

Herausforderungen und Möglichkeiten medienbasierter Bürgerkriegsdatensätze

In jüngster Zeit wurden zahlreiche neue globale Datensätze zu politischer Gewalt erhoben, die häufig auf Medienberichten basieren. Datensätze, die ausschließlich oder überwiegend auf Zeitungsartikeln aufbauen, werden häufig kritisch bewertet, denn Medienberichte können aufgrund selektiver und/oder unvollständiger Berichterstattung verzerrt sein, sodass bestimmte Informationen gar nicht oder falsch dargestellt werden. In diesem Aufsatz gehe ich auf die unterschiedlichen Arten der medialen Verzerrungen ein und diskutiere die damit einhergehenden Herausforderungen für die Anwendung medienbasierter Konfliktdatensätze. Es gibt sicherlich keine Patentlösung für die skizzierten Herausforderungen. Jedoch tragen Transparenz in der Datenerhebung – wie wurden die Daten erhoben und welche Quellen wurden genutzt – und ein verantwortungsbewusster Umgang mit den Daten durch die Nutzer_innen dazu bei, die aus der selektiven Berichterstattung resultierenden Probleme für statistische Analysen abzufedern.

1. Einleitung

In jüngster Zeit wurden zahlreiche neue globale Datensätze zu politischer Gewalt erhoben. Diese Datensätze zeichnen sich dadurch aus, dass sie weltweit bzw. für mehrere Kontinente Informationen zu kollektiv organisierter Gewalt zur Verfügung stellen sowie räumlich und zeitlich disaggregierte Angaben zu einzelnen Gewaltereignissen enthalten. Die systematische Erhebung neuer Daten erlaubt vergleichende Studien zu den Ursachen und Konsequenzen von Bürgerkriegen und zu den damit verbundenen Gewaltformen wie Terrorismus, Gewalt gegen Zivilist_innen oder kommunale Konflikte (Schneider 2015). Diese Daten stehen allen Forschenden als kollektives Gemeingut zur Verfügung und ermöglichen ihnen, neuen Fragestellungen nachzugehen und empirisch anspruchsvolle Analysen durchzuführen. Aufgrund der Datenmenge ist es begrüßenswert, dass z. B. Chojnacki und Reisch (2012) herausgearbeitet haben, welche Informationen die einzelnen Datensätze enthalten und welche Vor- und Nachteile sie für bestimmte Fragestellungen mit sich bringen. Die deutschsprachige Politikwissenschaft hat jedoch der Quellenauswahl und deren Implikationen für die Datensätze bisher wenig Beachtung geschenkt. In diesem Aufsatz beleuche ich, ob die Quellenauswahl Einfluss auf die Datenerhebung und nachfolgende Analysen hat, und wenn ja, welchen.

Die prominentesten nationalen und internationalen Datensätze verwenden überwiegend Tageszeitungen und Nachrichtenagenturen als Grundlage für ihre Datener-

hebung (s. Tab. 1 auf S. 139/140).¹ Datensätze, die ausschließlich oder überwiegend auf Zeitungsartikeln aufbauen, werden häufig kritisch bewertet, denn Medienberichte können aufgrund selektiver und/oder unvollständiger Berichterstattung verzerrt sein, sodass bestimmte Informationen gar nicht oder falsch dargestellt werden. Zeitungsbasierte Datensätze sind deshalb dem Risiko ausgesetzt, dass sie ein Konfliktgeschehen nicht repräsentativ wiedergeben. Dadurch können die Ergebnisse statistischer Analysen, die mit medienbasierten Datensätzen durchgeführt werden, im schlimmsten Fall zu falschen Schlussfolgerungen führen (Earl et al. 2004). Was bedeutet mediale Verzerrung und welche Konsequenzen hat sie? Welche Maßnahmen können bei der Erhebung und Anwendung politischer Gewaltdaten die mediale Verzerrung und die damit verbundenen Fallstricke mindern? Die Antworten auf diese Fragen sind für Anbieter_innen und Nutzer_innen von globalen Konflikt-datensätzen gleichermaßen relevant. Zum einen kann eine konstruktive Quellenkritik dazu beitragen, die Datenerhebung zu verbessern und deren Reliabilität zu erhöhen. Zum anderen verhindert ein reflektierter Umgang mit medienbasierten Daten, dass voreilige oder falsche Schlussfolgerungen in Bezug auf Bürgerkriege und damit verbundene politische Gewalt gezogen werden.

Dieser Aufsatz ordnet sich in die Debatte um die Datenqualität ein, die in der Politikwissenschaft seit kurzem international (z. B. Salehyan 2015) aber auch national in Gang gekommen ist (z. B. Pickel/Pickel 2012).² Ich gliedere ihn so, dass ich zunächst auf die Vorteile der medienbasierten Datenerhebung eingehe und dann die damit einhergehenden Herausforderungen darstelle. Im Weiteren diskutiere ich, ob alternative Formen der Datenerhebung – speziell die Nutzung von Primärquellen – zur Überwindung der skizzierten Problematik beitragen können. Ich schließe den Beitrag mit einigen Implikationen für Forschende, die medienbasierte Datensätze erheben und anwenden, ab.

2. Medienbasierte Datenerhebung von Konflikten

Globale Datensätze tragen dazu bei, Aussagen über die Entwicklung und Tendenzen von Bürgerkriegen und den damit verbundenen Gewaltformen zu treffen. Forschende können aufgrund logistischer und materieller Ressourcenbeschränkungen sowie ethischer Herausforderungen die einzelnen politischen Gewaltakte und -episoden, die weltweit ausgetragen werden, nicht selbst beobachten. Deswegen greifen sie in der Regel auf schon vorhandene Informationen zurück. Medienberichte sind zu einer der meistgenutzten Quellen für die Erhebung politischer Gewaltdaten avanciert (z. B. Schneider/Bussmann 2013; s. Tab. 1). Im Folgenden werde ich zunächst die Vorteile der medienbasierten Datenerhebung erläutern und dann die aus

1 Im Folgenden verwende ich den Begriff Medienberichte, wenn ich Zeitungsartikel und Meldungen von Nachrichtenagenturen meine.

