ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Saidani, Younes; Bohnensteffen, Sarah; Hadam, Sandra

**Article**

# Qualität von Mobilfunkdaten – Projekterfahrungen und Anwendungsfälle aus der amtlichen Statistik

WISTA - Wirtschaft und Statistik

**Provided in Cooperation with:**
Statistisches Bundesamt (Destatis), Wiesbaden

*Suggested Citation:* Saidani, Younes; Bohnensteffen, Sarah; Hadam, Sandra (2022) : Qualität von Mobilfunkdaten – Projekterfahrungen und Anwendungsfälle aus der amtlichen Statistik, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 74, Iss. 5, pp. 55-67

This Version is available at:
https://hdl.handle.net/10419/265346

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

# QUALITÄT VON MOBILFUNKDATEN – PROJEKTERFAHRUNGEN UND ANWENDUNGSFÄLLE AUS DER AMTLICHEN STATISTIK

Younes Saidani, Sarah Bohnensteffen, Sandra Hadam

**Schlüsselwörter:** Mobilfunkdaten – Qualität – Neue Digitale Daten – Bevölkerungsstatistik – Mobilität

## ZUSAMMENFASSUNG

Mobilfunkdaten gelten als vielversprechende Datenquelle, ihre Qualität im Hinblick auf den Einsatz in der amtlichen Statistik ist jedoch bisher nicht systematisch untersucht worden. Dieser Artikel bietet einen Überblick über die bislang durchgeführten Projekte mit Mobilfunkdaten im Statistischen Bundesamt und beschreibt anhand des Entstehungs- und Aufbereitungsprozesses der Daten deren wichtigste Qualitätsaspekte. Auf Basis der bisherigen Erfahrungen ist festzustellen, dass die Datenquelle bestens für kleinräumige Sonderauswertungen und Auswertungen am aktuellen Rand geeignet ist, die mit vorhandenen amtlichen Daten nicht realisiert werden können. Eine Statistikproduktion unter Berücksichtigung amtlicher Qualitätskriterien ist derzeit jedoch nicht möglich. Ein wichtiger Grund hierfür ist der bislang fehlende rechtliche Zugang zu Daten aller Anbieter von Mobilfunkdaten am deutschen Markt.

***Keywords:*** *mobile network data – quality – new digital data – population statistics – mobility*

## *ABSTRACT*

*Mobile network data are considered a promising new data source, but its quality for use in official statistics has hitherto not been systematically studied. This article provides an overview of the projects using mobile network data which have so far been carried out at the Federal Statistical Office and, based on the generation and processing of the data, describes the major quality aspects. On the basis of this experience, the following conclusion is reached: the data source is well suited for special evaluations of small-area data and for evaluations of very recent data that cannot be made using the existing official data. However, the production of statistics in line with the official quality criteria is currently not possible. A key reason is the lack of legislation permitting access to data of all providers of mobile network data in the German market.*

**Younes Saidani**

hat Volkswirtschaftslehre und Philosophie, Politik und Ökonomie an der Universität Oxford (Vereinigtes Königreich) studiert. Im Referat „Erforschung neuer digitaler Daten“ des Statistischen Bundesamtes entwickelte er seit Ende 2020 Produkte auf Basis von Mobilfunkdaten. Als Data Scientist im Referat „Künstliche Intelligenz, Big Data“ befasst er sich aktuell mit der Qualitätsbeurteilung von maschinellen Lernmodellen.

**Sarah Bohnensteffen**

hat Soziologie und amtliche Statistik an den Universitäten Heidelberg und Complutense de Madrid (Spanien) studiert. Beim Statistischen Bundesamt befasste sie sich im Referat „Erforschung neuer digitaler Daten“ zunächst mit Mobilitätsanalysen auf Basis von Mobilfunkdaten und forscht derzeit im Rahmen des MikroSim-Projekts im Referat „Forschungsdatenzentrum, Methoden der Datenanalyse“ zur Simulation von Wohnraum.

**Sandra Hadam**

hat Volkswirtschaftslehre an der Freien Universität Berlin studiert. Als Wissenschaftliche Mitarbeiterin im Referat „Erforschung neuer digitaler Daten“ des Statistischen Bundesamtes und Promovierende an der Freien Universität Berlin befasst sie sich seit Ende 2017 mit den Nutzungsmöglichkeiten von Mobilfunkdaten in der amtlichen Statistik.

## 1

## Einleitung

Mobilfunkdaten bieten aufgrund ihrer zeitlich und räumlich hohen Auflösung eine potenziell hochinteressante Datenquelle zur Unterstützung verschiedener amtlicher Statistiken wie der Bevölkerungs- und Tourismusstatistik (Grassini/Dugheri, 2021) und versprechen vielfältiges Analysepotenzial im Zusammenhang mit der Bevölkerungsmobilität. In Deutschland teilen sich die drei Mobilfunknetzbetreiber (englisch: mobile network operator) Deutsche Telekom, Telefónica Deutschland und Vodafone mit jeweils etwa einem Drittel Marktanteil den deutschen Mobilfunkmarkt im aktuellen 4G-Netz; im vierten Quartal 2021 waren in Deutschland rund 161 Millionen SIM-Karten registriert (Bundesnetzagentur, 2022).

Das Statistische Bundesamt erforscht seit Ende 2017 die Datenquelle Mobilfunkdaten hinsichtlich ihrer Nutzungsmöglichkeiten für die amtlichen Statistik. Damit kommt es seiner Aufgabe nach § 3 Absatz 1 Nummer 1 Bundesstatistikgesetz nach, Bundesstatistiken methodisch und technisch im Benehmen mit den Statistischen Ämtern der Länder vorzubereiten und weiterzuentwickeln. Seitdem wurden Kooperationen mit mehreren Datenanbietern etabliert, die Daten aus dem Netz der Deutschen Telekom und der Telefónica Deutschland aufbereiten und gemäß eines mit dem Bundesbeauftragten für Datenschutz und Informationssicherheit vereinbarten Konzepts anonymisieren. Auf Grundlage dieser anonymisierten und aggregierten Daten wurden mit verschiedenen Machbarkeitsstudien mögliche Anwendungsfelder für die Integration dieser Datenquelle in die amtliche Statistik ausgearbeitet.

Dieser Artikel bietet erstmals eine systematische qualitative Untersuchung der Datenquelle für den Einsatz in der amtlichen Statistik. Da Mobilfunkdaten das Ergebnis komplexer technischer Entstehungs- und Aufbereitungsprozesse sind, ist ein tiefergehendes Verständnis dieser Prozesse für eine Qualitätsbewertung unabdingbar. Kapitel 2 dieses Beitrags erläutert daher die Genese und die Verarbeitung der Daten und leitet daraus relevante Qualitätsaspekte ab. In Kapitel 3 werden Qualitätsprobleme anhand von Projekterfahrungen mit Mobilfunkdaten im Statistischen Bundesamt illustriert und beispielhaft Nutzungs- und Analysepotenziale herausgearbeitet. Aufbauend auf diese Erkenntnisse beschreibt Kapitel 4 Nutzungsmöglichkeiten von Mobilfunkdaten, anhand derer künftige Projektideen gemessen werden können. Ein Fazit zu den aktuellen Einsatzmöglichkeiten von Mobilfunkdaten in der amtlichen Statistikproduktion schließt den Artikel.

