



# Das Groß-N-Design

Anne-Kathrin Fischer und Achim Goerres

## Inhalt

1	Einleitung .....	128
2	Das Vorgehen bei und die Anwendung von Groß-N-Studien .....	129
3	Potenziale und Problemfelder von Groß-N-Studien .....	138
4	Kommentiertes Literaturverzeichnis .....	144
	Literatur .....	144

## Zusammenfassung

In diesem Beitrag werden die Logik und Anwendungsfelder von quantitativen Studien mit einer großen Fallzahl (Groß-N-Studien) in der Politikwissenschaft vorgestellt. Durch eine systematische Analyse mit möglichst vielen Fällen gilt es, ein Modell mit hoher Erklärungskraft, entweder für alle Variablenzusammenhänge oder für einen ausgewählten Teil der Variablen, zu finden. Dabei steht das Aufdecken und Verstehen der Gesamtheit der empirischen Muster in den vorliegenden Daten über alle Beobachtungen hinweg im Vordergrund, während die Analyse einzelner Fälle in den Hintergrund tritt. Nach einer Einführung in den Prozess und die Anwendung von Groß-N-Studien in der Politikwissenschaft werden den Leserinnen und Lesern die breiten Möglichkeiten sowie die Grenzen der Vorgehensweisen aufgezeigt.

## Schlüsselwörter

Groß-N-Studien · Soziale Mediendaten · Geodaten · Dateninfrastruktur · Internationale Beziehungen · Vergleichende Politikwissenschaft

Beide Autoren haben gleich viel zu dem Kapitel beigetragen. Die Auflistung folgt dem Alphabet. Wir danken Arndt Leininger und Claudius Wagemann für hilfreiche Kommentare.

A.-K. Fischer (✉) · A. Goerres (✉)  
Institut für Politikwissenschaft (IfP), Universität Duisburg-Essen, Duisburg, Deutschland  
E-Mail: [anne-kathrin.fischer@uni-due.de](mailto:anne-kathrin.fischer@uni-due.de); [achim.goerres@uni-due.de](mailto:achim.goerres@uni-due.de)

## 1 Einleitung

Dieser Beitrag stellt die Logik und die Praxis von quantitativen Studien mit einer großen Fallzahl in der Politikwissenschaft vor. Ein solcher Ansatz (im Englischen Large-N-Design) orientiert sich in seiner Logik an naturwissenschaftlichen Studien. Die Forscherin oder der Forscher intendiert durch die Nutzung einer maximalen Menge von Informationen, ein statistisches Modell zu finden, das möglichst viel (viele Daten) mit möglichst wenig (wenige Eigenschaften des statistischen Modells) erklärt. Das Ziel ist die Entwicklung von gesetzmäßigen Zusammenhängen durch die Abstrahierung aus dem Partikularen. Dieses Design wurde auch „nomothetisch“ getauft, wörtlich „Gesetz schaffend“, um diese Vorgehensweise von der „idiografischen“ Forschung, wörtlich „das Eigene beschreibend“, abzugrenzen (Windelband 2017 [1894]). Für eine Diskussion in der Politikwissenschaft siehe Berg-Schlusser und Stamm (2003 [1974]) (siehe auch den Beitrag von Berg-Schlusser in diesem Band).

Zusätzliche Daten werden, solange sie mit den anderen Daten vergleichbar sind, in diesem Design grundsätzlich immer positiv gesehen. Es kann nicht absolut definiert werden, wann ein N groß ist, weil es sich letztendlich um eine grundsätzliche Logik handelt, die umso wirksamer ist, je größer der Datensatz ist. Man könnte also dieselbe Logik sogar auf eine Handvoll Beobachtungen anwenden, was aber so viel Unsicherheit mit sich brächte, dass sich der Aufwand nicht lohnen würde. Trotzdem gibt es sie: faktische Klein-N-Studien mit einem Groß-N-Design. Beispielsweise basiert eine der meistzitierten Arbeiten zur Wahlbeteiligung im internationalen Vergleich auf der multivariaten statistischen Analyse von 19 Beobachtungen (Jackman 1987). Leider ist die Politikwissenschaft aufgrund ihres wichtigen Erkenntnisinteresses an Kollektiven häufig gezwungen, mit nur relativ kleinen Fallzahlen, z. B. Staaten in einer bestimmten Region, arbeiten zu können, wenn eigentlich ein Groß-N-Design angewandt wird. Es gibt also eine strukturell bedingte Tendenz in manchen Teilen der Politikwissenschaft, Gefahr zu laufen, mit zu wenig Daten in diesem Design zu viel zu wollen. In diesem Beitrag konzentrieren wir uns jedoch auf politikwissenschaftliche Studien vor allem aus den Bereichen der politischen Verhaltensforschung, Internationalen Beziehungen und Vergleichenden Politikwissenschaft, die durch eine besonders hohe Anzahl von Beobachtungen geprägt sind.

Ziel des Groß-N-Designs ist es, ein Modell zu finden, das so gut wie möglich zu den vorliegenden Daten passt. Es geht nicht darum, die Muster in den Daten einzelner Merkmalsträger gut zu erkennen, sondern die Gesamtheit der identifizierbaren Zusammenhänge über alle Beobachtungen möglichst gut zu verstehen und diese durch Variablenzusammenhänge zu beschreiben. Es interessiert also hier nicht, wie sich ein bestimmter Merkmalsträger von einem anderen unterscheidet, sondern wie diese Unterschiedlichkeit durch Variablenzusammenhänge systematisch beschrieben werden kann. In der Vergleichenden Politikwissenschaft wurde diese Strategie auf den Nenner gebracht, Ländernamen durch Variablen und ihre Ausprägungen zu ersetzen (Przeworski und Teune 1970).

Groß-N-Designs können sowohl in einer deduktiven wie auch in einer induktiven Logik eingebettet sein, d. h. es kann ein vorab formuliertes theoretisches Modell auf

seine Adäquatheit getestet werden, oder es werden aus den Daten Muster herausgearbeitet, die man anschließend für eine theoretische Modellbildung nutzt. Weiterhin kann das Design entweder Teil einer beobachtenden oder experimentellen Studie sein. Ausschlaggebend ist insgesamt die Suche nach einem bestmöglichen Modell über einen großen Datensatz hinweg.

In Abschn. 2.1 beschreiben wir die zentralen Grundelemente des Designs. Abschn. 2.2 gibt einen Überblick über unterschiedliche Datenarten und exemplarische Studien im Groß-N-Design. Abschn. 3.1 stellt die Potenziale des Designs in der Politikwissenschaft dar, bevor Abschn. 3.2 die Problemfelder diskutiert. In Abschn. 3.3 beschreiben wir aktuelle Entwicklungen und unsere Wahrnehmung der Zukunft dieses Designs in der Politikwissenschaft.

---

## **2 Das Vorgehen bei und die Anwendung von Groß-N-Studien**

### **2.1 Erläuterung zentraler Grundelemente**

Bei einer Groß-N-Studie liegt eine möglichst große Anzahl von Beobachtungen in einer theoretisch begründeten Population vor. Die Forscherin oder der Forscher definiert aufgrund von theoretischen, forschungsökonomischen und pragmatischen Überlegungen, was den Datensatz einer Studie definiert. Dabei können viele Beobachtungen von wenigen Merkmalsträgern stammen, wie es in Zeitreihenanalysen üblich ist, oder wenige Beobachtungen von einer Vielzahl von Merkmalsträgern. Doch grundsätzlich gilt: mehr Daten, solange vergleichbar, sind immer besser als weniger Daten.

Die Merkmalsträger in Groß-N-Studien können ganz unterschiedlicher Natur sein und auf verschiedenen Aggregationsebenen vorliegen: Menschen, Organisationen wie Parteien, Ereignisse, Austauschbeziehungen zwischen zwei Staaten und vieles mehr. Aufgrund der Natur der sozialen und politischen Realität sind Groß-N-Studien zu Individuen einfacher zu generieren als beispielsweise zu Staaten, weil bestimmte Merkmalsträger in der Politikwissenschaft einfach häufiger auftreten als andere. Dabei ist die unterschiedliche Homogenität von Merkmalsträgern, die man im Englischen *Unit Homogeneity* nennt, üblicherweise eine große Herausforderung. Daten von Kollektiven wie Parteien, Staaten oder Regionen sind in der Tendenz schwieriger zu vergleichen als einheitlich erhobene Daten von Individuen (Goerres und Prinzen 2012).

Die systematische Sammlung von Daten für eine Groß-N-Studie beinhaltet wie bei allen empirischen Studien zwei Entscheidungen: (1) Welche Art von Merkmalsträgern steht im Zentrum der Analyse? (2) Wie werden die Merkmalsträger ausgewählt, die in die Analyse einfließen? Die zweite Entscheidung wird bei Groß-N-Analysen so getroffen, dass möglichst viele vergleichbare Daten vorliegen, die eine sinnvolle Inferenz auf ein genau definiertes Universum zulassen. (siehe den Beitrag von Schnapp im vorliegenden Sammelband).