2 Ich konzentriere mich auf die Datenerhebung und damit verbundene Messprobleme. Somit ist dieser Aufsatz komplementär zu Aufsätzen, die die quantitativen Methoden in den Internationalen Beziehungen beleuchten (z. B. Ruhe et al. 2017).

ihr resultierenden Herausforderungen beleuchten. Dieser Aufsatz setzt sich mit den Verzerrungsarten, die der Medienberichterstattung inhärent sind, auseinander. Es würde den Rahmen des Aufsatzes sprengen, wenn ich weitere Herausforderungen wie beispielsweise Vor- und Nachteile von numerischen Schwellen (Kende 1978; Kreutz 2015) oder die Schwierigkeiten, Bürgerkriege zu definieren und zu operationalisieren (z. B. Sambanis 2004), diskutieren würde.

2.1. Die Vorteile der medienbasierten Datenerhebung

Im Vergleich zu alternativen Informationsquellen wie z. B. Archiven, historischen Fallstudien oder Kriegstagebüchern, haben Medienberichte fünf Vorteile: *Erstens* sind Zeitungen und Nachrichtenagenturen die einzigen Quellen, die täglich und weltweit über Bürgerkriege berichten. Sie beschreiben, welche Akteur_innen wann und wo Gewalt ausgeübt haben. Somit liefern Medienberichte die detaillierten Informationen, die benötigt werden, um umfangreiche Konfliktdatensätze zu erheben (Öberg/Sollenberg 2011). *Zweitens* ermöglicht die Verwendung von Medienberichten den Forschenden, die Daten auf Grundlage eigener theoretischer Konzepte und Definitionen zu erheben. Dadurch ist die Datenerhebung nicht nur unabhängig von vorgegebenen Definitionen und Konzepten anderer Quellen, wie z. B. nationalen Statistikämtern, sondern auch konsistent, wenn Daten für unterschiedliche Konflikte erhoben werden (Koopmans 1999: 95, 96). Eine einheitliche Datenerhebungsmethodik ist notwendig, um Konflikte zwischen einzelnen Ländern, aber auch auf subnationaler Ebene miteinander vergleichen zu können. *Drittens* schaffen die permanente Berichterstattung und die öffentlich zugänglichen Zeitungsartikel die Voraussetzung dafür, Datensätze zeitnah zu den Konflikten zu erheben (z. B. Otto 2013; Chojnicki/Reisch 2012). Im Gegensatz zu anderen Quellen, beispielsweise Militärdatenbanken, die häufig erst nach Jahrzehnten deklassifiziert werden, können medienbasierte Datensätze auf jährlicher Basis aktualisiert (z. B. das *Uppsala Conflict Data Program*, UCDP) oder in Echtzeit erhoben werden (z. B. das *Armed Conflict Location & Event Data Project*, ACLED). Aktuelle und systematisch aufgearbeitete Daten sind vor allem von Nutzen, um Vorhersagen über Konflikte treffen zu können und Risikoszenarien zu erstellen. *Viertens* ist Datenerhebung in Konfliktländern gefährlich, teuer und ethisch teilweise problematisch. Die medienbasierte Datenerhebung stellt hier eine preiswerte und sichere Alternative dar. *Fünftens* erhalten medienbasierte Konfliktdatensätze Daten, die auch von Entscheidungsträger_innen wahrgenommen werden. Das heißt, dass solche Datensätze die relevanten Informationen erfassen, die die Entscheidungen und Handlungslogiken von Staatsoberhäuptern und Konfliktparteien beeinflussen (Schneider/Bussmann 2013).

2.2. Die Herausforderungen der medienbasierten Datenerhebung

Obwohl die auf Medienberichten basierende Datenerhebung vielversprechend ist und eine Grundlage für Quer- und Längsschnittanalysen bildet, ist diese Erhebungsmethode auch mit Problemen behaftet. Konfliktdata, die auf Grundlage von Zeitungsberichten erhoben wurden, sind der Kritik ausgesetzt, dass sie die Konfliktgeschehen verzerrt darstellen. In der Literatur wird zwischen zwei Arten medialer Verzerrung (*media bias*) unterschieden: Beschreibungsverzerrung und Berichterstattungsverzerrung (Earl et al. 2004).

Die Beschreibungsverzerrung (*description bias*) betrifft die Richtigkeit und Vollkommenheit der Berichterstattung (Earl et al. 2004). Für die Erhebung von Konfliktdata sind insbesondere die sogenannten »hartes« Fakten wie Zeitpunkt, Ort, Täter_innen und Anzahl der Todesopfer relevant. Zeitungsartikel gelten als eine verlässliche Quelle für »harte« Fakten, weil solche Beschreibungen in der Regel keiner Interpretation bedürfen (Earl et al. 2004: 72). Nichtsdestoweniger können auch diese »harten« Fakten verzerrt wiedergegeben werden. Das kann z. B. der Fall sein, wenn die Informationen direkt von den Konfliktparteien bezogen werden. Besonders Aussagen über die Anzahl von Todesopfern und darüber, wer die Opfer der Gewalt waren, sind anfällig für Verzerrungen, weil diese Angaben häufig für Propagandazwecke der Konfliktakteure benutzt werden (Chojnacki et al. 2012) und somit zu einer Über- oder Untertreibung der eigenen Todesopfer oder der Gegenseite führen können. Andererseits besteht die Möglichkeit, dass ein Konfliktgeschehen nicht ausreichend beschrieben wird. Medienberichte über Gewaltereignisse variieren meist hinsichtlich der verfügbaren Details. Ungenaue Angaben zum Ort der Gewaltausübung oder über die Täter_innen sind keine Seltenheit (Otto 2013: 561, 562).