## 2

## Datengenese und Qualitätsaspekte

Die folgenden Abschnitte geben einen Überblick über die Schritte, die Mobilfunkdaten von ihrer Entstehung über die Aufbereitung bis zur finalen Nutzung durchlaufen. Zugleich gehen sie darauf ein, welche Implikationen die einzelnen Schritte auf die Qualität der Mobilfunkdaten haben. Eine Zusammenfassung dieser Qualitätsaspekte bietet ↘ **Übersicht 1**.

### Entstehung von Signaldaten

Wenn Mobilfunkgeräte das Mobilfunknetz nutzen, so geschieht dies durch das Übertragen von Signalen zwischen der SIM-Karte des Mobilfunkgerätes und einem Sendemast eines Mobilfunknetzbetreibers. Grundsätzlich wird zwischen Mobiltelefondaten – sogenannten Call Detail Records (CDR) – und Mobilfunkdaten[1] unterschieden. Die primär zu Abrechnungszwecken erstellten CDR enthalten ereignisbasierte Informationen zu „aktiven Events“, die durch die Nutzenden initiiert werden, etwa beim Tätigen von Anrufen, dem Versenden von SMS, oder der Nutzung des mobilen Internets. Mobilfunkdaten im Sinne dieses Artikels umfassen neben aktiven Events weitere Signaldaten, nämlich passive Events, die systembedingt und ohne aktive Beteiligung der Nutzenden entstehen. Darunter fallen beispielsweise regelmäßige Abfragen des bereits verbrauchten Datenvolumens durch den Netzbetreiber. Informationen zur Dauer und Art des Events oder den Nutzenden selbst sind nicht enthalten. Grundvoraussetzung für das Senden und Empfangen von Signaldaten ist, dass die mobilen Endgeräte nicht ausgeschaltet oder im Flugmodus sind. In allen erwähnten Projekten in diesem Artikel

1 Mobilfunkdaten in der nicht aufbereiteten Form werden im Folgenden auch als Signaldaten bezeichnet.

**Übersicht 1**

**Qualitätsaspekte bei Mobilfunkdaten nach Aufbereitungsschritten**

| Prozessschritt | Beschreibung | Qualitätsaspekt | | Fehlertyp |
|---|---|---|---|---|
| Entstehung von Signaldaten | Generierung von Signaldaten aus aktiven und passiven Events | A1 | Irreversible Datenverluste bei Netzausfall oder Hardwareproblemen | Messfehler; Datenverlust |
| | | A2 | Signaldaten entstehen als „Nebenprodukt“, Frequenz und Art der Signale sind daher nicht für die Aufzeichnung von Aktivitäten optimiert und sind stark vom Nutzungsverhalten abhängig | Genauigkeitsunterschiede bei Subgruppen |
| | | A3 | Abweichungen bei erzeugten Events durch unterschiedliche Praktiken zwischen Mobilfunknetzbetreibern | Genauigkeitsunterschiede zwischen Mobilfunknetzbetreibern |
| Pseudonymisierung | Vergabe von (neuen) Pseudo-IDs (Rehashing) | B1 | Auswertungszeitraum begrenzt, wenn beispielsweise alle 24 Stunden ein Rehashing stattfindet, längsschnittliche Analysen nicht möglich | Zeitlicher Umfang; Fehlen von Längsschnittinformationen |
| Positionsschätzung | Positionen müssen geschätzt werden | C1 | Kleinste räumliche Dimension durch unterschiedliche Zellabdeckung regional sehr verschieden groß und Positionsschätzung sehr unterschiedlich präzise | Räumliche Genauigkeit |
| | | C2 | Die Annahme, dass die Antenne, mit der eine SIM-Karte kommuniziert, immer die räumlich nächste sei, ist empirisch nicht haltbar: „Trugschluss der nächstgelegenen Antenne“ (Ogulenko und andere, 2022) | Validität |
| Deduplizierung | Signale verschiedener SIM-Karten werden auf Ebene von SIMs (Ziel: Personen) zusammengefasst | D1 | Dopplungen durch mehrere Mobilfunkgeräte je Person möglich | Doppelzählungen |
| | | D2 | Unvollständige Identifikation von SIM-Karten aus anderen Geräten wie Laptops, Tablets oder Kraftfahrzeugen | Overcoverage |
| Ableitung weiterer Merkmale | Je nach Erkenntnisinteresse werden aus den Mobilfunkdaten weitere Merkmale abgeleitet | E1 | Abgrenzung einzelner Gruppen (zum Beispiel Berufstätige, Tagestouristinnen und -touristen) über die Merkmale Raum und Zeit nur unzureichend möglich | Validität |
| | | E2 | Soziodemografische Merkmale Altersgruppe und Geschlecht stark verzerrt | Verzerrung von Merkmalen; Datenverfügbarkeit |
| | | E3 | Einheitliche Parameter bei der Konstruktion von definitionsbedürftigen Konzepten wie „Bewegungen“ können zwischen Mobilfunknetzbetreibern variieren und sind zudem nicht immer bekannt | Operationalisierung; einheitliche Methodik |
| | | E4 | Unterrepräsentation von kurzen Distanzen durch die Notwendigkeit von Funkzellenwechseln für die Identifikation von Positionsänderungen (je geringer die Netzabdeckung, desto stärker) | Räumliche Genauigkeit (Folgeproblem von C1) |
| Aggregation | Daten werden nach festgelegten Kriterien aggregiert | F1 | Bezugsgrößen mit geringer Aktivität überproportional von Anonymisierungsverlusten betroffen | Asymmetrische Datenverluste |
| Hochrechnung | Da die Daten jedes Anbieters nur einen Teil der Bevölkerung abbilden, Hochrechnung auf Gesamtbevölkerung notwendig | G1 | Unsicherheit bei der Schätzung von Hochrechnungsfaktoren, wenn Marktanteile nicht kleinräumig genug bekannt sind und daher geschätzt werden müssen | Datenverfügbarkeit; Modellgüte |
| | | G2 | Verzerrung der Kundenstämme hinsichtlich relevanter Einflussfaktoren auf die Mobilität, Daten zur Korrektur nicht ausreichend vorhanden | Datenverfügbarkeit; Modellgüte |
| | | G3 | Nicht alle Teile der Bevölkerung nutzen Mobilfunk, durch die Opt-Out-Möglichkeit werden zudem nicht die Daten aller Nutzenden berücksichtigt | Undercoverage; Nonresponse |

wurden Mobilfunkdaten[2] genutzt, welche sowohl aktive als auch passive Events umfassen. Diese bieten gegenüber CDR eine deutlich breitere und weniger stark vom Nutzungsverhalten abhängige Datenbasis (Zhao und andere, 2016). Bei der Entstehung von Signaldaten sind folgende Qualitätsaspekte zu nennen:

A1. Die ordnungsgemäße Aufzeichnung der Signaldaten unterliegt Störungen, die durch Hardwareprobleme am Sendemast oder Ausfälle des Mobilfunknetzes

2 Der Begriff „Mobilfunkdaten“ wird in diesem Artikel zum einen als Oberbegriff und zum anderen für die fertig aufbereiteten Daten benutzt.

hervorgerufen werden können. Irreversible Datenverluste traten in einem Beispieldatensatz an etwa 1 % der beobachteten Tage auf. In den meisten Anwendungsfällen ist eine solche Ausfallfrequenz problemlos tolerierbar.