Die einfachste Vorgehensweise ist die Vollerhebung, also die Erhebung aller Beobachtungen in einem genau definierten Universum. Ein Beispiel dafür ist die Studie von Kittel und Obinger (2003), bei der alle 21 OECD-Staaten, für die Daten zwischen 1982 und 1997 verfügbar waren, in die Stichprobe einfließen. Die zweite wichtige Vorgehensweise ist die zufallsbasierte Auswahl, unabhängig davon, ob sie einfach oder komplex durchgeführt wird. Hier werden mit kontrolliertem Zufall aus einer genau definierten Population Elemente derart gezogen, dass unter Berücksichtigung der Regeln der Statistik statistische Inferenz über die Population möglich ist (am häufigsten verbreitet ist die klassische Inferenzstatistik, wobei die Bayes'sche Statistik auch zunehmend genutzt wird; siehe hierzu das Kapitel von Traummüller und Gill in diesem Sammelband).

Schließende Statistik wird typischerweise in der Politikwissenschaft auch bei Vollerhebungen angewendet. Obwohl die grundsätzliche Annahme einer Zufallserhebung nicht zutrifft, nutzen Forscherinnen und Forscher häufig klassische inferenzstatistische Tests (oder Bayesianische Methoden), um Unsicherheitsschätzungen über die gefundenen Zusammenhänge vorweisen zu können. Für eine Diskussion um die Bedingungen, unter denen diese Vorgehensweise durchgeführt werden sollte, siehe Behnke (2005) sowie Broscheid und Gschwend (2005).

Eine theoretisch begründete Auswahl von Elementen kommt in der Groß-N-Analyse nur nachrangig vor. So ist es vorstellbar, in einem mehrschrittigen Auswahlverfahren auf höheren Aggregationsebenen die Kontexte der Elemente theoretisch auszuwählen, um dann innerhalb dieser Auswahl wieder voll oder nach dem Zufallsverfahren zu erheben. So wählen Andreß und Heien (2001) beispielsweise vier verschiedene wohlfahrtsstaatliche Regimes aus, die jeweils durch ein Land oder einen Landesteil repräsentiert werden, um dann innerhalb dieser Regimes mit zufallsbasierten Umfragestichproben zu arbeiten.

Alle Daten werden in numerische Informationen übersetzt und sind somit in einer Matrix oder einer Vielzahl von Matrizen darstellbar. Dabei werden auch ursprünglich nicht-numerische Informationen wie beispielsweise von Befragten offen genannte politische Probleme in eine nominale Variable transferiert, deren Ausprägungen eindeutig Zahlen zugeordnet bekommen. Eine Festlegung auf das metrische Skalenniveau für die Variablen ist entgegen der oftmals geäußerten Meinung nicht notwendig.

Nach dieser Überführung in numerische Informationen analysiert man die Datenmatrix ausschließlich mit statistischen Verfahren. Diese Verfahren können einfach sein wie in der univariaten Statistik (z. B. die Berechnung von Anteilen oder Mittelwerten) oder in der bivariaten Statistik (z. B. die Berechnung von Assoziationsgrößen zwischen zwei Variablen), oder multivariat wie Regressionsanalysen, Strukturgleichungsmodelle und andere Methoden (siehe hierzu auch die Beiträge von Ziller, Pötschke und Berning in diesem Band). Ziel der statistischen Analyse ist die Aufdeckung systematischer Beziehungen, um generalisierende Aussagen über die vorliegenden Daten, aber zumeist auch über die dahinterliegende Population zu treffen. Dabei sind inferenzstatistische Analysen in einem Groß-N-Design der klassischen frequentistischen Tradition genauso denkbar wie der bayesianischen Tradition. Der Unterschied ist lediglich, dass die frequentistische Tradition hauptsächlich

von Zufallsstichproben ausgeht, und die bayesianische Tradition die Daten zu einer Verbesserung einer Wahrscheinlichkeitsverteilung über den Parameter nutzt.

Bezüglich der Suche nach Kausalität folgen Groß-N-Studien entweder einem experimentellen Design wie z. B. bei der Facebook-Studie (siehe Abschn. 2.2.2) mit der bewussten Manipulation hypothetisierter Ursache-Wirkungszusammenhänge oder der in-der-Natur-vorkommenden Als-ob-Zuteilung im natürlichen Experiment. Alternativ folgen sie einem beobachtenden Design, in dem entweder im Nachhinein Überlegungen zu einem Quasi-Experiment angestellt werden oder durch eine geschickte Berücksichtigung von Drittvariablen eine möglichst große Kontrolle über alternative Erklärungsgrößen erzielt wird.

Grundsätzlich unterstellt der Groß-N-Ansatz, dass es möglich ist, ein relativ bestes Modell für die vorliegenden Daten zu finden. Dieses relativ beste Modell beinhaltet entweder ein kausales Modell für alle Elemente aus einer großen Gruppe wie z. B. im Standard-Regressionsansatz oder für die Zugehörigkeit aller Elemente in unterschiedliche Gruppen, innerhalb derer unterschiedliche kausale Konfigurationen vorliegen. Im letzteren Fall muss die kausale Erklärung dazu beitragen, sagen zu können, welche Merkmalsträger zu welcher Gruppe gehören und wie die unterschiedlichen gruppenspezifischen kausalen Untermodelle konfiguriert sind. So wird beispielsweise in der modernen Wahlforschung zunehmend von heterogenen Wahlmodellen ausgegangen (Bartle 2005), d. h. Modellen, in denen Gruppen von Wählern unterschiedlichen kausalen Dynamiken unterliegen. Die Modellierung solcher Heterogenität erfolgt entweder über eine Interaktionsanalyse (Heterogenität im kleineren Umfang) oder über Strukturgleichungsmodelle.

## 2.2 Anwendungsfelder von Groß-N-Studien in der Politikwissenschaft

Die Anwendungsfelder von Groß-N-Studien in der Politikwissenschaft sind vielfältig, da den Studien verschiedenste Datentypen zugrunde liegen können. Dazu zählen klassische Individualdatenquerschnitte, Aggregatdatenquerschnitte und Paneldaten. Bei dem erstgenannten Datentyp sind Individuen die Merkmalsträger, deren Merkmale zu einem gegebenen Zeitpunkt erhoben wurden. Ändert sich die Beobachtungseinheit vom Individuum auf einen aggregierten Merkmalsträger wie beispielsweise auf den eines Landes, spricht man von einem Aggregatdatenquerschnitt. Als klassische Datensätze lassen sich u. a. der internationale World Value Survey (WVS) oder auch die German Longitudinal Election Study (GLES) anführen. GLES bietet sowohl Querschnitte von Individual- als auch Aggregatdaten: So können mit diesem Datensatz gleichermaßen individuelle Daten zum Wahlverhalten als auch Merkmale ganzer Wahlkreise untersucht werden. Werden Daten für eine oder mehrere Variable(n) nicht nur für einen gegebenen Zeitpunkt, sondern über mehrere Beobachtungspunkte über die Zeit hinweg erhoben, handelt es sich um gepoolte Querschnitts- oder Paneldaten. Während bei gepoolten Querschnitten Daten von verschiedenen Merkmalsträgern für mehrere Zeitpunkte erhoben werden (beispielsweise beim WVS und GLES), werden bei Paneldesigns Daten von glei-

chen Merkmalsträgern für mehrere Zeitpunkte erfasst. Das vom Deutschen Institut für Wirtschaftsforschung durchgeführte Sozio-ökonomische Panel (SOEP) liefert solche Paneldaten für Deutschland.

Groß-N-Studien in der Politikwissenschaft unterscheiden sich neben den angeführten Datenarten auch nach Datenzugang. Neben klassischen Datenquellen der amtlichen Statistik (siehe u. a. Daten der Statistischen Bundes- und Landesämter für den deutschen oder Eurostat für den europäischen Raum) sowie Befragungen (siehe u. a. SOEP), kommen ebenfalls die nichtamtliche Statistik oder groß angelegte Experimente als Datenquellen für Groß-N-Studien infrage. Zur nichtamtlichen Statistik gehören Daten, die beispielsweise von der OECD, UNO oder auch der Weltbank zur Verfügung gestellt werden.

Im Folgenden bieten wir einen Überblick über verschiedene Arten von Groß-N-Studien mit Beispielen, die wir anhand der Datenarten strukturiert haben: international vergleichende Umfragedaten, soziale Mediendaten, Zeit-Länder-Daten und Geodaten. Die vorgestellten Beispiele stellen natürlich nur eine Auswahl dar, die wir aufgrund ihrer Unterschiedlichkeit vorgenommen haben. Die Leserin oder der Leser soll hierin die breiten Möglichkeiten eines quantitativen Groß-N-Designs in der Politikwissenschaft ersehen.