Die Berichterstattungsverzerrung (*reporting bias*) hingegen bezieht sich auf die Auswahl der berichteten Ereignisse, das heißt, dass Zeitungen und Nachrichtenagenturen nicht alle Gewaltbereignisse berichten, die stattgefunden haben (Earl et al. 2004: 69, 70; Weidmann 2016: 207). Aufgrund dessen repräsentieren sie möglicherweise nur einen Ausschnitt der politischen Gewalt. Die erhobenen Daten stellen somit häufig nicht die Grundgesamtheit der ausgeübten politischen Gewaltakte dar, sondern nur eine Stichprobe. Selektive Berichterstattung kann durch ein allgemein geringes Medieninteresse an bestimmten Konflikten begründet sein. Hawkins (2011) hat beispielsweise aufgezeigt, dass über einige der gewaltsamsten Bürgerkriege selten berichtet wurde. Ein weiterer Grund für die selektive Berichterstattung resultiert aus der »Nachrichtentauglichkeit« des Ereignisses. Nachrichtenagenturen und Herausgeber_innen entscheiden, ob ein beobachtetes Gewaltbereignis als berichtenswert eingestuft wird. Solche Entscheidungen sind oftmals vom Interesse der Leser_innenschaft abhängig (Öberg/Sollenberg 2011: 56). Die Untererfassung (*underreporting*) von Gewaltbereignissen resultiert außerdem aus dem Stadt-Land-Gefälle. Der Mangel an Journalist_innen in abgelegenen Regionen und der eingeschränkte Zugang zu Telefonnetzen und Internet können dazu führen, dass Gewaltbereignisse einfach nicht berichtet werden (Kalyvas 2004: 164).

2.3. Warum und unter welchen Umständen ist mediale Verzerrung ein Problem?

Nachdem zwei Arten medialer Verzerrung unterschieden werden können, stellt sich die Frage, welche Probleme sich aus diesen Verzerrungen für die Datensätze zu politischer Gewalt ergeben. Verkürzt könnte man sagen: Es kommt drauf an. Im Folgenden gehe ich deshalb auf die Herausforderungen ein, die aus der Berichterstattungs- und Beschreibungsverzerrung resultieren.

Medienbasierte Konfliktdatensätze bilden oftmals nur einen Ausschnitt der Gewaltereignisse ab, die stattgefunden haben. Statistisch ausgedrückt bedeutet das, dass die erhobenen Daten eine Stichprobe aus der Grundgesamtheit der Gewalt darstellen. Wenn die Datensätze die Kriterien einer Zufallsstichprobe erfüllten, wäre die Anwendung der Daten für deskriptive Darstellungen und statistische Analysen unproblematisch. Aufgrund der Berichterstattungsverzerrung stellen medienbasierte Datensätze jedoch häufig sogenannte willkürliche Stichproben (*convenience samples*, z. B. Price/Ball 2015) dar. Das bedeutet, dass der Erhebungsprozess nicht nach dem Zufallsprinzip erfolgte, sondern dass nur die jeweils verfügbaren, beobachtbaren Informationen erfasst wurden. Somit sind medienbasierte Konfliktdatensätze aufgrund von Messproblemen willkürliche Stichproben. Das wiederum bedeutet, dass angesichts der selektiven Berichterstattung einige Ereignisse, die stattgefunden haben, nicht erhoben und somit als »Nicht-Gewaltereignisse« erfasst werden. Das heißt, dass über Gewalt, die eigentlich stattgefunden hat, nicht berichtet wird und dass sie nicht in medienbasierten Datensätzen auftaucht.

Ereignisdatensätze sind also aufgrund der Berichterstattungsverzerrung mit Messfehlern behaftet. Die Herausforderung dabei ist nicht, dass die Datensätze nicht das gesamte Universum der Gewalt erfassen, sondern dass die Stichproben systematisch verzerrt sein können. Die Erhebung der Daten korreliert möglicherweise mit Variablen, die das Auftreten politischer Gewalt erklären sollen (dazu ausführlich: Weidmann 2016). In Regionen, in denen mehr Menschen leben oder in denen Menschen Zugang zu Informationstechnologien haben, ist die Wahrscheinlichkeit höher, dass Zeug_innen Informationen über Gewalt an Journalist_innen übermitteln, als im Vergleich zu dünner besiedelten oder Regionen ohne Zugang zu Internet oder Mobilfunknetzen. Somit können höhere Gewaltniveaus, die in den Datensätzen erfasst werden, eine Folge besserer Berichterstattung sein und sind nicht zwingend gleichbedeutend mit der höheren Gewaltbereitschaft von bewaffneten Akteuren. Bei Analysen, die auf willkürlichen Stichproben basieren, besteht also die Gefahr, dass sie Aussagen über die Berichterstattungsmuster treffen statt über Gewaltmuster. Price und Ball (2015: 243-244) geben dafür ein Beispiel: Im Fall der Wahrheitskommission in Peru ergaben Zeug_innenaussagen, dass der Staat eine höhere Anzahl von Menschenrechtsverletzungen begangen hat als die Rebellengruppe »Leuchtender Pfad« (*Sendero Luminoso*). Als die Selektionsverzerrung mittels statistischer Modelle berücksichtigt wurde, stellte sich heraus, dass der Staat nicht mehr Gewalt verübt hatte, sondern dass staatlicherseits verübte Gewalt mit einer höheren Wahrscheinlichkeit berichtet wurde als Gewaltakte der Rebellen_innen. Eine naive Interpretation dieser Daten wäre zu dem Schluss gekommen, dass staat-

liche Sicherheitsbehörden für einen größeren Teil der Gewalt verantwortlich waren und hätte nicht das »tatsächliche« Gewaltmuster erkannt. Ein anderes Beispiel liefert der Datensatz ACLED. Ein Mitarbeiter des Datenerhebungsteams verweist darauf, dass die Daten die Gewalt in urbanen Regionen im Vergleich zu ländlichen Gegenden überrepräsentieren (Wigmore-Shepherd 2014: 7). Eine unreflektierte Anwendung der Daten könnte zu dem Urteil kommen, dass städtische Gebiete stärker als abgelegene Regionen von Konflikten und Gewalt betroffen sind.