A2. Aktive Events sind vollständig vom Verhalten der Mobilfunknutzenden abhängig. Es ist davon auszugehen, dass einige wenige, sehr aktive Nutzende eine große Zahl von Events erzeugen, während eine breite Masse nur verhältnismäßig wenige aktive Events verursacht (Zhao und andere, 2016). Passive Events entstehen gewissermaßen als „Nebenprodukt" der Betreibung des Mobilfunknetzes. Ihre Regularität hinsichtlich Frequenz, Region und Art der Nutzenden ist nicht gewährleistet, denn auch bei passiven Events können durch bestimmte Nutzungsmuster mehr Datenpunkte erzeugt werden. Beide Effekte können in der Konsequenz zu Verzerrungen in den Daten führen.[3] Ob solche Verzerrungen systematisch vorkommen oder ob sich zufällige Verzerrungen in der Summe ausgleichen ist unklar: Die Messung und Quantifizierung des Effekts ist aufgrund der Komplexität der technischen Systeme auf Seiten der Mobilfunknetzbetreiber weder möglich, noch besteht ein wirtschaftlicher Anreiz dafür.

A3. Verschiedene Mobilfunknetze nutzen unterschiedliche Systeme zur Datenerfassung und generieren daher nicht zwangsläufig die gleichen Ereignisse mit der gleichen Frequenz. Werden in einem Netz etwa häufiger automatisierte Abfragen getätigt, die passive Events generieren, erhöht sich die Informationsdichte und damit die Genauigkeit der Daten. In der Praxis erschwert dies nochmals die Evaluation und Quantifizierung eines möglichen Effekts bei der Nutzung von Daten aus mehreren Mobilfunknetzen (was aus Gründen der Präzision erstrebenswert ist, siehe den Abschnitt „Hochrechnung").

## Pseudonymisierung

Vor der weiteren Verarbeitung werden individuelle Signaldaten durch die Vergabe von Pseudo-IDs pseudonymisiert. Damit werden potenziell sensible Informationen, beispielsweise die Telefonnummer, unzugänglich gemacht.

B1. Pseudo-IDs sind – abhängig von den geltenden Datenschutzbestimmungen – in regelmäßigen Abständen neu zu vergeben. In Deutschland hat der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit dieses sogenannte Rehashing alle 24 Stunden vorgeschrieben. Dadurch kann eine Abfolge von Aktivitäten nur innerhalb dieses Zeitraums eindeutig einem bestimmten (pseudonymisierten) Gerät zugeordnet werden. Dieses Vorgehen limitiert die Auswertungsmöglichkeiten in Deutschland (anders als in anderen Ländern) auf nur einen Tag; langfristige, longitudinale Analysen sind somit nicht möglich.

## Positionsschätzung

In einem nächsten Schritt werden auf Grundlage der erfassten Events in jeder Mobilfunkzelle unter Verwendung von Hilfsinformationen (wie der Ausrichtung einer Antenne und der Sendeleistung) die Positionen der SIM-Karten geschätzt. Da eine exakte Schätzung nicht möglich ist, wird das Mobilfunknetz häufig in kleinstmögliche Geometrien, beispielsweise Voronoi-Polygone[4], aufgeteilt, um die Position der SIM-Karten auf dieser Ebene zu schätzen.

C1. Hier entstehen große Unterschiede hinsichtlich der Genauigkeit der Positionsschätzung durch die unterschiedliche Abdeckung mit Sendemasten: Vor allem in ländlichen Regionen gibt es deutlich weniger Sendemasten, die aber eine größere Reichweite haben. Dagegen ist die Abdeckung in verstädterten Regionen deutlich dichter. Folglich sind Funkzellen in Städten in der Regel deutlich kleiner als auf dem Land; die Genauigkeit der Positionsschätzung variiert regional stark. Dies ist bei der Analyse klein-

3 Wenn beispielsweise jüngere Nutzende aufgrund der Buchung von mehr Zusatzdienstleistungen häufiger passive Events erzeugen, dann ist eine genauere Schätzung des Bewegungsverlaufs möglich. Da bei nicht geradliniger Bewegung mehr Datenpunkte zu einer längeren geschätzten Bewegungsdistanz führen, können Bewegungsdaten durch diesen Effekt systematisch verzerrt werden.

4 Methode zur Zerlegung von geografischen Räumen in Polygone, wonach Flächen dem jeweils nächstgelegenen Punkt (in diesem Fall den Sendemasten) zugeordnet wird. Entspricht der Annahme, dass sich jede SIM-Karte mit dem nächstgelegenen Sendemast verbindet.

räumiger Ergebnisse zu berücksichtigen und kann zu Folgeproblemen führen (siehe E4 im Abschnitt „Ableitung weiterer Merkmale“).

C2. Die Validität der Positionsschätzung basiert auf der Annahme, dass sich Geräte in der Regel mit dem nächstgelegenen Sendemast verbinden. Jüngste Studien stellen dies infrage: Ogulenko und andere (2022) etwa kommen im Abgleich mit GPS-Daten zu dem Schluss, dass sich die untersuchten Mobilfunkgeräte (unter anderem aufgrund wesentlicher Überschneidungen zwischen den Versorgungsgebieten) in 60 % der Fälle nicht mit dem Sendemast verbinden, der geografisch am nächsten liegt. Konkret bedeutet dies, dass sich Geräte oft nicht innerhalb der vermuteten Voronoi-Polygone befunden haben. Dieser „Trugschluss der nächstgelegenen Antenne“ (Ogulenko und andere, 2022) lässt auf eine stark erhöhte Standortunsicherheit bei der Positionsschätzung schließen. Wie groß diese erhöhte Unsicherheit ist, und ob damit auf aggregierter Ebene auch eine Verzerrung der Daten einhergeht, darüber lässt sich auf Grundlage der dem Statistischen Bundesamt zur Verfügung stehenden Daten keine Aussage treffen.

## Deduplizierung

Signaldaten werden auf Ebene von SIM-Karten erhoben, die gewünschte Beobachtungseinheit der aufbereiteten Mobilfunkdaten ist allerdings die Gesamtbevölkerung. Um von SIM-Karten auf Geräte und schließlich auf Personen schließen zu können, ist eine Deduplizierung erforderlich.

D1. Während die Zuordnung verschiedener Signale zu einer SIM-Karte unproblematisch ist, können Doppelzählungen auf Personenebene entstehen, wenn in einem Gerät mehrere SIM-Karten verwendet werden oder eine Person mehrere Mobilfunkgeräte mit SIM-Karte mit sich führt.

D2. Außer in Mobiltelefonen und Smartphones sind SIM-Karten unter anderem auch in Tablets, Laptops und Kraftfahrzeugen zu finden. Laut Aussage der Mobilfunknetzbetreiber ist es meist nur mit hohem Aufwand möglich, diese zu identifizieren und herauszurechnen, allerdings würde nur ein kleiner Teil der Events von anderen Geräten als Mobiltelefonen oder Smartphones erzeugt.