### **2.2.1 Umfragedaten ländervergleichend**

Internationale Umfragedatensätze sind üblicherweise das Ergebnis einer komplexen, teuren und langjährigen Zusammenarbeit von Hunderten von Sozialwissenschaftlern. Pro Länderzeitpunkt gibt es eine länderspezifische Stichprobenziehung, die national umgesetzt und zentral koordiniert werden muss. Typische Stichprobengrößen pro Länderzeitpunkt betragen 1000 realisierte Interviews, so dass große internationale Umfragedatensätze wie der European Social Survey (ESS) schnell über 30.000 Beobachtungen umfassen. Querschnittsbefragungen dominieren die politikwissenschaftlichen Studien in diesem Kontext. Harmonisierte Panelbefragungen, wie es sie in der Soziologie und den Wirtschaftswissenschaften gibt (beispielsweise EU-SILC), fehlen bisher in der Politikwissenschaft. Diese internationalen Umfragedatensätze erlauben die Analyse von inter-individuellen Unterschieden, von Unterschieden zwischen nationalen oder regionalen Kontexten im Querschnitt und bei einer Kumulation auch von Kontexten im Längsschnitt. Die Informationen der Individuen aus einem Land sind typischerweise unabhängig voneinander und erlauben eine einfache statistische Auswertung. Bei einer Analyse von Umfragedaten aus mehreren Ländern in solch einer harmonisierten Studie müssen entsprechende technische Anpassungen geprüft und gegebenenfalls getroffen werden (siehe hierzu den Beitrag von Ziller in diesem Band).

Professionelle Infrastrukturen durch nationale und internationale Datenarchive erlauben die multiple Nutzung dieser Datensätze. Typischerweise reichen eine Anbindung an eine wissenschaftliche Einrichtung und ein Internetzugang aus, um dieser Daten habhaft zu werden.

*Bayram (2015): „What Drives Modern Diogenes? Individual Values and Cosmopolitan Identity“*

Große Datensätze beruhend auf Umfragedaten sind unter anderem im Forschungsfeld der international vergleichenden Politikwissenschaft zu finden. Beispielhaft lässt sich die Studie von Bayram (2015) heranziehen, bei der auf Grundlage des World Value Surveys (WVS) ländervergleichend Einstellungen der Bürger untersucht werden. Die Studie prüft, welche Werte (wie Moral, Diversität, Freiheit, Konservatismus und auch Selbstinteresse) das Selbstbild als Weltbürger fördern oder behindern. Die Befragten des WVS (Welle von 2005–2008) stellen die Untersuchungseinheiten dar und machen ein N von fast 25.000 Fällen aus. Mit der Nutzung des WVS wird eine Datengrundlage gewählt, die es aufgrund ihrer umfangreichen und weiträumigen Befragungen ermöglicht, Aussagen zum Status sowie Wandel menschlicher Werte in verschiedensten Ländern zu treffen. Bayram kann in dieser Studie damit Einstellungen in sehr unterschiedlichen Gesellschaften und nationalen Kontexten vergleichend analysieren.

*Shayo (2009): „A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution“*

Als Beispiel für ländervergleichende Umfragestudien mit zusätzlichen Kontextdaten lässt sich die Untersuchung von Shayo (2009) anführen, bei der die Bedeutung von sozialer Identität für Einstellungen zur Einkommensverteilung eines Landes beleuchtet wird. Die Mikro-Daten zur Erfassung politischer Präferenzen und politischer Identifikation stammen aus dem International Social Survey Program (ISSP; 1995er Welle zur nationalen Identität) sowie den Wellen 1–3 des WVS. Die Daten auf der Makro-Ebene haben ihren Ursprung in der Luxembourg Income Study (LIS), die Auskunft über das Ausmaß der Umverteilungsmaßnahmen der untersuchten Länder gibt. Mit dieser Datenbasis kann die Analyse für mehr als 20 demokratische Staaten in den 1990er-Jahren durchgeführt werden. Shayo schafft es damit, Interaktionseffekte zwischen nationaler und klassenspezifischer Identifikation, Einkommensungleichheit und politischen Präferenzen über Länder und Individuen hinweg zu vergleichen. Damit zeigt sie u. a., dass sich Länder mit bereits etablierten Demokratien von erst kürzlich demokratisierten Staaten hinsichtlich der nationalen Identität ihrer Bürger sowie der politischen Präferenzen in Bezug auf staatliche Umverteilung unterscheiden.

### 2.2.2 Soziale Mediendaten

Soziale Mediendaten entstehen in kommerziellen Plattformen, die es Nutzerinnen und Nutzern ermöglichen, nicht nur Inhalte zu konsumieren, sondern auch direkt (meistens schriftlich) auf diese zu reagieren. Im Jahr 2017 waren die am häufigsten gebrauchten Plattformen Facebook und Twitter (Facebook verweist auf 2,01 Milliarden (Facebook Inc 2017) und Twitter auf 328 Millionen aktive Nutzer pro Monat (Twitter Inc 2017)).

Durch die hohe Anzahl an Nutzerinnen und Nutzern und damit der Interaktionen entstehen so ungeheure Mengen an Kommunikationsdaten zwischen Individuen, Gruppen und Organisationen. Zudem haben diese Daten im Vergleich zu den künstlichen Befragungssituationen der Umfrageforschung den Vorteil, dass sie tatsächli-

ches Verhalten abbilden. Jedoch gibt es auch ethische sowie rechtliche Bedenken. Bei dieser Art von Daten stellt sich erst einmal die rechtliche Frage, wem sie gehören. Sie gehen einerseits auf kommerzielle Plattformen zurück, deren Ziel Gewinnmaximierung ist. Andererseits schlagen sie sich in der öffentlichen Sphäre nieder und sind – zum großen Teil – jeder Nutzerin und jedem Nutzer zugänglich (vgl. für eine Diskussion über Twitter-Daten Puschmann und Burgess 2014). Soziale Medienfirmen stellen häufig, aber nicht immer, einen kleinen Teil ihrer Daten über eine API-Schnittstelle zur Verfügung (*Application Programming Interface*), wobei unklar ist, nach welchen Kriterien diese Auswahl zur Verfügung gestellt wird. Diese Schnittstelle erlaubt es, die Daten für wissenschaftliche Groß-N-Analysen aufzubereiten. Jedoch gilt es hierbei zu berücksichtigen, dass die Mechanismen, nach denen die Daten zur Verfügung gestellt werden, meist unbekannt sind. Von einer Vollerhebung aller Nutzer der Medienplattform kann deshalb nicht ohne Weiteres ausgegangen werden.

Ein Nachteil von sozialen Mediendaten besteht in ihrem Repräsentationsproblem. Nicht hinter jeder Nutzerin bzw. jedem Nutzer steht zwingendermaßen ein einzelner Mensch. Zum einen kann ein Individuum mehrere Accounts gleichzeitig führen. Zum anderem besteht die Gefahr, dass sich hinter einem angeblichem Nutzer ein Bot-Account verbirgt (Boyd und Crawford 2012, S. 669). Neben möglichen Problemen mit den Merkmalsträgern können zudem Einschränkungen bei der Erfassung der eigentlichen Merkmale bestehen. Die Datengrundlage kann verzerrt sein, da durch die Betreiber sozialer Medienseiten wie Twitter oder Facebook Posts mit problematischem Inhalt gelöscht werden.

*Colleoni et al. (2014): „Echo Chamber or Public Sphere: Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data“*

Der Artikel von Colleoni et al. (2014) analysiert politische Kommunikation im Internet und geht der Frage nach, ob sich auch online das Phänomen der politischen Homophilie finden lässt. Politische Homophilie beschreibt die Tendenz, dass politisch ähnlich gesinnte Individuen eher Bindungen mit Gleichgesinnten suchen. Für ihre Untersuchung in den USA greifen die Autoren auf zwei bestehende Datensets (*Twitter Content* und *Twitter Graph*)<sup>1</sup> zurück. Um diese große Masse an Daten auszuwerten, nutzen die Autoren bei der Textanalyse ein induktives Machine Learning-Verfahren. Der politische Inhalt der Tweets wird so auf automatisierte Weise demokratischen oder republikanischen Diskursen zugeordnet. Die Untersuchung zeigt, dass die verschiedenen Formen politischer Partizipation auch online in sozialen Netzwerken wie Twitter existieren. Zudem wirft die Studie bereits erste kritische Fragen im Umgang mit Twitter-Daten auf – so nämlich die Frage, ob

---

<sup>1</sup>Twitter Content wurde von Yang und Leskovec (2011) zusammengestellt und umfasst ein repräsentatives Sample von 467 Millionen Tweets von 20 Millionen Usern für einen Zeitraum von sieben Monaten (Juni bis Dezember 2009). Twitter Graph wurde von Kwak et al. (2010) erstellt und zeigt alle Nutzer als Knotenpunkte ihrer Verbindungen mit anderen Nutzern von Twitter im Jahre 2009. Diese Datenbank umfasst 1,47 Milliarden Verbindungen.