Deskriptive Darstellungen sind genauso wie statistische Analysen von den potenziellen Konsequenzen der willkürlichen Stichproben betroffen. Price und Ball (2015: 243-244) zeigen, dass selbst deskriptive Trends von Gewalt in Syrien die Realität falsch abbilden können, wenn die Daten auf willkürlichen Stichproben basieren. Für statistische Analysen sind Messprobleme dann ein Problem, wenn der Messfehler des zu untersuchenden Phänomens (abhängige Variable) systematisch mit dem zu erklärenden Faktor (unabhängige Variable) korreliert. Unter solchen Umständen sind die Schätzungen der Regressionsmodelle verzerrt (Wooldridge 2013: 318). Weidmann (2016) zeigt die Problematik an dem Beispiel von Mobiltelefonen und Rebellengewalt auf. Eine Analyse, der ein medienbasierter Datensatz zugrunde liegt, kommt zu dem Ergebnis, dass besserer Mobiltelefonempfang die Gewaltanwendung durch Rebellen_innen erhöht. Wenn jedoch die gleiche Analyse mit einem nicht-verzerrten Datensatz durchgeführt wird, findet sich kein statistischer Zusammenhang zwischen den Variablen. Zusammengefasst heißt das: Wenn Forschende die Berichterstattungsverzerrung in medienbasierten Datensätzen auf niedriger Aggregationsebene nicht ernst nehmen, können ihre Analysen Gewaltmuster fehlerhaft interpretieren und schlimmstenfalls zu falschen Schlussfolgerungen führen.

Welche Konsequenzen ergeben sich aus der Beschreibungsverzerrung? Diese Art der Verzerrung wird umso spürbarer, je mehr detaillierte Informationen für die Datenerhebung benötigt werden. Der Informationsbedarf sinkt mit zunehmender Aggregationsebene und somit der Grad der Verzerrung. Wenn der Bürgerkrieg die Analyseeinheit der Datenerhebung ist, dann wird die Information benötigt, dass ein Staat sich mit mindestens einer Rebellengruppe in einem bewaffneten Konflikt befindet. Hierbei ist es unerheblich, wie viele Rebellengruppen den Staat bekämpfen und wann und wo genau die militärischen Interaktionen innerhalb eines Landes ausgetragen wurden. Für Ereignisdatensätze hingegen bedarf es präziser Informationen dazu, wer wann, wo und wie Gewalt ausgeübt hat. Wenn diese Angaben fehlen, können möglicherweise bestimmte Ereignisse nicht erhoben oder bestimmte Aspekte der Ereignisbeschreibung nicht spezifiziert werden. UCDP beispielsweise erhebt nur Daten, wenn die ausgeübte Gewalt einer spezifischen bewaffneten Gruppe zugeordnet werden kann. ACLED hingegen erhebt auch Daten zu »nicht-identifizierten bewaffneten Gruppen«. Solche Unterschiede können den Eindruck vermitteln, dass verschiedene Datensätze unterschiedliche »Welten der Gewalt« (Chojnacki/Reisch 2012: 40) abbilden.

Eine weitere Herausforderung der Beschreibungsverzerrung ist die Diskrepanz zwischen theoriegeleiteten Definitionen und den Informationen, die Medienberich-

ten zu entnehmen sind. Operationalisierung und Erhebung theoriegeleiteter Definitionen setzen, unabhängig vom Aggregationsniveau, ein Mindestmaß an empirisch beobachtbarer Information voraus, das häufig nicht gegeben ist. UCDP beispielsweise definiert Gewalt gegen Zivilist_innen als die direkte und intentionale Anwendung bewaffneter Gewalt durch die Regierung eines Staates oder eine formal organisierte Gruppe gegen Zivilist_innen (Eck/Hultman 2007: 235). Otto (2013) zeigt beispielsweise, dass Zeitungsartikel häufig keinen Rückschluss auf die beabsichtigte und direkte Gewaltanwendung zulassen. Aufgrund dessen werden Gewaltereignisse mit zivilen Todesopfern nicht in der Kategorie Gewalt gegen Zivilist_innen, sondern in der Kategorie militärische Todesopfer kodiert, wenn in den Medienberichten kein Nachweis der Intentionalität der Gewaltanwendung zu finden ist. Zusammengekommen bedeutet das, dass der Einfluss der Beschreibungsverzerrung auf die Datenqualität vom Aggregationsniveau abhängig ist. Des Weiteren kann die Beschreibungsverzerrung zu einer konservativen und selektiven Datenerhebung führen. Konfliktdatensätze erfassen häufig nur die Ereignisse, die ausreichend Informationen enthalten, um den theoretischen Konzepten gerecht zu werden. Die Eingrenzung der Medienberichte auf die, die nur detaillierte Informationen enthalten, ist eine weitere Ausprägung der nicht randomisierten Datenerhebung.