## Ableitung weiterer Merkmale

Die so aufbereiteten Mobilfunkdaten lassen Rückschlüsse über die Verteilung der Bevölkerung auf verschiedenen geografischen Ebenen und in bestimmten zeitlichen Intervallen zu. Die flexible Festlegung der zeitlich-räumlichen Dimensionen eröffnet umfassende Möglichkeiten, Auswertungen kleinräumig (zum Beispiel auf Stadtteilebene) und zeitlich genau (zum Beispiel stündlich) durchzuführen.

E1. Schwieriger ist die Abgrenzung einzelner Gruppen wie der Berufstätigen oder Tagestouristinnen und -touristen (Grassini/Dugheri, 2021). Neben den Merkmalen Raum, Zeit und Roaming-Status[5] liegen praktisch keine weiteren Informationen zur SIM-Karte vor, die bei der Eingrenzung eines bestimmten Personenkreises behilflich sein könnten.

E2. Zwar sind die soziodemografischen Merkmale Altersgruppe und Geschlecht verfügbar, jedoch nur für Vertragskundinnen und -kunden, da nur diese im Customer-Relationship-Management-System (CRM-System) hinterlegt sind; Prepaidkundinnen und -kunden sind nicht repräsentiert. Zudem sind die Daten stark verzerrt, da die Vertragspersonen in vielen Fällen nicht mit den Nutzenden der SIM-Karten übereinstimmen (beispielsweise aufgrund von Familienverträgen) sowie die Verteilung der Prepaidkundinnen und -kunden nicht bekannt ist. Insbesondere fehlen Nichtvertragsmündige in den Daten komplett (Statistisches Bundesamt, 2021a).

Neben diesen statischen Daten lassen sich aber auch Aussagen über Mobilitätsströme ableiten. Zur Generierung dieser dynamischen Daten werden Positionsänderungen von SIM-Karten genutzt, um Bewegungen zu identifizieren. Dafür werden Signale im Zeitverlauf in Nicht-Bewegungs-Phasen und Positionsänderungen unterteilt. Welche zeitlich-räumliche Positionsveränderung als Bewegung gilt, bis wann Bewegungsunterbrechungen als Pausen gelten und ab wann von separaten Bewegungen auszugehen ist – all dies muss definiert werden. Zudem müssen die Daten weiter aufbereitet werden, um etwa zu verhindern, dass häufige Funkzellenwechsel von Endgeräten, die sich zwischen zwei Sendemasten befinden, fälschlicherweise als Bewegungen identifiziert werden.

5 Das heißt ob es sich um eine SIM-Karte aus dem Ausland handelt.

E3. Methodische Entscheidungen dieser Art sind zu dokumentieren und bei der Nutzung von Daten mehrerer Datenanbieter unbedingt zu vereinheitlichen. Qualitätsbedenken bestehen, da die Datenanbieter in der Regel aufgrund von Geschäftsgeheimnissen einige Details vorenthalten und Parameter nicht immer flexibel angepasst werden können.

E4. Regionale Unterschiede bei der Genauigkeit der Positionsschätzung (siehe C1) führen bei der Bewegungsidentifikation zu Verzerrungen: Da Bewegungen erst bei einem Wechsel der Funkzelle identifiziert werden können, findet in ländlichen Regionen ein Funkzellenwechsel erst bei größeren Positionsveränderungen statt. Bewegungen über kurze Distanzen sind folglich dort unterrepräsentiert, wo die Sendemasten deutlich weiter auseinanderliegen. Dies kann algorithmisch interpoliert werden; die Genauigkeit einer solchen Interpolation müsste jedoch noch durch externe Studien validiert werden.

## Aggregation

Um auf Basis der aufbereiteten Daten Analysen durchführen zu können, werden sie auf eine festgelegte räumliche und zeitliche Ebene sowie gegebenenfalls nach weiteren Merkmalen aggregiert. Dies reduziert deutlich die Größe der Daten, die sonst mit herkömmlicher Hardware nicht nutzbar wären. Zudem erfüllt die Aggregation den Zweck, dass keine Aussage über den Aufenthaltsort oder die Bewegungen einzelner Personen gemacht werden können. Nach den Vorgaben des Bundesbeauftragten für Datenschutz und Informationssicherheit werden Einträge, die sich nicht auf eine bestimmte Mindestanzahl von Personen beziehen (aktuell: $N > 5$), bei der Aufbereitung aus dem Datensatz gelöscht.

F1. Daher sind Bezugsgrößen mit geringer Aktivität, beispielsweise die Auswärtsbewegungen aus einer ländlichen Region nachts zwischen 03:00 und 04:00 Uhr, überproportional von Anonymisierungsverlusten betroffen. Dies kann teilweise interpoliert werden, wenn Informationen über das Ausmaß der Anonymisierungsverluste vorliegen.

## Hochrechnung

Um präzise Aussagen nicht nur über die Nutzenden eines bestimmten Netzanbieters, sondern über die Gesamtbevölkerung treffen zu können, ist die Hochrechnung der Mobilfunkdaten notwendig.

G1. Hierbei ist zu berücksichtigen, dass sich die Marktanteile der Netzbetreiber regional stark unterscheiden. Diese sind aus Wettbewerbsgründen aber nicht kleinräumig verfügbar und müssen ihrerseits geschätzt werden – eine weitere Fehlerquelle, die insbesondere beim Vergleich von Daten verschiedener Mobilfunkanbieter zu berücksichtigen ist.

G2. Eine Hochrechnung anhand von Marktanteilen nimmt implizit an, dass sich die Kundenstämme der Mobilfunknetzbetreiber hinsichtlich der interessierenden Variablen nicht unterscheiden. Es muss davon ausgegangen werden, dass soziodemografische Unterschiede zwischen den Kundenstämmen der Anbieter bestehen (wie in Statistisches Bundesamt [2021a] am Beispiel des Bundeslandes Nordrhein-Westfalen exemplarisch dargestellt wurde) und dass sich zum Beispiel Unterschiede in der Altersstruktur stark auf das Bewegungsverhalten auswirken. Auf Grundlage der vorhandenen Datenquellen besteht bisher keine zufriedenstellende Möglichkeit, Verzerrungen zwischen den Mobilfunkanbietern modellbasiert zu schätzen und auszugleichen. Deshalb wäre der Zugang zu Daten aller Mobilfunkanbieter wichtig, um verlässliche Aussagen über die Gesamtbevölkerung treffen zu können.

G3. Auch wenn Daten aller Anbieter genutzt werden können, wird ein kleiner Teil der Bevölkerung nicht repräsentiert: 2,4 % der privaten Haushalte nutzen kein Mobilfunkgerät (Statistisches Bundesamt, 2021b), zudem können die Nutzenden der Verarbeitung ihrer Daten widersprechen (Opt-Out). Diese Ausfallquoten sind gering und stellen kein ernst zu nehmendes Qualitätsproblem dar.