Twitter als ein soziales Medium betrachtet und analysiert werden sollte oder als eine Form eines Nachrichtenmediums.

*Bond et al. (2012): „A 61-million-person experiment in social influence and political mobilization“*

Die wachsende Verfügbarkeit von kostengünstigen und groß angelegten Online-Social-Network-Daten ermöglicht, dass Experimente leichter im Feld durchgeführt werden können. Als ein solches groß angelegtes Forschungsdesign lässt sich das Facebook-Experiment von Bond et al. (2012) anführen. In ihrem Forschungsvorhaben mit über 61 Millionen Facebook-Nutzern untersuchen sie, ob die bewusste Wahlmobilisierung über diese Online-Plattform während der Kongresswahlen 2010 in den USA zu einer höheren Wahlbeteiligung führt. Dafür führen die Forscher ein randomisiertes kontrolliertes Experiment zur Wählermobilisierung bei Facebook-Nutzern durch. Es zeigt, dass politische Online-Mobilisierung tatsächlich zu einer höheren Wahlbeteiligung führt. Dieses größte publizierte Experiment der Politikwissenschaft bis August 2017 besticht durch seine sehr hohe Validität, seine sehr präzise Schätzung aufgrund der großen Fallzahl und seine beispiellose Zusammenarbeit zwischen Unternehmen und Wissenschaftlern. Gleichzeitig kann man an dieser Studie auch die ethischen Schwierigkeiten sehen, wenn Wissenschaftler über soziale Medien direkt in Wahlen eingreifen: Menschen werden getäuscht, da sie keine Möglichkeit haben, nicht mitzumachen, und da sie bezüglich ihres realen Verhaltens manipuliert werden.

### 2.2.3 Zeit-Länder-Daten

Wenden wir uns nun den Möglichkeiten von Groß-N-Studien jenseits der Individualdaten zu. Die Internationalen Beziehungen und die Vergleichende Politikwissenschaft beschäftigen sich häufig mit Erklärungen für das Verhalten von Regierungen, Staaten oder anderen nationalen Entitäten. Diese Ansätze laufen Gefahr, mit zu wenigen Merkmalsträgern zu arbeiten, deren Informationen in eine Analyse fließen können. Im Jahr 2017 waren selbst bei Einbeziehung von Territorien nicht allgemein anerkannter Staaten weniger als 240 Entitäten existent (Central Intelligence Agency 2017). Die Anzahl der unabhängigen Merkmalsträger ist gering; zudem kann je nach Forschungsgegenstand auch deren Unabhängigkeit beispielsweise aufgrund von Policy-Diffusion angezweifelt werden. Damit kann die Logik für eine Groß-N quantitative Analyse nur bis zu einem gewissen Grad angewendet werden.

Das Problem dieser geringen Anzahl von Staaten kann umschifft werden. Zum einen können die Merkmalsträger anders definiert werden, wenn man sich zum Beispiel für Zweier-Beziehungen zwischen Staaten interessiert, sogenannte Dyaden. Zum anderen können wiederholte Beobachtungen desselben Staates über Zeit die Anzahl der Beobachtungen deutlich erhöhen.

*Lundsgaarde et al. (2010): „Instrumental Philanthropy: Trade and the Allocation of Foreign Aid“*

Als ein Beispiel für Groß-N-Studien in den Internationalen Beziehungen lässt sich die Studie von Lundsgaarde et al. (2010) anführen, bei der es um Entwicklungshilfeströme zwischen Geber- und Empfängerländern geht. Die Autoren gehen der Frage nach, welche Systematik hinter dem Handel und der Zuteilung von Entwicklungshilfen in bilateralen Beziehungen in der Praxis vorherrscht. Um die Entwicklungshilfeströme im Zeitraum von 1980 bis 2002 von 22 Geberländern in 187 Empfängerländer zu untersuchen, greifen die Autoren auf Daten der OECD zurück. Im Bereich für Entwicklungszusammenarbeit stellt das sog. „Development Assistance Committee“ (DAC) der OECD u. a. Statistiken über die Entwicklungshilfeströme zur Verfügung. Mit diesen Daten kann die Studie auf ein N von 34.856 zurückgreifen. Die Analyse zeigt, dass arme Länder, die ohnehin aufgrund fehlender handelbarer Ressourcen bereits beim internationalen Handel benachteiligt sind, zusätzlich auch weniger fremde Investitionshilfen erhalten. Damit verstärken sich die zuvor bestehenden strukturellen Ungleichheiten der Entwicklungsländer.

*Dafoe et al. (2013): „The Democratic Peace: Weighing the Evidence and Cautious Inference“*

Als ein weiteres Beispiel aus dem Forschungsfeld der Internationalen Beziehungen lässt sich die Studie von Dafoe et al. (2013) anführen, die sich mit einer zentralen Erkenntnis der Friedens- und Konfliktforschung auseinandersetzen. Dieser zufolge ist Frieden zwischen demokratischen Staaten stabiler als zwischen demokratischen und autokratischen Staaten. Zur Überprüfung dieser Erkenntnis nehmen die Forscher eine Zeitreihenanalyse der Daten von 1816 bis 2001 vor und erreichen aufgrund der dyadischen Datenstruktur (d. h. Krieg oder Frieden zwischen zwei Staaten) ein sehr großes N von ~500.000. Die Autoren greifen dafür auf eine Datenbank zurück, die Informationen über Konflikte zwischen verschiedenen Staaten zusammenträgt. Diese sog. Militarized Interstate Dispute (MID) Datensammlung umfasst entsprechende Informationen von 1816 bis 2010 und ist frei zugänglich.<sup>2</sup> Auf Grundlage dieser Daten kommen die Autoren zu dem Ergebnis, dass der Frieden bei demokratischen Dyaden immer noch am stabilsten ist. Auch zeigt sich, dass autokratische Paare untereinander friedfertiger sind als Dyaden von demokratischen und autokratischen Staaten.

## 2.2.4 Geodaten

Als letztes Beispiel für Groß-N-Studien in der Politikwissenschaft stellen wir Studien vor, die auf geografischen Daten beruhen. Hier werden geografische Einheiten, z. B. Flächeneinheiten oder Wohnadressen zugrunde gelegt, um Daten zu erheben (siehe hierzu ausführlich den Beitrag von Weidmann und Gleditsch in diesem Band). Neuere Datenquellen bestehen aus Satellitendaten, digitalisierten kartografischen Daten sowie einem zunehmenden Fundus an Daten aus dem Mikromarketing, das für die Forschung nutzbar gemacht wird wie z. B. auf Adressebene vorliegende Milieuvariablen. Streng genommen sind Geodaten Individualdaten, die aber häufig in aggregierter Form genutzt werden.

<sup>2</sup><http://www.correlatesofwar.org/news/mid-data-set-v4-01-available>.

*Weidmann and Schutte (2017): „Using night light emissions for the prediction of local wealth“*

Oftmals sind gerade für Entwicklungsländer die gewünschten Daten nicht verfügbar oder lediglich von eingeschränkter Qualität. Eine Möglichkeit, dennoch Forschung zu betreiben, liegt in der Nutzung von Proxys anstelle der eigentlichen Variable. Als ein Beispiel lässt sich die Studie von Weidmann und Schutte (2017) anführen, die die Nachtbeleuchtung als einen Proxy für Wohlstand in den Entwicklungsländern nutzen. Für die Analysen greifen die Autoren auf zwei Datensätze zurück. Bei dem ersten handelt es sich um den amerikanischen Archivdatensatz der DMSP-OLS,<sup>3</sup> der Satellitenaufnahmen zur Erfassung der Lichtemissionen der Straßenbeleuchtung enthält. Zudem nutzen die Autoren Daten des Demographic and Health Surveys (DHS),<sup>4</sup> welche durch den Rückgriff auf über 300 Umfragen in über 90 Ländern repräsentative Bevölkerungsdaten liefern. Mit Variablen aus diesen beiden Datensätzen können die Autoren ihre Untersuchung über 39 weniger entwickelte, meist nicht-demokratische Länder mit einer Stichprobe von 34.000 Beobachtungen durchführen. Sie können zeigen, dass Lichtemissionen als Prädiktor für die Schätzung des ökonomischen Wohlstands von Ländern genutzt werden kann. Dieser Proxy stellt damit einen adäquaten Prädiktor für die ökonomische Reichtumsschätzung von Regionen in Entwicklungsländern und somit einen Mehrwert für die Erforschung von Entwicklungsländern dar.