3. Implikationen für die Forschung

Welche Implikationen lassen sich aus der Diskussion ableiten? Eine Option wäre, den medienbasierten Daten den Rücken zu kehren und andere Quellen zu verwenden. Primärquellen sind beispielsweise Daten aus »erster Hand« und liefern demzufolge ungefilterte Informationen. Eine mögliche Alternative könnte demnach die Nutzung von Primärquellen sein, um die Konsequenzen medialer Verzerrungen in Datensätzen zu vermeiden. Die populärsten Primärdaten in der Konfliktforschung sind Surveys, Interviews, offizielle Statistiken und Berichte von Wahrheitskommissionen.

Surveys und Interviews ermöglichen es, reichhaltige Daten zu erheben, deren Informationsgehalt weit über den von Zeitungsberichten oder anderen Sekundärquellen hinausgeht. Nichtsdestoweniger sind beide Erhebungsmethoden nicht frei von Verzerrungen. Obwohl bei korrekter Durchführung Surveys einer Zufallsstichprobe gleichkommen, ist das in Konflikt- oder Post-Konfliktländern selten der Fall. Aufgrund des Sicherheitsrisikos für alle Beteiligten werden die Teilnehmer_innen selten nach dem Zufallsprinzip ausgewählt.³ Ähnlich verhält es sich mit der Auswahl von Interviewpartner_innen. Häufig liegen dem Auswahlprozess pragmatische Entscheidungen zugrunde, die nicht notwendigerweise methodischen Anforderungen genügen. Die Daten, die mithilfe von Surveys und Interviews gewonnen werden, sind deswegen ebenfalls dem Risiko ausgesetzt, dass die Auswahl der Teilnehmen-

3 Schneider und Bussmann (2013: 637) diskutieren unterschiedliche Probleme der Surveydatenerhebung in Konfliktländern.

den mit der erklärenden Variable korreliert und somit die Daten verzerrt sind (z. B. Höglund/Oberg 2011).⁴ Offizielle Statistiken und Wahrheitskommissionen sind in ähnlichem Maße von Berichterstattungsverzerrungen betroffen wie Medienberichte. Erstere erheben Daten zu bekannt gewordenen Ereignissen – über die Dunkelziffer lässt sich nur spekulieren. Außerdem sind offizielle Statistiken anfällig für Manipulationen, beispielsweise wenn Staaten ihre eigenen Statistiken zu Menschenrechtsverletzungen schönen (Salehyan 2015). Wahrheitskommissionen, wie in Peru oder Südafrika, basieren auf Stichproben. Wenn die Zeug_innen nicht zufällig ausgewählt wurden, unterliegen auch diese Daten einer Berichterstattungsverzerrung und erlauben eher Rückschlüsse auf die Muster der Berichterstattung als auf die Gewaltmuster (Price/Ball 2015). Es steht außer Frage, dass Surveys, Interviews, offizielle Statistiken und Wahrheitskommissionen essenzielle Quellen für die Friedens- und Konfliktforschung sind. Gleichwohl sind die aus ihnen resultierenden Daten in unterschiedlichem Maße von Verzerrungen betroffen und können somit nicht per se als sichere Alternative zu medienbasierten Konfliktdaten gelten.

4 Surveys und Interviews sind mit weiteren Verzerrungen behaftet wie beispielsweise mit Erinnerungsverzerrung (*recall bias*), sozialer Erwünschtheit (*social desirability bias*) oder Ermüdungserscheinungen durch Überforschung (Höglund/Oberg 2011).

Tabelle 1: Überblick der prominentesten medienbasierten Konfliktdatensätze inklusive Quellengrundlage und Umgang mit medialer Verzerrung

Name	Genutzte Quellen für Daten-erhebung	Transparenz über Verzerrungen in Publikationen
<i>Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED)</i>	<p><i>Hauptsächliche Quellen:</i></p> <ul style="list-style-type: none"> – weltweite Nachrichtenagenturen – Übersetzungen lokaler Zeitungen basierend auf BBC <p><i>Zusätzlich Quellen:</i></p> <ul style="list-style-type: none"> – NGO-Berichte, Fachbücher, Berichte von Wahrheitskommissionen etc. <p>60% der Ereignisse sind basierend auf Medienberichten kodiert (Croicu/Sundberg 2017: 12)</p>	<p><i>Codebook:</i> Croicu/Sundberg (2017)</p> <ul style="list-style-type: none"> – Informationen zu Quellenauswahl und Kodierungsprozess (S. 12, 13) – Verweis auf mediale Verzerrungen und wie mediale Verzerrungen statistisch adressiert werden können (S. 12, 13, FN 2) <p><i>Datensatzpapier:</i> Sundberg/Melander (2013) inklusive Appendix</p> <ul style="list-style-type: none"> – Verweis auf willkürliche Stichprobe und mediale Verzerrung (S. 527; Appendix S. 4, 5) – Vor- und Nachteile bestimmter Datenerhebungsprozedere (S. 527; Appendix S. 6, 7) – Kodierungsprozess (S. 526) <p><i>Zusätzliche Veröffentlichungen:</i></p> <ul style="list-style-type: none"> – Zur Datenerhebung: Sundberg/Harbom (2011) und Kreutz (2015) – Zur medialen Verzerrung allgemein: Öberg/Sollenberg (2011) – Herausforderungen medialer Verzerrungen und empirische Anwendung: Croicu/Kreutz (2017)
<i>Armed Conflict Location & Event Data Project (ACLED)</i>	<p><i>Hauptsächlich Quellen:</i></p> <ul style="list-style-type: none"> – lokale und nationale Medienberichte <p><i>Zusätzliche Quellen:</i></p> <ul style="list-style-type: none"> – NGO-Berichte – Afrika spezifische Berichterstattung (Raleigh/Dowd 2017) <p>Von den 10 meist genutzten Quellen, die 72% aller Ereignisse zugrunde liegen, sind 6 Nachrichtenagenturen (ACLED 2012a)</p>	<p><i>Codebook:</i> Raleigh/Dowd (2017)</p> <ul style="list-style-type: none"> – allgemeine Quellenauswahl <p><i>Handbuch für media user:</i> ACLED (2017)</p> <ul style="list-style-type: none"> – Verweis auf mediale Verzerrung und willkürliche Stichprobe (S. 9, 10) <p><i>Datensatzpapier:</i> Raleigh et al. (2010)</p> <ul style="list-style-type: none"> – Information zur Quellenauswahl (S. 656) – Verweis auf mediale Verzerrung Quellenauswahl (S. 656) <p><i>Zusätzliche Veröffentlichungen:</i></p> <ul style="list-style-type: none"> – Zur medialen Verzerrung in ACLED: ACLED (2015) – Zur Quellenangabe: ACLED (2012a) – Zu nicht-identifizierten Gruppen: ACLED (2012b)