Zusammenfassend wird klar, dass der Entstehungs- und Verarbeitungsprozess von Mobilfunkdaten aus Qualitätssicht eine Vielzahl von inhärenten Herausforderungen mit sich bringt. Weitere Fehlerquellen entstehen aus datenschutzrechtlich vorgeschriebenen Arbeitsschritten zur Pseudonymisierung und Anonymisierung. Sollen Aussagen über die Gesamtbevölkerung getroffen werden

– was Ziel der amtlichen Statistik ist –, folgen darüber hinaus wichtige Qualitätseinschränkungen aus der Tatsache, dass der Markt in mehrere Anbieter segmentiert ist. Die gemeinsame Betrachtung der Daten aller drei Anbieter kann marktspezifische regionale Verzerrungen verhindern. Bei der Nutzung der Daten aller Anbieter müsste zusätzlich eine einheitliche Methodik sichergestellt werden.

# 3

# Projekterfahrungen

Im Zuge der fortschreitenden Digitalisierung hat das Statistische Bundesamt Ende 2017 damit begonnen, die Datenquelle Mobilfunkdaten für Zwecke der amtlichen Statistik zu erforschen. Chancen werden vor allem in einer Ergänzung des bisherigen Datenangebots gesehen: Den Nutzenden können über neue Datenquellen Informationen zu Themengebieten bereitgestellt werden, welche bislang durch die amtliche Statistik nicht zur Verfügung stehen – und dies auf kleinräumiger Ebene und hochaktuell. Des Weiteren werden mit der Forschung Qualitätsverbesserungen bei amtlichen Statistiken hinsichtlich Aktualität und Genauigkeit erhofft sowie eine Entlastung von Auskunftspflichtigen, beispielsweise wenn Befragungsmerkmale durch Mobilfunkdaten ersetzt werden könnten.

Seit Beginn der Forschungen mit Mobilfunkdaten wurden Kooperationen mit Datenanbietern etabliert, die Daten aus dem Netz der Deutschen Telekom und der Telefónica Deutschland generieren; insgesamt wurden sechs Projekte abgeschlossen. Im Folgenden sollen anhand dieser Projekte beispielhaft Nutzungs- und Analysepotenziale herausgearbeitet werden und illustriert werden, inwiefern die in Kapitel 2 vorgestellten Qualitätsprobleme aufgetreten sind.

## Analysen zur Bevölkerungsverteilung

Mobilfunkdaten ermöglichen potenziell eine dynamische und kleinräumige Abbildung der Bevölkerungsverteilung und sind daher prädestiniert, Fragen der Bevölkerungsforschung zu beantworten. Erste Projekte zum Nutzungspotenzial von Mobilfunkdaten im Statistischen Bundesamt untersuchten daher, inwieweit die Tages- und Wohnbevölkerung valide, aktuell und kleinräumig mit Mobilfunkdaten abgebildet werden kann (Statistisches Bundesamt, 2019; Hadam und andere, 2020a).

Eine Validierung mit den Ergebnissen des Zensus 2011 zeigte vielversprechende Resultate. Auf Basis dieser Daten kann eine experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung für die Berichtsjahre 2019 und 2020 erstellt und veröffentlicht werden. Damit lässt sich die Zeit bis zur Veröffentlichung der ersten amtlichen georeferenzierten Bevölkerungszahl Ende 2022 überbrücken (Statistisches Bundesamt, 2022a). Während des Projekts zeigte sich, dass die verwendeten Mobilfunkdaten nicht ohne Weiteres auf Ebene eines den Anforderungen der amtlichen Statistik entsprechenden, INSPIRE[6]-konformen 1x1-km-Rasters geliefert werden konnten: Da die Abdeckung mit Mobilfunkmasten insbesondere in ländlichen Regionen weniger fein ist (siehe C1), musste in Zusammenarbeit mit dem Datenanbieter eine Methodik entwickelt werden, um größere Mobilfunkzellen auf das gewünschte 1x1-km-Raster zu verteilen. Eine naive, proportionale Verteilung anhand von Flächenanteilen führte zu positiven Bevölkerungsschätzungen in (unbewohnten) Wald- und Seegebieten und war daher nicht zufriedenstellend. Deshalb wurden Geodaten des Landbedeckungsmodells des Bundesamtes für Kartographie und Geodäsie verwendet, um unbewohnte Regionen im Datenaufbereitungsprozess zu lokalisieren, diese nicht weiter einzubeziehen und so letztlich alle mobilen Aktivitäten sinnvoll zu verorten.

Sind Eingriffe in den Datenaufbereitungsprozess nicht möglich und liegen die Mobilfunkdaten für ein Mischraster[7] vor, so müssen Umrechnungsmethoden angewandt werden, um die Daten auf eine einheitliche Geometrie zu projizieren. Bei der Bevölkerungsdarstellung in Hadam und andere (2020a) sowie Statistisches Bundesamt (2019) wird hierfür beispielsweise eine Kerndichteschätzung genutzt. Dieses Vorgehen erzeugt eine zusätzliche Unsicherheit in den Mobilfunkdaten und folglich auch in der daraus abgeleiteten Schätzung der Tages- und Wohnbevölkerung.

6 Zu INSPIRE, dem Vorhaben für eine gemeinsame Geodateninfrastruktur in Europa, siehe auch Brenzel/Gebers (2020, hier: Seite 49 f.).

7 Darunter sind Gitterzellen mit unterschiedlichen Gitterweiten zu verstehen.

Basierend auf den Erkenntnissen der Bevölkerungsdarstellung wurde – ausgehend vom ESSnet-Projekt[8] „City Data from LFS and Big Data" – die Erwerbslosenquote aus der Arbeitskräfteerhebung[9] mittels eines Small-Area-Verfahrens auf kleinräumiger Ebene geschätzt (Europäische Kommission, 2019; Hadam und andere, 2020b). Erstmals wurden hier Mobilfunkdaten als Hilfsinformationen zur Schätzung einer amtlichen Statistik integriert, wodurch sich die Ergebnisse nachweislich verbesserten. Darauf aufbauend nutzten Hadam und andere (2020b) aus den Daten abgeleitete Pendlerbewegungen und implementierten eine Bias-Korrektur, um die Schätzung weiter zu verbessern und eine alternative Erwerbslosenquote für das Testgebiet Nordrhein-Westfalen zu produzieren. Trotz grundlegend positiver Ergebnisse wurde die Machbarkeitsstudie nicht weiterverfolgt: Nach Einschätzung des Fachbereichs waren alternative Daten der Bundesagentur für Arbeit besser geeignet; zudem bestanden Zweifel an der Qualität der Mobilfunkdaten. Zum einen wurde vermutet, dass der Einbezug der soziodemografischen Merkmale verzerrte Ergebnisse provozieren würde (siehe E2). Zum anderen wurde beanstandet, dass Ungenauigkeiten und Verzerrungen vage und nicht quantifizierbar seien (siehe A2, A3, C2 und E4).