*Haspel and Knotts (2005): „Location, location, location: Precinct placement and the costs of voting“*

Die Studie von Haspel und Knotts (2005) weicht von vielen früheren Studien zur Wahlforschung in zwei wichtigen Punkten ab. Zunächst verwenden die Autoren Verwaltungs- statt Umfragedaten, die ihnen Auskunft über das tatsächliche Verhalten und die geografische Lage einzelner Personen liefern. Zudem machen sie sich die technologische Verbesserung im Bereich der Geoinformationssysteme (GIS) zunutze, mit denen räumliche Daten erfasst, bearbeitet, organisiert und analysiert werden können. Sie kombinieren dieses Werkzeug mit Daten aus dem Wählerverzeichnis der Stadt Atlanta, in dem neben Namen und Anschrift zudem soziodemografische Informationen sowie – im Gegensatz zu deutschen Wählerverzeichnissen – ebenfalls Informationen zum bisherigen Wahlverhalten enthalten sind. Für jede Adresse eines Wahlberechtigten in Atlanta für die Bürgermeisterwahl 2001 wird mithilfe der Capiler's Maptitude Mapping Software<sup>5</sup> die Entfernung (als Wegdistanz über die Straße) zum jeweiligen Wahllokal errechnet. Insgesamt können so Daten von über 198.000 Wahlberechtigten für die Analysen nutzbar gemacht werden.

---

<sup>3</sup>Der DMSP-OLS (Defense Meteorological Satellite Program's Operational Linescan Systems) wird von der US National Oceanic and Atmospheric Administration zur Verfügung gestellt. Siehe <https://data.noaa.gov/dataset/dmsp-ols-operational-linescan-system>.

<sup>4</sup><http://dhsprogram.com>.

<sup>5</sup>Die GIS Software Maptitude ist ein geographisches Analysesystem, das auch für wissenschaftliche Zwecke genutzt werden kann, um die eigenen Daten kartographisch darstellen zu können. <http://www.caliper.com/maptovu.htm>.

Mithilfe der geografischen Daten zeigen die Autoren, dass bereits kleine Unterschiede in der Entfernung vom Wahllokal zum eigenen Wohnort einen erheblichen Einfluss auf die Wahlbeteiligung haben können.

---

## 3 Potenziale und Problemfelder von Groß-N-Studien

### 3.1 Potenziale

Groß-N-Studien werden mit der Intention durchgeführt, alle relevanten und verfügbaren Informationen von möglichst vielen Merkmalsträgern in einer Analyse zu berücksichtigen. Dadurch, dass das relativ beste statistische Modell für eine große Datenmenge gefunden wird, haben die Ergebnisse für die vorliegenden Daten eine große Generalisierbarkeit. Darüber hinaus versuchen die Forscherinnen und Forscher über breitere theoretische Universen Aussagen zu treffen; d. h. das theoretische Universum geht typischerweise über das empirische hinaus.

Aufgrund des rapiden Wachstums in der digitalen Aufbereitung von Daten und den besseren allgemein zugänglichen Archiven war es niemals so leicht, eine Groß-N-Studie in der Politikwissenschaft durchzuführen, wie jetzt. Die Kosten pro Merkmalsträger in einer Studie, so können wir spekulieren, sind im Laufe der Zeit deutlich gesunken. Besonders ausgeprägt ist die Möglichkeit einer Groß-N-Studie in der Verarbeitung neuer sozialer Mediendaten. Hier können Forscherinnen und Forscher bereits mit wenigen Ressourcen sehr große Datenmengen verarbeiten und einer quantitativen Groß-N-Studie zuführen (siehe hierzu den Beitrag von Munzert und Nyhuis in diesem Band).

Quantitative Groß-N-Studien erlauben weiterhin eine sehr intuitive Verknüpfung mit qualitativen Analyseabschnitten als Teil einer durchdachten mixed-oder multi-method-Strategie (siehe den Beitrag von Maggetti in diesem Band). Beispielsweise kann eine Regressionsanalyse einer Groß-N-Studie genutzt werden, um einzelne Merkmalsträger oder Gruppen von Merkmalsträgern zu identifizieren, die nur schlecht durch das statistische Modell erklärt werden können. Dann kann eine Analyse dieser Merkmalsträger (*deviant cases* genannt) das theoretische Modell verbessern. Idealerweise kann eine gute Fallstudie die Schwächen einer Groß-N-Studie ausgleichen und umgekehrt (Fearon und Laitin 2008, S. 758; Lieberman 2005, im Kontrast dazu Rohlfing 2008). Mit einem solchen Vorgehen kann man auch auf die Kritik antworten, dass in einer Groß-N-Studie niemals Aussagen über die kausalen Mechanismen bei einzelnen Merkmalsträgern getroffen werden können.

Mit ihrem 1994 erschienen Lehrbuch „Designing Social Inquiry“ gehen King, Keohane und Verba noch einen Schritt weiter. Sie versuchen, grundlegende Konzepte des quantitativen Designs auch für Klein-N-Studien fruchtbar zu machen. Beispielsweise diskutieren sie Probleme von Messfehlern, kausaler Inferenz, Endogenität, Unsicherheit in Aussagen, Unbestimmtheit eines Forschungsdesigns, usw. Für eine Reihe von Kritikpunkten an King, Keohane und Verbas Ausführungen seien die Leserinnen und Leser auf Brady und Collier (2010) verwiesen.

## 3.2 Problemfelder

Groß-N-Studien können potenziell an fünf Dingen krankem: der Abhängigkeit von früheren Datensammlungen in Bezug auf (a) Datenqualität, (b) Variablenverfügbarkeit und (c) Stichprobenkomposition, (d) der mangelnden Vergleichbarkeit der Merkmalsträger und (e) der Scheinobjektivierung eines Forschungsdesigns. Siehe dazu auch einen sehr kernigen Überblicksartikel über häufige Fehler (Schrodt 2014).

Die wenigsten Groß-N-Studien beruhen auf einer Primärdatenanalyse. Die Nutzung von Sekundärdaten ist vermutlich die Norm, d. h. Forscherinnen und Forscher nutzen bereits (digital) abgelegte Daten für ihre eigene Forschung. Dies ist aus Sicht des wissenschaftlichen Fortschritts auch wünschenswert, da Fortschritt fast nur inkrementell erfolgen kann. Insofern ist eine systematische Nutzung bereits vorhandenen Wissens und vorliegender Daten sinnvoll. Mit diesem Ansatz gehen allerdings drei Probleme einher. Ein erstes Problem sind Messprobleme (a). Der Nutzer muss letztendlich darauf vertrauen, dass die Messfehler über alle Merkmalsträger im Schnitt null sind, oder versuchen, diesen Messfehler in einem komplexen Messmodell mit zu modellieren (daraus ergibt sich aber die Notwendigkeit weiterer Variablen). Eine explizite Nachprüfung der Messung für jeden Merkmalsträger ist allein aus ökonomischer Sicht unmöglich. Die eindeutige Darstellung numerischer Daten in der Datenmatrix kann unvorsichtige Forscherinnen und Forscher dazu verleiten, eine Präzision in den Daten zu sehen, die hier gar nicht vorliegt. Anekdotisch haben wir beobachtet, dass Surveyforscherinnen und -forscher, die das erste Mal selbst eine Umfrage erheben, nachdem sie vorher nur als Sekundäranalysten gearbeitet haben, wesentlich kritischer bezüglich Validität, Objektivität und Reliabilität der Daten sind als vorher. Grundsätzlich muss man sich bei einer Groß-N-Analyse von Sekundärdaten die scheinbar banale Frage stellen, wie die Zahlen in die Matrix gekommen sind.

Das Groß-N-Design hat seinen wissenschaftstheoretischen Ursprung in den Naturwissenschaften. Diese sind davon gekennzeichnet, dass sich viele Dinge direkt messen lassen, also einfach als manifeste Variablen zu operationalisieren und zu messen sind. Messfehler auf der Konstruktebene gibt es dadurch seltener als in der Politikwissenschaft. Die Politikwissenschaft hat als Erkenntnisinteresse viele Konzepte, die man nicht einfach direkt messen kann, wie z. B. Demokratiequalität oder politische Xenophobie. Es gibt keinen offensichtlichen Weg festzustellen, wie demokratisch ein Land ist, oder wie verbreitet politisch relevante Abneigung gegenüber Fremden ist. Die Wirtschaftswissenschaft als Nachbardisziplin der Politikwissenschaft ist dahingehend schon besser aufgestellt, als dass sich viele ihrer Konstrukte durch monetäre Variablen direkt messen lassen. Die Psychologie als weitere Nachbardisziplin hat wiederum den Umgang mit latenten Variablen professionalisiert, indem in Strukturgleichungsmodellen Mess- und Kausalmodelle explizit modelliert werden (siehe den Beitrag von Berning in diesem Band). Die Psychologie profitiert jedoch davon, dass ihre Merkmalsträger fast ausschließlich Individuen in sehr großer Anzahl sind, anders als beispielsweise in der vergleichenden Staatsforschung in der Politikwissenschaft. So ist das Groß-N-Design in der Politikwissenschaft in großen Teilen der Disziplin bei der Messung latenter Variablen durch geringe Falluniversen gekennzeichnet. Vergleiche Paxton (2002) für eine der seltenen internationalen verglei-

chenden Makro-Studien mit einer expliziten Messung latenter Variablen über die interdependente Beziehung zwischen Sozialkapital und Demokratie.