Name	Genutzte Quellen für Daten-erhebung	Transparenz über Verzerrungen in Publikationen
<i>Social Conflict in Africa Database (SCAD)</i>	<p><i>Hauptsächliche Quellen:</i></p> <ul style="list-style-type: none"> – Nachrichtenagenturen <p>Alle Ereignisse wurden basierend auf <i>Associated Press</i> und <i>Agence France Presse</i> kodiert (Salehyan/Hendrix 2016)</p>	<p><i>Codebook:</i> Salehyan/Hendrix (2016)</p> <ul style="list-style-type: none"> – Informationen zu Quellenauswahl und Kodierungsprozess (S. 1-7) <p><i>Datensatzpapier:</i> Salehyan et al. (2012)</p> <ul style="list-style-type: none"> – Informationen zur Quellenauswahl und Kodierungsprozess (S. 505) – Verweis auf willkürliche Stichprobe (S. 505, 506) <p><i>Zusätzliche Veröffentlichungen</i></p> <ul style="list-style-type: none"> – <i>Best practice</i> zur Datenerhebung: Salehyan (2015) – Mediale Verzerrungen und empirische Anwendung: Hendrix/Salehyan (2015)
<i>Disaggregated Conflict Dataset at Heidelberg University (DISCON)</i>	<p><i>Hauptsächliche Quellen:</i></p> <ul style="list-style-type: none"> – Nachrichtenagenturen und Zeitungen <p><i>Zusätzliche Quellen:</i></p> <ul style="list-style-type: none"> – NGO-Berichte, akademische Publikationen (Trinn/Wencker 2016) 	<p><i>Codebook:</i> HIIK (2016)</p> <p><i>Datensatzpapier:</i> Trinn/Wencker (2016)</p> <ul style="list-style-type: none"> – Information zu Kodierungsprozess (S. 5) – Verweis auf mediale Verzerrung und Beobachtungseinheit Region-Monat statt Ereignis, um mediale Verzerrung zu reduzieren (S. 5)
<i>Konstanz One-Sided Violence Event Dataset (KOSVED)</i>	<p><i>Hauptsächliche Quellen:</i></p> <ul style="list-style-type: none"> – Nachrichtenagenturen und Zeitungen <p><i>Zusätzliche Quellen:</i></p> <ul style="list-style-type: none"> – NGO-Berichte (Schneider/Bussmann 2013) 	<p><i>Codebook:</i> Bussmann/Schneider (2012)</p> <p><i>Datensatzartikel:</i> Schneider/Bussmann (2013) inklusive Appendix</p> <ul style="list-style-type: none"> – Information zu Quellenauswahl und Kodierungsprozess (S. 638; Appendix S. 13-14) – Verweis auf mediale Verzerrung (S. 637)

Eine andere Option ist, die Herausforderung der medialen Verzerrung anzunehmen. Für die oben skizzierten Herausforderungen gibt es keine Patentlösung. Es stehen jedoch unterschiedliche Möglichkeiten bereit, wie Forschende, die Daten erheben und anwenden, mit medienbasierten Datensätzen umgehen können. Die *erste* Implikation betrifft die Forschungsteams, die Daten erheben. Transparenz hinsichtlich des Datenerhebungsprozesses ist unerlässlich. Das heißt, dass vorzugsweise im *codebook* die Quellenauswahl dokumentiert und auf die potenziellen Verzerrungen aufgrund der Quellenauswahl verwiesen werden sollte. Tabelle 1 zeigt einen Überblick über die Quellenauswahl und Dokumentation der Datenerhebungsprozesse für die fünf prominentesten nationalen und internationalen medienbasierten Datensätze zu Bürgerkriegen und politischer Gewalt. Die linke Spalte zählt die Quellen auf, die den einzelnen Datensätzen zugrunde liegen. Die rechte Spalte fasst zusammen, wie transparent Herausgeber_innen mit den Quellenangaben und den Herausforderun-

gen der medialen Verzerrungen umgehen.⁵ Im Großen und Ganzen machen die Herausgeber_innen von medienbasierten Konfliktdatensätzen ihre Hausaufgaben. Sie listen in der Regel ihre Quellen auf und verweisen zumindest darauf, dass mediale Verzerrung ein Problem sein könnte. Somit stellen sie die Informationen für die Anwender_innen bereit, die zur Einschätzung der medialen Verzerrung notwendig sind. Tabelle 1 verdeutlicht aber auch, dass die drei großen Datenprojekte (UCDP GED, ACLED und SCAD), die auch über größere finanzielle und personelle Ressourcen verfügen als DISCON und KOSVED, umfangreicher ihre Quellenauswahl, den Kodierungsprozess und die Herausforderungen und Lösungsmöglichkeiten herausarbeiten.

