kann so nicht gänzlich gelöst werden, wird aber geringer je kleiner die Rasterzellen gewählt werden. Dieses Vorgehen kann mit aktuellen GIS-Programmen relativ bequem für eigene Daten durchgeführt werden. Darüber hinaus stellt auch die amtliche Statistik (z.B. *destatis*) inzwischen regionale statistische Informationen auf der Basis solcher Raster zur Verfügung.

Geoinformationen bieten für zukünftige, kleinräumige Analysen in der Sozialberichterstattung einen deutlichen Mehrwert. Zum einen haben sie das Potential die familialen Nahumwelten detailliert zu beschreiben. Dabei kann mit den beschriebenen Verfahren auf die bisherige Verwendung administrativer Raumeinheiten verzichtet werden. Die Analyseebene kann dabei grundsätzlich bis auf individuelle Distanzen runter gebrochen werden,

wobei Datenschutz, Datenqualität und Datenverfügbarkeit hierbei die begrenzenden Faktoren sind. Zum anderen führt diese Einschränkung aber auch zu der Erkenntnis, dass für manche Auswertungen mittelfristig auch weiterhin auf amtliche regional aggregierte Daten zurückgegriffen werden muss. Auch hier besteht Potential für die Verbesserung kleinräumiger Analysen durch moderne statistische Verfahren wie Small-Area-Estimation.

Anmerkung

- 1 Um diesem Umstand entgegenzuwirken können u.a. Methoden der Small – Area – Statistik angewendet werden. Durch die Anpassung bereits vorhandener Schätzmethoden können die „Lücken“, welche durch die Erhebung entstehen verringert werden (vgl. Münnich et al. 2013). Derartige Verfahren kamen in den hier vorliegenden Analysen nicht zur Anwendung, sind aber für die Zukunft geplant.
- 2 Alternativ gibt es für die Berechnung der Rasterwerte auch deutlich elaboriertere statistische Verfahren, wie z.B. die Verwendung von Kerndichteschätzern (vgl. Schräpler, 2011).

Literatur

- Aust, F./Burg, J. von der/Hess, D./Knecht, B. (2015): Methodenbericht. KiföG-Regionalstudie Winter 2014/2015 und Zusatzuntersuchung: „Kinder in öffentlicher Tagesbetreuung“. Hg. v. infas Institut für angewandte Sozialwissenschaft GmbH. – Bonn.
- Berghammer, C./Verwiebe, R. (2015): Die Verbreitung des Doppelernährer-/Doppelbetreuermodells in fünf Ländern Europas. WSI-Mitteilungen (1), S. 116-124.
- Keller, M./Haustein, T. (2014): Vereinbarkeit von Familie und Beruf. Ergebnisse des Mikrozensus 2013. Wirtschaft und Statistik (Dezember), S. 733-753.
- Kirchner, T./Pflanz, F./Techen, A./Wagenknecht, L. (2014): Kleinräumige Gliederung, Georeferenzierung und Rasterdarstellung im Zensus. Zeitschrift für amtliche Statistik Berlin-Brandenburg 3, S. 28.
- Knittel, T./Henkel, M./Krämer, L./Lopp, R./Schein, C. (2014): Dossier Müttererwerbstätigkeit: Erwerbstätigkeit, Erwerbsumfang und Erwerbsvolumen 2012. Prognos AG, Berlin.
- Lange, N. de (2013): Geoinformatik. Theorie und Praxis. – Springer-Verlag.
<https://doi.org/10.1007/978-3-642-34807-5>
- Münnich, R./Burgard, J.P./Vogt, M. (2013): Small area-statistik: methoden und anwendungen. AStA Wirtschafts-und Sozialstatistisches Archiv 6 (3-4), S. 149-191.
<https://doi.org/10.1007/s11943-013-0126-1>
- Rüling, A./Kassner, K. (2007): Familienpolitik aus der Gleichstellungsperspektive. Ein europäischer Vergleich, Friedrich Ebert Stiftung. – Berlin.
- Schräpler, J. P. (2011). Konstruktion von SGB II–Dichten als Raumindikator und ihre Verwendung als Indikator im Rahmen der Sozialberichterstattung am Beispiel der „sozialen Belastung“ von Schulstandorten in NRW – ein Kerndichte-Ansatz. AStA Wirtschafts-und Sozialstatistisches Archiv 5 (2), S. 97-124.
- Strohmeier, K.P./Schultz, A./Lersch, P. (2011): Familiengerechte Kommunen–Gemeindetypisierung familienrelevanter Rahmenbedingungen in nordrheinwestfälischen Kommunen. Audit familiengerechte Kommunen in Nordrhein-Westfalen. – Bochum.
- Wengler, A./Trappe, H./Schmitt, C. (2008): Partnerschaftliche Arbeitsteilung und Elternschaft: Analysen zur Aufteilung von Hausarbeit und Elternaufgaben auf Basis des generations and gender survey.