Um Erkenntnisse über die Aussagekraft der Mobilfunkdaten zu gewinnen und die Qualitätseinschränkungen G1 und G2 zu prüfen, müssen Daten verschiedener Mobilfunkanbieter für das gleiche Testgebiet und den gleichen Zeitraum zusammengeführt und verglichen werden. Dies wurde in einem Projekt umgesetzt: Anhand nicht hochgerechneter Mobilfunkdaten zweier Mobilfunkanbieter (die etwa zwei Drittel des Mobilfunkmarkts abdecken) wurde untersucht, inwieweit eine Kombination zweier Datensätze zur Steigerung der Präzision der Mobilfunkdaten führen kann (Statistischen Bundesamt, 2021a). Im Ergebnis erbrachte eine Zusammenführung von mobilen Aktivitäten verschiedener Mobilfunkanbieter eine deutliche Steigerung der Qualität, da sich die Mobilfunkdaten entsprechend ergänzen konnten und Unsicherheiten durch die jeweilige Hochrechnung weniger zum Tragen kamen – trotz der unterschiedlichen Datenaufbereitungsprozesse der Datenanbieter. Die starke Verzerrung der soziodemografischen Angaben (siehe E2) war durch das Zusammenführen der Datensätze erwartungsgemäß nicht abzumildern.

## Analysen zur Bevölkerungsmobilität

Neben Aussagen zur Bevölkerungsverteilung sind mit Mobilfunkdaten auch Analysen zur Bevölkerungsmobilität möglich. Hadam (2021) führte im Projekt „Pendler Mobil" in Kooperation mit IT.NRW – Statistisches Landesamt Nordrhein-Westfalen Analysen zur Bevölkerungsmobilität in Nordrhein-Westfalen durch, um Bereiche zu identifizieren, in denen Mobilfunkdaten zur amtlichen Pendlerrechnung beitragen können. Hohe Korrelationen zwischen der anhand von Mobilfunkdaten geschätzten Pendlermobilität und den Ergebnissen der amtlichen Pendlerrechnung machen deutlich, dass Mobilfunkdaten im Allgemeinen das Potenzial haben, die amtliche Pendlerrechnung zu unterstützen. Dabei war die Identifikation von Pendlerinnen und Pendlern ausschließlich auf Grundlage des Bewegungsprofils[10], das heißt ohne Nutzung soziodemografischer Merkmale, möglich. In diesem Fall stellte E2 also keine größere Qualitätseinschränkung dar. Jedoch zeigte sich, dass ein Zielkonflikt zwischen der Genauigkeit und dem Informationsgehalt der Daten besteht (siehe F1): Werden Pendlerverflechtungen auf Gemeindeebene geschätzt, so ist dies nur dann ohne große Anonymisierungsverluste möglich, wenn die Pendlerdefinition verallgemeinert wird. Auch wurde trotz guter Korrelationen die Gesamtzahl der Pendlerinnen und Pendler aufgrund des ungenügenden Detailgrads bei der Hochrechnung (siehe G2) unterschätzt. Insgesamt konnte zwar die Forschungsfrage zufriedenstellend beantwortet werden, detailliertere Fragestellungen konnten mit dem gegebenen Datensatz allerdings nicht flexibel untersucht werden.

Wenn nicht nach Subgruppen unterschieden werden muss, sondern die allgemeine Mobilität der Bevölkerung von Interesse ist, dann bieten Mobilfunkdaten ein brei-

8 ESSnet-Projekte sind ein Instrument im Europäischen Statistischen System (ESS), um in den ESS-Organisationen verteiltes Fachwissen zusammenzuführen sowie spezifische Maßnahmen zu entwickeln, die dem gesamten System zugutekommen.

9 In der Arbeitskräfteerhebung (Labour Force Survey — LFS) der Europäischen Union (EU) erheben alle EU-Mitgliedstaaten differenzierte Daten zur Erwerbstätigkeit in gleicher Form und ermöglichen so eine harmonisierte Arbeitsmarktberichterstattung und Vergleiche der Beschäftigungssituation. In Deutschland ist die LFS in den Mikrozensus integriert.

10 Folgendes Bewegungsprofil wurde zur Pendlerbestimmung genutzt: Bewegung über Gemeindegrenze hinweg, Eintritt in Zielgemeinde vor 09:00 Uhr, Mindestverweildauer vor Rückkehr in Startgemeinde von mindestens zwei Stunden.

tes Analysepotenzial: So wurde mit dem Projekt „Mobilitätsindikatoren auf Basis von Mobilfunkdaten“ während der Corona-Pandemie ein bedarfsorientiertes Datenangebot geschaffen. Täglich aktualisierte Datenprodukte und anlassbezogene Pressemitteilungen stellten eine Datengrundlage für eine faktenbasierte öffentliche Diskussion bereit. Inhaltlich lieferte etwa eine Analyse der Mobilität in Corona-Hotspots wichtige Einblicke in die Wirkung der Maßnahmen und das Mobilitätsverhalten der Bevölkerung (Statistisches Bundesamt, 2022b; Bohnensteffen und andere, 2021). Dabei wurden auch wichtige praktische Erkenntnisse zur Qualität von Mobilfunkdaten gesammelt: In enger Kooperation mit dem Datenanbieter wurden methodische Entscheidungen diskutiert und in einer Methodendokumentation festgehalten (siehe E3). Dies ermöglicht ein vertieftes Verständnis der Daten und eine fundierte Einschätzung von Effekten und Besonderheiten der Datenquelle. Erfolgt die Neuschätzung der regionalen Marktanteile beziehungsweise Hochrechnungsfaktoren beispielsweise monatlich statt jährlich, werden saisonale Abweichungen der üblichen Bevölkerungsstruktur in touristischen Regionen zu stark berücksichtigt und generieren große Verzerrungen der geschätzten Bevölkerungsmobilität (siehe G2). Damit zeigte sich auch, wie wichtig vermeintlich kleine methodische Entscheidungen sind: Wurden solche Effekte in den Daten bemerkt, erfolgte umgehend eine Korrektur seitens des Datenanbieters.

Der Abschluss der Projekte und Eignungsprüfungen zu Mobilfunkdaten im Statistischen Bundesamt führt zu folgendem Resümee: Mobilfunkdaten haben als junge Datenquelle noch keine endgültige Reife erreicht, sondern befinden sich in der Weiterentwicklung. Hinsichtlich ihrer Qualität sind bereits in kurz- bis mittelfristiger Zukunft deutliche Verbesserungen zu erwarten. In der Praxis beruht ein Großteil der beobachteten Qualitätseinschränkungen darauf, dass lediglich mit prozessierten Daten gearbeitet werden kann, deren Aufbereitung andere Akteure vornehmen. Diese vertraglich begründeten Datenzugänge verursachen eine erhebliche Abhängigkeit vom jeweiligen Datenanbieter. Sie erschweren es maßgeblich, die in Kapitel 2 „Datengenese und Qualitätsaspekte“ genannten Qualitätseinschränkungen zu quantifizieren sowie die Ergebnisse der Machbarkeitsstudien zu bewerten. Hierbei ist zu betonen, dass die Kooperationen nicht mit den Mobilfunkanbietern selbst, sondern mit Tochterunternehmen oder weiteren Drittanbietern bestanden oder noch bestehen. Das bedeutet, dass im Falle des Marktaustritts oder der Auflösung der Datenanbieter die Gefahr realistisch ist, den Datenzugang zu verlieren. Das Fehlen eines uneingeschränkten, langfristig gesicherten Wissens- und Datenzugangs macht das Überführen von experimentellen Produkten auf Basis von Mobilfunkdaten in die amtliche Statistikproduktion derzeit nahezu unmöglich. Für die dauerhafte Nutzung aussagefähiger Daten wird somit dringend ein rechtssicherer Zugang zu den anonymisierten Daten aller Mobilfunkanbieter in Deutschland für Statistikzwecke benötigt.