Die zweite Implikation betrifft diejenigen, die die Daten nutzen. Es liegt in der Verantwortung der Forschenden, darauf zu verweisen, ob die mithilfe von medienbasierten Datensätzen gewonnenen Ergebnisse durch mediale Verzerrungen beeinflusst sind. Forschende können theoretisch überlegen, ob und inwiefern ihre Analysen von der medialen Verzerrung betroffen sind. In einigen Fällen haben die aus der medialen Verzerrung resultierenden Messfehler nämlich keine Auswirkungen für die Anwendung medienbasierter Konfliktdatensätze. Für Regressionsmodelle sind Messfehler beispielsweise in der abhängigen Variable unproblematisch, wenn der Messfehler nicht mit der unabhängigen Variable korreliert. Die Schätzungen einfacher Regressionsmodelle bleiben in solchen Fällen unverzerrt und konsistent (Wooldridge 2013: 318). Des Weiteren sind Mikro-Studien in der Regel anfälliger für die Berichts- und Beschreibungsverzerrung als Makro-Studien. Eine Möglichkeit ist daher, die Analysen auf dem Meso-Level durchzuführen (z. B. Provinzebene), um die Verzerrungen gering zu halten, aber trotzdem sub-staatlichen Dynamiken nachgehen zu können (s. Weidmann 2013: 573; 2015: 1146).

Darüber hinaus gibt es statistische Möglichkeiten, die Konsequenzen der medialen Verzerrung aufzuzeigen bzw. abzufedern. Forschende können in ihren statistischen Modellen die Gründe der medialen Verzerrung kontrollieren, wie z. B. den Zugang zu Mobilfunknetzen und Internet oder die Nähe zu urbanen Zentren. Auch theoriegeleitete Diagnosetests können dabei helfen, herauszufinden, ob Ergebnisse möglicherweise durch mediale Verzerrung beeinflusst sind. Weidmann (2016) wendet beispielsweise einen Test an, der auf der Idee beruht, dass sehr gewaltreiche Ereignisse weniger von medialer Verzerrung betroffen sind. Eine weitere Möglichkeit sind sogenannte *Multiple System Estimations* (MSE). MSE nutzen die Überlappung zwischen mehreren willkürlichen Stichproben zu politischer Gewalt, um die Gesamtanzahl der Gewaltereignisse zu schätzen (z. B. Price/Ball 2015). Hendrix und Salehyan (2015) demonstrieren mithilfe solch eines Schätzverfahrens, dass die in SCAD enthaltenen Ereignisse 76 Prozent aller Gewaltereignisse in Afrika entsprechen. Diese Rate verbessert sich zunehmend, wenn einflussreichere Ereignisse berücksichtigt werden. Solche Verfahren gehören zu den wenigen, die es erlauben, die

⁵ Die Grundlage der Zusammenfassung bilden die aktuellen *codebooks*, die Datensatzpapiere, die mit den Daten in Fachzeitschriften veröffentlicht wurden sowie weitere Veröffentlichungen von Mitgliedern der Datenerhebungsteams.

Grundgesamtheit zu schätzen. Sie sind aber aufwendiger in ihrer Durchführung als die zuvor genannten Möglichkeiten.

Zusammenfassend lässt sich sagen, dass es wohl keine Datenquellen in der Friedens- und Konfliktforschung gibt, die nicht von Verzerrungen betroffen sind. Analog zur Debatte in der qualitativen Forschung über einen reflektierten Umgang mit beispielsweise aus Interviews gewonnenen Daten (Oettler 2008), plädiere ich für eine Quellenkritik in der quantitativen Forschung. Medienbasierte Konfliktdatensätze haben ihre Schwächen. Jedoch erlaubt ein reflektierter Umgang mit solchen Daten über Raum und Zeit vergleichende Analysen durchzuführen, die dann Generalisierungen über Friedens- und Konfliktprozesse zulassen.

Literatur

- Armed Conflict Location & Event Data Project (ACLED)* 2012a: Data Sources (ACLED Working Paper, July 2012), https://www.acleddata.com/wp-content/uploads/2014/12/ACLED_Sources-Working-Paper_July-2012_updated.pdf; 22.2.2018.
- ACLED* 2012b: Unidentified Armed Groups (ACLED Working Paper, July 2012), https://www.acleddata.com/wp-content/uploads/2012/07/ACLED_Unidentified-Armed-Groups-Working-Paper_July-2012.pdf; 22.2.2018.
- ACLED* 2015: Reporting Sources (ACLED Working Paper No. 5, 2015), Sussex.
- ACLED* 2017: Armed Conflict Location & Event Data Project (ACLED), Guide for Media Users, http://www.acleddata.com/wp-content/uploads/2017/01/ACLED_User-Guide-for-Media_2017.pdf; 22.2.2018.
- Bussmann, Margit/Schneider, Gerald* 2012: Konstanz One-Sided Violence Event Dataset (KOSVED), Codebook Version 1, Konstanz.
- Chojnacki, Sven/Ickler, Christians/Spies, Michael/Wiesel, John* 2012: Event Data on Armed Conflict and Security: New Perspectives, Old Challenges, and Some Solutions, in: International Interactions 38: 4, 382-401.
- Chojnacki, Sven/Reisch, Gregor* 2012: Perspectives on War. Generation Eventdata. Was zeigen Ereignisdaten, worin unterscheiden sie sich und was können wir lernen?, in: Sicherheit und Frieden 30: 1, 36-43.
- Croicu, Mihai/Kreutz, Joakim* 2017: Communication Technology and Reports on Political Violence: Cross-National Evidence Using African Events Data, in: Political Research Quarterly 70: 1, 19-31.
- Croicu, Mihai/Sundberg, Ralph* 2017: UCDP Georeferenced Event Dataset Codebook Version 17.1. Department of Peace and Conflict Research, Uppsala.
- Earl, Jennifer/Martin, Andrew/McCarthy, John D./Soule, Sarah A.* 2004: The Use of Newspaper Data in the Study of Collective Action, in: Annual Review of Sociology 30: 1, 65-80.
- Eck, Kristine/Hultman, Lisa* 2007: One-Sided Violence Against Civilians in War: Insights from New Fatality Data, in: Journal of Peace Research 44: 2, 233-246.
- Hawkins, Virgil* 2011: Media Selectivity and the Other Side of the CNN Effect: The Consequences of Not Paying Attention to Conflict, in: Media, War & Conflict 4:1, 55-68.
- Hendrix, Cullen S./Salehyan, Idean* 2015: No News Is Good News: Mark and Recapture for Event Data When Reporting Probabilities Are Less Than One, in: International Interactions 41: 2, 392-406.
- Heidelberger Institut für Internationale Konfliktforschung (HIIK)* 2016: Disaggregated Conflict Dataset (DISCON) 1.0 (Codebook Version 2-2016), Heidelberg.