Wegen dieser Gefahr der unbewussten Weiterbenutzung von mit Messfehlern behafteten Daten ist eines der Desiderata in der Groß-N-Forschung eine stetige Auseinandersetzung und Verbesserung bestehender Datensätze. Hier gibt es in der Politikwissenschaft gerade im Bereich der Messung von Parteipositionen eine rege Diskussion und eine stetige Verbesserung durch alternative Messungen (beispielsweise Manifestdaten, Experteneinschätzungen, Medienberichte oder Wählereinschätzungen) (Dinas und Gemenis 2010; Volkens 2007).

Weiterhin (b) verleitet die Sekundärnutzung von vorliegenden Variablen dazu, das eigene theoretische Konzept so weit anzupassen, dass es mit den vorliegenden Indikatoren gemessen werden kann. Neben einer inhaltlichen Anpassung kann es dabei zu einer bewussten Reduktion der Messgenauigkeit kommen, indem vielleicht ein metrisches Phänomen nur binär gemessen wird. Beispielsweise benutzt Rabuza (2016) einen einfachen Indikator bezüglich des Vorkommens von Civic Education in der Schulausbildung im Ländervergleich, obwohl sein theoretisches Konzept eigentlich die Inhalte von Civic Education und die Intensität der Ausbildung messen soll. Auch hier ist eine stetige Verbesserung politikwissenschaftlicher Daten notwendig und möglich: So wurde die Validität von manchen wohlfahrtsstaatlichen Indikatoren für politikwissenschaftliche theoretische Konzepte durch das Comparative Welfare Entitlement Dataset erhöht (Scruggs et al. 2017).

Ferner bestimmt (c) die Datenverfügbarkeit gerade in der ländervergleichenden Forschung häufig die Stichprobenzusammensetzung. So ist ein Großteil der vergleichenden Wohlfahrtsstaatsforschung reine OECD-Welt-Forschung, weil die OECD qualitativ hochwertige Statistikerarbeit leistet. Welche Länder allerdings Teil der OECD sind, ist eine politische Entscheidung, die somit implizit die Forschungsergebnisse beeinflusst (Kittel 2006). So gehören die meisten OECD-Staaten zu den reichsten der Welt, aber auch hier gibt es Ausnahmen wie Mexiko. Da davon auszugehen ist, dass wirtschaftlich-politische Faktoren bestimmen, welche Regierungen ihr Land in die OECD-Gruppe bringen wollen, und welche Länder von der OECD aufgenommen werden, ist in Studien mit OECD-Daten die Zusammensetzung der Stichprobe im Vergleich zu einer hypothetischen Zufallsstichprobe aller Länder verzerrt. Darüber hinaus ist die Datenverfügbarkeit im Zeitverlauf begrenzt, so dass die Stichprobe durch die Analysierbarkeit bestimmter Zeitperioden weiter atheoretisch definiert wird.

Die drei hier beschriebenen Probleme der Sekundäranalyse, Datenqualität, Variablenkomposition und Stichprobenzusammensetzung, finden sich interessanterweise auch bei Big-Data-Ansätzen. Gerade in der Nutzung von Internetdaten besteht die Gefahr, dass die sozial verzerrte Produktion dieser Daten (beispielsweise bei Twitter) eine mangelnde Passgüte zu theoretischen Konzepten und Messfehler durch ein selektives Eingreifen der kommerziellen Firmen (beispielsweise durch Zensur) die gleichen Datenprobleme verursachen (Boyd und Crawford 2012).

Ein viertes Problem (d) ist die mangelnde Vergleichbarkeit von Merkmalsträgern in der Groß-N-Analyse. Dieses Problem besteht weniger beim Vergleich von Indivi-

duen als von Kollektiven – so etwa bei der Analyse von Staaten (Goerres und Prinzen 2012). Kann man Staaten in bestimmten historischen Kontexten sinnvoll als homogene Einheiten vergleichen? Methodisch gesehen kann ein Merkmalsträger, auch ein Staat, mit einer Anzahl von  $n$  Informationen beschrieben werden, die für die vorliegende Forschungsfrage relevant sind. Wenn man jeden Merkmalsträger mit diesen  $n$  Informationen beschreiben kann, ist die sog. konditionale Homogenität der Merkmalsträger hergestellt. In anderen Worten: unter der Berücksichtigung aller relevanten Drittvariablen können Staaten oder andere komplexe Kollektive verglichen werden.

In diesem Kontext läuft man Gefahr zu unterstellen, man habe nie die perfekte Informationsbasis für den Vergleich von spezifischen Merkmalsträgern. Jede Realisation eines Merkmalsträgers aus einem theoretischen Universum in einer Stichprobe ist idiosynkratisch, nicht nur in Bezug auf die Kombination aller Ausprägungen, sondern auch in Bezug auf die Variablen, die wichtig sind, um den Merkmalsträger zu beschreiben. Dieser Unterstellung kann man leicht entgegenzutreten, indem man durch ein Groß-N-Design Ergebnisse produziert, die theoretisch sinnvoll zu interpretieren sind. Denn wenn es wirklich überhaupt keine Vergleichbarkeit gäbe, müsste die Anwendung des Groß-Designs chaotische Zufallsergebnisse bringen.

Die fünfte und letzte Grenze eines Groß-N-Designs (e) ist eher übergreifend und diffus, nämlich die scheinbare Objektivierung der Daten. Viele Daten, die in Zahlenform vorliegen, sind kein Garant für ein hochwertiges empirisches Projekt. Zahlen wirken objektiver als Worte, obwohl beide nur eindeutige Repräsentationen von bestimmten Inhalten sind. Diese Feststellung ist gleichzeitig banal und wichtig. Sie ist banal, weil sie jedem offensichtlich sein müsste, der sich ernsthaft mit empirischer Forschung beschäftigt. Sie ist aber auch wichtig, weil Zahlen und numerische Effektbeschreibung zumindest für den nicht ganz geschulten Betrachter wissenschaftlicher wirken als andere Ergebnisse. Diese Erkenntnis ist besonders dann relevant, wenn politikwissenschaftliche Studien aus ihrem Kontext gerissen werden, oder wenn wissenschaftliche Studien in der Politik genutzt werden (Mayntz 2017).

Manche Autoren verneinen grundsätzlich die Möglichkeit, gesetzmäßige Regelmäßigkeiten in der Sozialwissenschaft zu entwickeln und für die Politikberatung aufzubereiten, und legen den Wert auf die analytische Beschreibung und ihre Bedeutung in der Politikberatung (Streeck 2009). Andere betonen die Notwendigkeit zur Darstellung multipler Interpretationen in der wissenschaftlichen Politikberatung, um dieses Problem zu umgehen oder zumindest abzumildern (Renn 2017). Schließlich gibt es Situationen, in denen die Nutzung von wissenschaftlichen Studien in der Politik ausschließlich über Groß-N-Studien erfolgen kann. So beauftragte 2010 das bundesdeutsche Familienministerium das ifo-Institut München, zusammen mit dem DIW in Berlin und dem ZEW in Mannheim (alles wirtschaftswissenschaftliche Institute wohlgermerkt), die Wirksamkeit der verschiedenen Familienpolicies vor allem in Bezug auf Fertilität, weibliche Partizipation am bezahlten Arbeitsmarkt und Armut zu überprüfen. Die 2013 veröffentlichten Studien (Rainer et al. 2013a, b), deren Ergebnisse die Wirkungslosigkeit deutscher Familienpolitik entlarvten, sind ausnahmslos statistische Analysen, teilweise mit quasi-experimentellem Design. Jede Art von anderer Analyse wäre – gegeben die gute Messbarkeit der individuellen Eigenschaften (Einkommen, Fertilität, Bildung) – nicht sinnvoll gewesen.