## 4

## Nutzungsmöglichkeiten von Mobilfunkdaten

Die Betrachtung der Qualität und des Potenzials von Mobilfunkdaten anhand ihres Entstehungs- und Verarbeitungsprozesses liefert ein Bild der Vor- und Nachteile dieser Datenquelle. Nach der Diskussion der technisch-methodischen Grenzen und der praktischen Erfahrungen mit Mobilfunkdaten soll nun beschrieben werden, für welche Anwendungsfälle sich Mobilfunkdaten besonders gut nutzen lassen.

Mobilfunkdaten eignen sich insbesondere für Auswertungen, deren Grundgesamtheit Personen im Gebiet der Bundesrepublik Deutschland sind (nicht: Unternehmen, Haushalte). Den höchsten Informationsgehalt liefern die vorhandenen Informationen, wenn die Statistik mit hoher Frequenz (zum Beispiel täglich oder stündlich) und hoher Aktualität (bis zu t + 2) veröffentlicht wird. Im günstigsten Fall weist die Statistik zudem eine räumliche Auflösung unter der Bundes- oder Landesebene auf (zum Beispiel Kreise, Gemeinden, Stadtteile). Gleichzeitig sollte die gewünschte Auflösung aber nicht kleinräumiger als die Dichte des Mobilfunknetzes sein. Beispielsweise können einige Datenanbieter verlässliche Daten für bis zu 100 000 feinräumige Gebiete mit durchschnittlich 400 Haushalten bereitstellen; eine genauere räumliche Zuordnung ist derzeit nicht möglich. Damit eignen sich Mobilfunkdaten kaum für Statistiken mit einem Fokus auf sehr kleine räumliche Regionen (zum Beispiel Einkaufszentren, Straßen, Parks). Füllen gewünschte räumliche Einheiten zumindest den Großteil von Mobilfunk-

zellen, dann ist es möglich, Aktivitäten zum Beispiel proportional anhand des Flächenanteils oder der Bevölkerungsdichte zu verteilen. Dies führt allerdings zu Unsicherheiten und Fehlerquellen, für deren Quantifizierung zudem eine Methodik zu erarbeiten wäre.

Aufgrund von Unsicherheiten bei der Deduplizierung und Hochrechnung auf die Gesamtbevölkerung gibt es Vorbehalte, statische Mobilfunkdaten als „Personen" und dynamische Mobilfunkdaten als „Bewegungen"[11] zu interpretieren. Daher sollten Aussagen zu absoluten Aktivitäts- oder Bewegungsdaten vermieden werden. Stattdessen wird empfohlen relative Aussagen über die Entwicklung im Zeitverlauf – etwa relativ zum saison- und kalenderbereinigten Durchschnitt der Vorjahre oder eines fixen Referenzzeitraums zu treffen. Konkret ist dies der Fall, wenn auf Grundlage der absoluten Daten Veränderungsraten oder Indexwerte berechnet werden. Wenn eine Kalibrierung der absoluten Zahlen auf aggregierter Ebene[12] anhand anderer, verlässlicher Datenquellen möglich ist, können potenziell auch absolute Werte adäquate Informationen liefern.

Sollten Subgruppen von Interesse sein, so müssen sie über die vorhandenen Merkmale – Zeit und Ort, sowie bei Bewegungsdaten Geschwindigkeit und Verkehrsmittel – identifizierbar sein. Die soziodemografischen Merkmale von Vertragskundinnen und -kunden sollten – aufgrund der vorhandenen Verzerrungen beziehungsweise Auswahl – für die Filterung nach Subgruppen nur mit Vorsicht genutzt werden. Beispielsweise ließen sich Studierende und Mitarbeitende einer Campus-Universität mithilfe von Mobilfunkdaten gut identifizieren, wenn die geografische Restriktion des Hochschulgeländes mit einem zeitlichen Filter (zum Beispiel zwischen 08:00 und 20:00 Uhr an Werktagen) und einer Mindestaufenthaltsdauer von etwa 60 Minuten kombiniert wird. Im so gefilterten Datensatz ist zu erwarten, dass die gewünschte Subgruppe einen hinreichend großen Anteil der Beobachtungen ausmacht, um beispielsweise zeitliche Entwicklungen in den Daten auf Veränderungen in der Aktivität der Subgruppe zurückzuführen. Je größer (und dynamischer) die Personengruppen, die nicht über vorhandene Merkmale aus den Daten herauszufiltern sind, desto weniger verlässlich können Aussagen über die gewünschte Subgruppe getroffen werden. Dies wäre etwa der Fall, wenn eine Hochschule direkt neben einem Einkaufszentrum liegt und vonseiten des Datenanbieters keine hinreichend trennscharfe Aufteilung der Geografien möglich ist.[13]

## 5

## Fazit und Ausblick

Mobilfunkdaten stellen eine komplexe Datenquelle dar, die das bisherige Datenangebot ergänzen kann sowie die Qualität von amtlichen Statistiken verbessern könnte. Um diese sekundären, das heißt ursprünglich für einen anderen Zweck erstellten, Daten für statistische Zwecke nutzbar zu machen, sind umfangreiche Aufbereitungsschritte nötig.

Der Entstehungs- und Verarbeitungsprozess von Mobilfunkdaten bringt aus Qualitätssicht eine Vielzahl von inhärenten Herausforderungen mit sich. Datenschutzrechtlich vorgeschrieben sind die Pseudonymisierung und Anonymisierung dieser Daten. Dies beschränkt die räumliche und zeitliche Tiefe, in der Auswertungen möglich sind. Um hinreichend präzise Aussagen auf Ebene der Gesamtbevölkerung treffen zu können die den Qualitätsanforderungen der amtlichen Statistik entsprechen, müssen folgende Herausforderungen bewältigt werden:

a) Verzerrungen, die aus den unterschiedlichen Kundenstämmen der Mobilfunkanbieter entstehen, müssen quantifiziert und korrigiert werden, um eine präzise Hochrechnung zu ermöglichen. Sollte dies nicht möglich sein, sollten alternativ die Daten aller Mobilfunknetzbetreiber miteinbezogen werden.

11 Das Konzept einer „Bewegung" im allgemeinen Sprachgebrauch ist nicht scharf und unterliegt oft der subjektiven Einschätzung der Sprechenden: Wenn man auf dem Weg zur Arbeit noch die Apotheke besucht, handelt es sich dann um zwei Bewegungen oder um eine Bewegung mit Zwischenstopp? Ändert sich die Einschätzung, wenn der Apothekenbesuch „länger" dauert? Oder wenn für den Apothekenbesuch ein „kleiner" Umweg genommen wird? Die „technische" Bewegungsdefinition bei den Mobilitätsindikatoren entspricht also nicht unbedingt einem intuitiven, scharf abgegrenzten Konzept in der Alltagssprache.

12 Eine vollständig vergleichbare Datenquelle existiert nicht, dies ist schließlich eines der Kernargumente für die Nutzung von Mobilfunkdaten. Möglicherweise gibt es aber Daten mit höherer zeitlicher und räumlicher Aggregation, anhand derer eine Kalibrierung erfolgen kann.