- Höglund, Kristine/Öberg, Magnus (Hrsg.) 2011: Understanding Peace Research: Methods and Challenges, London.
- Kalyvas, Stathis N. 2004: The Urban Bias in Research on Civil Wars, in: Security Studies 13: 3, 160-190.
- Kende, Istvan 1978: Wars of Ten Years (1967-1976), in: Journal of Peace Research 3: 15, 227-239.
- Koopmans, Ruud 1999: The Use of Protest Event Data in Comparative Research: Cross-National Comparability, Sampling Methods and Robustness, in: Acts of Dissent: New Developments in the Study of Protest, Berlin, 90-110.
- Kreutz, Joakim 2015: The War that Wasn't There: Managing Unclear Cases in Conflict Data, in: Journal of Peace Research 52: 1, 120-124.
- Öberg, Magnus/Sollenberg, Margareta 2011: Gathering Conflict Information Using News Resources, in: Höglund, Kristine/Öberg, Magnus (Hrsg.): Understanding Peace Research: Methods and Challenges, London, 47-73.
- Oettler, Anika 2008: Do Qualitative Data Help in Addressing Central American Violence? Research Note on Data Collection (GIGA Working Paper 76/2008), Hamburg.
- Otto, Sabine 2013: Coding One-sided Violence from Media Reports, in: Cooperation and Conflict 48: 4, 557-566.
- Pickel, Susanne/Pickel, Gert 2012: Die Messung von Indizes in der Vergleichenden Politikwissenschaft – methodologische Spitzfindigkeit oder substantielle Notwendigkeit, in: Zeitschrift für Vergleichende Politikwissenschaft 6: 1, 1-17.
- Price, Megan/Ball, Patrick 2015: The Limits of Observation for Understanding Mass Violence, in: Canadian Journal of Law and Society 30: 2, 237-257.
- Raleigh, Clionadh/Dowd, Caitriona 2017: Armed Conflict Location and Event Data Project (ACLED) Codebook 2017, Sussex.
- Raleigh, Clionadh/Linke, Andrew/Hegre, Harvard/Karlsen, Joakim 2010: Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature, in: Journal of Peace Research 47: 5, 651-660.
- Ruhe, Constantin/Schneider, Gerald/Spilker, Gabriele 2017: Quantitative Methoden in den Internationalen Beziehungen, in: Sauer, Frank/Masala, Carlo (Hrsg.): Handbuch Internationale Beziehungen, Wiesbaden, 1-20.
- Salehyan, Idean 2015: Best Practices in the Collection of Conflict Data, in: Journal of Peace Research 52: 1, 105-109.
- Salehyan, Idean/Hendrix, Cullen 2016: Social Conflict Analysis Database (SCAD) Version 3.2, Codebook and Coding Procedures, Austin, TX.
- Salehyan, Idean/Hendrix, Cullen S./Hamner, Jesse/Case, Christina/Linebarger, Christopher/Stull, Emily/Williams, Jennifer 2012: Social Conflict in Africa: A New Database, in: International Interactions 38: 4, 503-511.
- Sambanis, Nicolas 2004: What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition, in: The Journal of Conflict Resolution 48: 6, 814-858.
- Schneider, Gerald/Bussmann, Margit 2013: Accounting for the Dynamics of One-Sided Violence: Introducing KOSVED, in: Journal of Peace Research 50: 5, 635-644.
- Schneider/Gerald 2015: Von Makro zu Mikro: Grundlagen und Perspektiven der Bürgerkriegsforschung, in: Zeitschrift für Friedens- und Konfliktforschung 4: 2, 308-329.
- Sundberg, Ralph/Harbom, Lotta 2011: Systematic Data Collection: Experiences from the Uppsala Conflict Data Program, in: Höglund, Kristine/Öberg, Magnus (Hrsg.): Understanding Peace Research: Methods and Challenges, London, 91-113.
- Sundberg, Ralph/Melander, Erik 2013: Introducing the UCDP Georeferenced Event Dataset, in: Journal of Peace Research 50: 4, 523-532.
- Trinn, Christoph/Wencker, Thomas 2016: Introducing the Heidelberg Approach to Conflict Research, in: European Political Science (Online First), <https://doi.org/10.1057/s41304-016-0093-4>.

- Weidmann, Nils B.* 2013: The Higher the Better? The Limits of Analytical Resolution in Conflict Event Datasets, in: Cooperation and Conflict 48: 4, 567-576.
- Weidmann, Nils B.* 2015: On the Accuracy of Media-based Conflict Event Data, in: Journal of Conflict Resolution 59: 6, 1129-1149.
- Weidmann, Nils B.* 2016: A Closer Look at Reporting Bias in Conflict Event Data, in: American Journal of Political Science 60: 1, 206-218.
- Wigmore-Shepherd, Daniel* 2014: Urban Bias in Media Reporting, <http://www.crisis.acleddata.com/urban-bias-in-media-reporting/>; 15.9.2017.
- Wooldridge, Jeffrey M.* 2013: Introductory Econometrics: A Modern Approach, 5. Auflage, Manson, OH.