Zusammenfassend sind Groß-N-Studien kein Allheilmittel in der Politikwissenschaft. Die teilweise problematische Datenqualität und die vielfach datengeleitete Operationalisierung der zu untersuchenden Merkmale sind Indizien dafür, dass Groß-N-Studien mit Bedacht interpretiert und genutzt werden sollten. Auch sollte die Gefahr einer Scheinobjektivierung von Zahlen als verlässliche Daten nicht unterschätzt werden. Ebenfalls gilt es, die Stichprobenszusammensetzung von Sekundärdaten vor der Nutzung kritisch zu hinterfragen und auf ihre Eignung für das eigene Forschungsvorhaben zu prüfen.

### **3.3 Die Zukunft und zentrale Entwicklungen von Groß-N-Studien in der Politikwissenschaft**

Es stehen goldene Zeiten für Groß-N-Designs in der Politikwissenschaft an. Dies hat mehrere Gründe: eine bessere Infrastruktur verfügbarer Daten für Sekundäranalysen, eine Professionalisierung der Publikationswege mit Einforderung von Replikationsdaten sowie die zunehmende Verbreitung von Big-Data-Techniken zur Aufbereitung großer Mengen unstrukturierter Daten.

Viele Staaten unterhalten inzwischen nationale Archive für wissenschaftliche Daten. In Deutschland hat das Leibniz-Institut für Sozialwissenschaften GESIS mit Sitzen in Köln und Mannheim diese Funktion. Wissenschaftlerinnen und Wissenschaftler können häufig durch einfache Onlinebestellung einen schnellen Zugang zu erhobenen Daten, vor allem Umfragedaten, bekommen. Gerade im Bereich der Umfrageforschung ist dies oftmals sehr einfach. Große Umfrageprojekte unterhalten eigene Onlinearchive wie beispielsweise der European Social Survey oder der World Values Survey. Während vergleichende Umfragedatenanalysen vor wenigen Jahrzehnten erhebliche Ressourcen verschlangen wie beispielsweise der Klassiker *Civic Culture* mit fünf Ländern (Almond und Verba 1963), kann zumindest für die Sekundäranalyse heutzutage jede Studentin und jeder Student eine solche Analyse in der Abschlussarbeit vornehmen.

In Deutschland gibt es darüber hinaus extern geförderte Forschungsdatenzentren, die häufig für einen sehr speziellen Forschungsstrang Daten für die wissenschaftliche Forschung zur Verfügung stellen. Beispielsweise unterhält das Institut für Arbeitsmarkt- und Berufsforschung in Nürnberg ein solches Zentrum für die Analyse von diversen Datensätzen mit teilweise sensiblen persönlichen Daten. Alle diese Archive stellen große Datensätze zur Verfügung, die sich mithilfe der Groß-N-Analyse neuen Forschungszielen zuführen lassen. Darunter fallen auch spezialisierte, digitale Infrastruktureinrichtungen wie CLARIN-D,<sup>6</sup> die sich Sprachressourcen im Bereich der Geistes- und Sozialwissenschaften verschrieben haben. Darin findet sich beispielsweise auch POLMINE,<sup>7</sup> eine Schnittstelle zur Analyse umfangreicher Textkorpora von Plenarprotokollen aus Bundestag und Landtagen.

<sup>6</sup><http://www.clarin-d.de/de/ueber/kurzuebersicht>.

<sup>7</sup><https://polmine.github.io>.



Forscherinnen und Forscher können ihre Daten inzwischen ganz einfach über nicht-kommerzielle Plattformen für Replikationsstudien zur Verfügung stellen: Dataverse in den USA<sup>8</sup> oder das Open Science Framework<sup>9</sup> sind Beispiele.

Auch auf der Makroebene gibt es eine zunehmende Anzahl von öffentlich zugänglichen Datensätzen. In dem Varieties of Democracies-Project (Coppedge et al. 2017) an der Universität Göteborg stehen globale Daten über diverse Aspekte demokratischer Systeme zur Verfügung. Ebenso in Göteborg findet sich die Quality of Government-Datenbank mit zahlreichen Indikatoren zur Regierungs- und Systemperformanz (Teorell et al. 2017). ParlGov trägt Daten zu Parlamenten, Regierungen und Wahlen für EU- und OECD-Staaten zusammen (Döring und Manow 2016). Das Comparative Political Data Set enthält zahllose Indikatoren über die politischen Institutionen moderner Demokratien. Amtliche oder quasi-amtliche Daten stellen die Weltbank, die OECD, der Internationale Währungsfonds und Eurostat zur Verfügung (Armingeon et al. 2016). Geodaten (siehe den Beitrag von Weidmann und Gleditsch in diesem Band) sowie von Geodaten abgeleitete Regionaldaten wie Milieudaten auf Straßenebene vervollständigen das Bild eines wachsenden Mekkas für Groß-N-Studien in der Politikwissenschaft.

Während einer der Größten unseres Faches, Robert Dahl, sich noch als Professor in die Statistik-Vorlesungen seiner Kollegen setzte, um statistische Analysen zu lernen (Dahl 1997), ist heute eine Befähigung zur Groß-N-Analyse eine Grundvoraussetzung für die Politikwissenschaft auch im Early-Career-Bereich. Diese Standardvoraussetzung wird zusätzlich durch eine zunehmende Verfügbarkeit von Datensätzen für die Reproduktion und Replikation gespiegelt.

Die Entwicklung zur Veröffentlichung von Wissenschaftsdaten in der Politikwissenschaft wird entscheidend davon vorangetrieben, dass wissenschaftliche Zeitschriften bereits bei der Einreichung eines Manuskripts die Zusage erwarten, dass die Daten schon bei der Begutachtung, aber spätestens nach der Veröffentlichung zur Verfügung gestellt werden. Dieser Trend zeigt sich unverkennbar im Bereich der quantitativen Studien, aber auch zunehmend bei qualitativen Studien, wenngleich hier die Infrastruktur und die Standards nicht so deutlich sind. Deutschsprachige Zeitschriften hinken hier dem internationalen Trend allerdings noch hinterher. So gab es im Juli 2018 für die Österreichische Zeitschrift für Politikwissenschaft keinerlei Vorgaben zur Datenverfügbarkeit. Bei der Swiss Political Science Review handelt es sich um eine Soll-Vorgabe. Lediglich in der Politischen Vierteljahresschrift ist die Vorgabe ein Muss. Das bedeutet, dass ein *circulus virtuosus* vorliegt: je mehr Groß-N-Studien durchgeführt werden, und je mehr Wert auf Replikationsdaten gelegt wird, desto mehr Folgestudien mit Groß-N-Design sind zu erwarten.

Schließlich existieren immer mehr Möglichkeiten, große Mengen unstrukturierter Daten (Big Data) zu strukturieren und eine Groß-N-Studie durchzuführen. Facebook und Twitter waren im Jahr 2018 die großen Plattformen mit entsprechenden Schnittstellen für die Forschung. Facebook hat sich 2018 sogar in eine explizite Partner-

---

<sup>8</sup><http://dataverse.org/about>.

<sup>9</sup><https://osf.io/>.

schaft mit der Wissenschaft begeben, um die eigenen Daten für Forschung zur Verfügung zu stellen. Obwohl die Plattformen und Strukturen vermutlich nicht langfristig stabil sind, zeigt sich ein unaufhaltsamer Trend zu mehr und größeren Datensätzen, die durch technisch Versierte gehoben werden können.

## 4 Kommentiertes Literaturverzeichnis

Das Kapitel von Pickel (2016) im Handbuch zur Vergleichenden Politikwissenschaft diskutiert ausführlich die quantitativen Ansätze im Ländervergleich. Boyd und Crawford (2012) stellen in ihrem Beitrag die Diskussion von eher grundsätzlichen epistemologischen und ethischen Diskursen im Kontext der ansteigenden Datenflut dar und geben einen sehr guten Einblick in die „neuen“ Probleme großer Datenmengen. Schrodt (2014) beschreibt in seinem Artikel eine Reihe von häufig gemachten Fehlern in der quantitativen Politikwissenschaft, deren Lektüre sehr hilfreich für Wissenschaftlerinnen und Wissenschaftler jedes Erfahrungsstands ist.