13 In diesem Fall erlauben die verbleibenden Merkmale Zeit und Mindestaufenthaltsdauer keine Unterscheidung zwischen Studierenden und einkaufenden Personen.

b) Verzerrungen, die aus unterschiedlichen Datenerfassungssystemen resultieren, müssen ausgeglichen werden.

c) Bei Nutzung der Daten mehrerer Mobilfunknetzbetreiber (siehe a) müssen methodische Aufbereitungsschritte harmonisiert werden.

Gelingt dies nicht, ist mit regionalen und subgruppenspezifischen Verzerrungen zu rechnen.

Die durchgeführten Machbarkeitsstudien zeigen grundsätzlich, dass Mobilfunkdaten das Potenzial haben, die amtliche Statistik zu unterstützen und konkrete Statistiken zu ergänzen. Um die Datenquelle jedoch adäquat nutzen zu können, sind tiefes technisches Wissen und ein detailliertes methodisches Verständnis unabdingbar. Kontinuierliche methodische Neuerungen im Bereich der Mobilfunktechnologie erfordern dafür die enge Kommunikation zwischen technischen Fachleuten und statistischen Nutzenden. Langfristig ist eine Fernverarbeitung der Daten durch die amtliche Statistik bei den Anbietern wünschenswert, um harmonisierte Mobilfunkdaten aufzubereiten und zudem qualitativ hochwertige Statistiken unter Einbezug amtlicher Qualitätskriterien produzieren zu können. Ein nachhaltiger Datenzugang bedarf zudem einer Rechtsgrundlage zur Nutzung externer Datenquellen.

Hier zeigt sich ein Dilemma: Die Etablierung einer Rechtsgrundlage ist ohne Anwendungsbeispiele, die direkt in die amtliche Produktion überführt werden können, nicht realistisch. Jedoch ist eine tiefergehende Prüfung ohne uneingeschränkten Datenzugang schwer durchführbar. Insbesondere werden die Mobilfunkdaten für die jeweiligen konkreten Fragestellungen und Anwendungszwecke aufbereitet. Andere Fragestellungen können damit nicht beantwortet werden. Dadurch entsteht ein Zielkonflikt zwischen den Machbarkeitsstudien für die Datenquelle und der Nutzbarkeit der beschafften Daten für konkrete Anwendungen in amtlichen Fachstatistiken.

Was bedeutet dies für die Nutzung der Datenquelle „Mobilfunkdaten“ in der amtlichen Statistik? Einerseits eignen sich Mobilfunkdaten sehr gut für Sonderauswertungen, die mit verfügbaren amtlichen Daten fachlich nicht abzudecken sind oder die als Reaktion auf aktuelle Ereignisse (wie die Corona-Pandemie) eine schnelle (möglichst tagesaktuelle) Reaktion auf kleinräumiger Ebene erfordern. Andererseits wird aber auch deutlich, dass eine verlässliche Statistikproduktion unter Berücksichtigung amtlicher Qualitätskriterien derzeit mit Mobilfunkdaten nicht möglich ist. Technische Weiterentwicklungen dieser Datenquelle und der Ausbau des 5G-Netzes werden mittelfristig eine Neubewertung nötig machen.

## LITERATURVERZEICHNIS

Bohnensteffen, Sarah/Mühlhan, Jannek/Saidani, Younes. *Mobilität während der Corona-Pandemie*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2021, Seite 89 ff.

Brenzel, Hanna/Gebers, Kathrin. *Werkstattbericht: Georeferenzierung im Statistischen Verbund*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2020, Seite 48 ff.

Bundesnetzagentur. *Teilnehmerentwicklung im Mobilfunk*. 2022. [Zugriff am 13. Juni 2022]. Verfügbar unter: www.bundesnetzagentur.de

Europäische Kommission. *City data from LFS and big data*. 2019. [Zugriff am 26. Juli 2022]. Verfügbar unter: ec.europa.eu

Grassini, Laura/Dugheri, Gianni. *Mobile phone data and tourism statistics: a broken promise?* In: National Accounting Review. Jahrgang 3. Ausgabe 1/2021, Seite 50 ff. [Zugriff am 26. Juli 2022]. Verfügbar unter: doi.org/10.3934/nar.2021002

Hadam, Sandra/Schmid, Timo/Simm, Joanna. *Kleinräumige Prädiktion von Bevölkerungszahlen basierend auf Mobilfunkdaten aus Deutschland*. In: Klumpe, Bettina/Schröder, Jette/Zwick, Markus (Herausgeber). Qualität bei zusammengeführten Daten. Wiesbaden 2020a, Seite 27 ff. [Zugriff am 26. Juli 2022]. Verfügbar unter: doi.org/10.1007/978-3-658-31009-7_3

Hadam, Sandra/Würz, Nora/Kreutzmann, Ann-Kristin. *Estimating Regional Unemployment with Mobile Network Data for Functional Urban Areas in Germany*. Refubium – Repositorium der Freien Universität Berlin. 2020b. [Zugriff am 26. Juli 2022]. Verfügbar unter: dx.doi.org/10.17169/refubium-26791

Hadam, Sandra. *Pendler Mobil: Die Verwendung von Mobilfunkdaten zur Unterstützung der amtlichen Pendlerstatistik*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 15. 2021. Seite 197 ff. [Zugriff am 26. Juli 2022]. Verfügbar unter: doi.org/10.1007/s11943-021-00294-z

Ogulenko, Aleksey/Benenson, Itzhak/Toger, Marina/Östh, John/Siretskiy, Alexey. *The fallacy of the closest antenna: Towards an adequate view of device location in the mobile network*. In: Computers, Environment and Urban Systems. Band 95, 101826. 2022. [Zugriff am 26. Juli 2022]. Verfügbar unter: doi.org/10.1016/j.compenvurbsys.2022.101826

Statistisches Bundesamt. *Bevölkerungsdarstellung mit Mobilfunkdaten*. 2019.

Statistisches Bundesamt. *Strukturvergleich von Mobilfunkdaten zweier Mobilfunkanbieter*. 2021a.

Statistisches Bundesamt. *Ausstattung mit Gebrauchsgütern – Daten aus den Laufenden Wirtschaftsrechnungen (LWR) zur Ausstattung privater Haushalte mit Informationstechnik*. 2021b.

Statistisches Bundesamt. *Experimentelle georeferenzierte Bevölkerungszahl auf Basis der Bevölkerungsfortschreibung und Mobilfunkdaten*. 2022a.

## LITERATURVERZEICHNIS

Statistisches Bundesamt. *Mobilitätsindikatoren auf Basis von Mobilfunkdaten*. 2022b.

Zhao, Ziliang/Shaw, Shih-Lung/Xu, Yang/Lu, Feng/Chen, Jie/Yin, Ling. *Understanding the bias of call detail records in human mobility researc*h. In: International Journal of Geographical Information Science. Jahrgang 30. Ausgabe 9/2016, Seite 1738 ff.

**Herausgeber**
Statistisches Bundesamt (Destatis), Wiesbaden

**Schriftleitung**
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

**Ihr Kontakt zu uns**
www.destatis.de/kontakt

**Erscheinungsfolge**
zweimonatlich, erschienen im Oktober 2022
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der Statistischen Bibliothek.

Artikelnummer: 1010200-22005-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2022
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.