## Literatur

- Almond, Gabriel A., und Sidney Verba. 1963. *The civic culture: Political attitudes and democracy in five nations*. Princeton: Princeton University press.
- Andreß, Hans-Jürgen, und Thorsten Heien. 2001. Four Worlds of Welfare State Attitudes? A comparison of Germany, Norway, and the United States. *European Sociological Review* 17(4): 337–356.
- Armingeon, Klaus, Christian Isler, Laura Knöpfel, David Weisstanner, und Sarah Engler. 2016. *Comparative political dataset 1960–2014*. Berne: Institute of Political Science, University of Berne.
- Bartle, John. 2005. Homogeneous models and heterogeneous voters. *Political Studies* 53(4): 653–675. <https://doi.org/10.1111/j.1467-9248.2005.00550.x>.
- Bayram, A. Burcu. 2015. What drives modern Diogenes? Individual values and cosmopolitan allegiance. *European Journal of International Relations* 21(2): 451–479. <https://doi.org/10.1177/1354066114541879>.
- Behnke, Joachim. 2005. Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. *Politische Vierteljahresschrift* 46(1): 1–15.
- Berg-Schlosser, Dirk, und Theo Stamm. 2003[1974]. *Einführung in die Politikwissenschaft*, 7., durchges. u. erw. Aufl. München: C. H. Beck.
- Bond, R. M., C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, und J. H. Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415): 295–298. <https://www.nature.com/articles/nature11421>. Zugegriffen am 29.09.2018.
- Boyd, Danah, und Kate Crawford. 2012. Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679. <https://doi.org/10.1080/1369118X.2012.678878>.
- Brady, Henry E., und David Collier, Hrsg. 2010. *Rethinking social inquiry: Diverse tools, shared standards*. Lanham: Rowman & Littlefield Publishers.
- Broscheid, Andreas, und Thomas Gschwend. 2005. Zur statistischen Analyse von Vollerhebungen. *Politische Vierteljahresschrift* 46(1): 16–26.
- Central Intelligence Agency. 2017. World fact book: Country comparison. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2119rank.html>. Zugegriffen am 10.01.2018.

- Colleoni, Elanor, Alessandro Rozza, und Adam Arvidsson. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication* 64(2): 317–332. <https://doi.org/10.1111/jcom.12084>.
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Laura Saxer, Brigitte Seim, Rachel Sigman, und Jeffrey Staton. 2017. V-Dem Codebook v7.1. Varieties of Democracy (V-Dem) Project.
- Dafoe, A., J. R. Oneal, und B. Russett. 2013. The democratic peace: Weighing the evidence and cautious inference. *International Studies Quarterly* 57(1): 201–214. <https://doi.org/10.1111/isqu.12055>.
- Dahl, Robert A. 1997. A brief intellectual biography. In *Comparative European Politics: The story of profession*, Hrsg. Hans Daalder, 68–78. London: Pinter.
- Dinas, Elias, und Kostas Gemenis. 2010. Measuring parties' ideological positions with manifesto data: A critical evaluation of the competing methods. *Party Politics* 16(4): 427–450.
- Döring, Holger, und Philip Manow. 2016. Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in modern democracies. Development version. <http://www.parlgov.org/>. Zugegriffen am 10.01.2018.
- Facebook Inc. 2017. Stats. <https://newsroom.fb.com/company-info>. Zugegriffen am 10.01.2018.
- Fearon, J. D., und D. D. Laitin. 2008. Integrating qualitative and quantitative methods. In *The Oxford handbook of political methodology*, Hrsg. J. M. Box-Steffensmeier, H. E. Brady, D. Collier, J. D. Fearon und D. D. Laitin. Oxford: Oxford University Press.
- Goerres, Achim, und Katrin Prinzen. 2012. Using mixed methods for the analysis of individuals: A review of necessary and sufficient conditions and an application to welfare state attitudes. *Quality & Quantity* 46(2): 415–450.
- Haspel, Moshe, und Gibss H. Knotts. 2005. Location, location, location: Precinct placement and the costs of voting. *The Journal of Politics* 67(02): 560–573.
- Jackman, Robert W. 1987. Political institutions and voter turnout in the industrial democracies. *American Political Science Review* 81(2): 405–423.
- King, Gary, Robert O. Keohane, und Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton: Princeton University Press.
- Kittel, Bernhard. 2006. A crazy methodology? *International Sociology* 21(5): 647–677. <https://doi.org/10.1177/0268580906067835>.
- Kittel, Bernhard, und Herbert Obinger. 2003. Political parties, institutions, and the dynamics of social expenditure in times of austerity. *Journal of European Public Policy* 10(1): 20–45.
- Kwak, H., C. Lee, H. Park, und S. Moon. 2010. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on world wide web*, April 26–30, Raleigh/North Carolina/USA, 591–600.
- Lieberman, Evan S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99(3): 435–452.
- Lundsgaarde, E., C. Breunig, und A. Prakash. 2010. Instrumental philanthropy: Trade and the allocation of foreign aid. *Canadian Journal of Political Science/Revue canadienne de science politique* 43(3): 733–761.
- Mayntz, Renate. 2017. Zählen – Messen – Entscheiden: Wissen im politischen Prozess. MPIfG Discussion Paper 17/12.
- Paxton, Pamela. 2002. Social capital and democracy: An interdependent relationship. *American Sociological Review* 67(2): 254–277. <https://doi.org/10.2307/3088895>.
- Pickel, Gert. 2016. Quantitative makroanalytische Verfahren in der Vergleichenden Politikwissenschaft. In *Handbuch Vergleichende Politikwissenschaft*, Hrsg. Hans-Joachim Lauth, Marianne Kneuer und Gert Pickel, 47–62. Wiesbaden: Springer Fachmedien Wiesbaden.
- Przeworski, Adam, und Henry Teune. 1970. *The logic of comparative social inquiry*. New York: Wiley-Interscience.
- Puschmann, Cornelius, und Jean Burgess. 2014. The politics of twitter data. In *Twitter and society*, Hrsg. Katrin Weller, Axel Bruns, Jean Burgess, Merja Maht und Cornelius Puschmann, 43–54. New York: Peter Lang.

- Rabuza, Florian. 2016. Formale Bildung und Wahlbeteiligung am Anfang des 21. Jahrhunderts in Europa: eine vergleichende Analyse der Kontextabhängigkeit ungleicher politischer Beteiligung mit Daten des European Social Survey. Duisburg, Essen, Universität Duisburg-Essen, Dissertation, 2016.
- Rainer, Helmut, Stefan Bauernschuster, Wolfgang Auer, Natalia Danzer, Mine Hancioglu, Bastian Hartmann, Timo Hener, Christian Holzner, Notburga Ott, Janina Reinkowski, und Martin Werdig. 2013a. Kinderbetreuung. *ifo Forschungsbericht* 59.
- Rainer, Helmut, Stefan Bauernschuster, Natalia Danzer, Timo Hener, Christian Holzner, und Janina Reinkowski. 2013b. Kindergeld. *ifo Forschungsbericht* 60.
- Renn, Ortwin. 2017. Kommunikation zwischen Wissenschaft und Politik. In *Forschungsfeld Wissenschaftskommunikation*, Hrsg. Heinz Bonfadelli, Birte Fähnrich, Corinna Lütjhe, Jutta Milde, Markus Rhomberg und Mike S. Schäfer, 183–205. Wiesbaden: Springer Fachmedien Wiesbaden.
- Rohlfing, Ingo. 2008. What you see and what you get: Pitfalls and principles of nested analysis in comparative research. *Comparative Political Studies* 41(11): 1492–1514.
- Schrodt, Philip A. 2014. Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research* 51(2): 287–300.
- Scruggs, Lyle, Detlef Jahn, und Kati Kuitto. 2017. „Comparative Welfare Entitlements Data Set 2, Version 2017-09.“ University of Connecticut & University of Greifswald. <http://cwed2.org/>. Zugegriffen am 29.09.2018.
- Shayo, M. 2009. A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review* 103(2): 147–174. <https://doi.org/10.1017/S0003055409090194>.
- Streeck, Wolfgang. 2009. Man weiß es nicht genau: Vom Nutzen der Sozialwissenschaften für die Politik. MPIfG Working Paper 09/11.
- Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko, und Richard Svensson. 2017. The Quality of Government Standard Dataset, version Jan 17. University of Gothenburg: The Quality of Government Institute. <http://www.qog.pol.gu.se>. <https://doi.org/10.18157/QoGStdJan17>. Zugegriffen am 10.01.2018.
- Twitter Inc. 2017. Twitter Nutzung/Fakten zum Unternehmen. <https://about.twitter.com/de/company>. Zugegriffen am 10.01.2018.
- Volkens, Andrea. 2007. Strengths and weaknesses of approaches to measuring policy positions of parties. *Electoral Studies* 26(1): 108–120.
- Weidmann, N. B., und S. Schutte. 2017. Using night light emissions for the prediction of local wealth. *Journal of Peace Research* 54(2): 125–140. <https://doi.org/10.1177/0022343316630359>.
- Windelband, Wilhelm. 2017 [1894]. Geschichte und Naturwissenschaft. [http://www.hs-augsburg.de/~harsch/germanica/Chronologie/19Jh/Windelband/win\\_rede.html](http://www.hs-augsburg.de/~harsch/germanica/Chronologie/19Jh/Windelband/win_rede.html). Zugegriffen am 10.01.2018.
- Yang, J., und J. Leskovec. 2011. Patterns of temporal variation in online media. In ACM international conference on web search and data mining (WSDM), 09–12 Feb, Hong Kong/China.