

# Zur Aussagefähigkeit von Web-basierten Datenerhebungen und -analysen

*Beurteilung und gute Handlungspraxis am Beispiel von Civey*

*Verfasst von:  
Prof. Dr. Bouncken<sup>1</sup>*

---

<sup>1</sup> Prof. Dr. Ricarda Bouncken hat den Lehrstuhl für Strategisches Management und Organisation der Universität Bayreuth inne. Ihre Forschung beschäftigt sich mit der empirischen Analyse von Wirkungsbeziehungen in und zwischen Organisationen sowie denen mit ihrem und innerhalb ihres Umfeldes. Weit über 200 Publikationen, meist in internationalen referierten Zeitschriften, aber auch in Büchern belegen ihre Expertise.

# Zur Aussagefähigkeit von Web-basierten Datenerhebungen und -analysen

*Beurteilung und gute Handlungspraxis am Beispiel von Civey*

## Inhaltsverzeichnis

Executive Summary .....	3
1. Einleitung .....	1
2. Hintergrund .....	2
3. Kurzbeschreibungen und -beurteilungen gängiger Erhebungsinstrumente .....	8
a. Face-to-Face Befragungen.....	8
b. Telefonbefragungen.....	9
c. Brieflich-postalische und E-Mail-Befragungen.....	10
d. Online-Befragungen .....	10
4. Kurzbeschreibung der Erhebungs- und Analysepraxis bei Civey .....	14
5. Kriterien-geleitete Beurteilung der typischen Vorgehensweise bei Civey.....	17
a. Vorbemerkungen zur Aussagequalität .....	17
b. Rücklauf bzw. Response Rate .....	18
c. Kosten und Aktualität.....	21
d. Technische Voraussetzungen, Multi-medialität und Darstellung.....	25
e. Mehrfacherhebung und Adaption.....	28
f. Zugang zu besonderen Gruppen .....	28
g. Repräsentativität im engeren Sinne .....	30
h. Soziale Erwünschtheit und Verzerrungen .....	34
i. Verbindung mit Online-Panels .....	35
j. Probabilistische und nicht-probabilistische Stichproben.....	37
6. Gesamtbeurteilung.....	41
Literaturverzeichnis .....	47

# Zur Aussagefähigkeit von Web-basierten Datenerhebungen und -analysen

*Beurteilung und gute Handlungspraxis am Beispiel von Civey*

## **Executive Summary**

*Online-Befragungen zur Meinungs- und Marktforschung bieten die Möglichkeit schnell und mit relativ hohen Responseraten Ergebnisse zu erzielen und daraus Aussagen und Trends abzuleiten. Allerdings wird Online-Befragungen auch oft entgegengehalten, dass sie nicht repräsentativ sein. Angesichts der hohen Vielfalt bei Online-Befragungen neigt diese Kritik tendenziell zu Pauschalisierung und ist angesichts der ständig ansteigenden Nutzung des Internets und der teils sehr ausgefeilten und gut kombinierten Verfahren bei Online-Befragungen auch oft nicht mehr zeitgemäß und zutreffend. Dieses Gutachten zieht verschiedenste wissenschaftliche und praxisgeprägte Kriterien sowie Studien heran, um die Aussagekraft von Online-Erhebungen und ihren gängigen Komponenten sowie um die Aussagekraft der typischen Systematik bei Civey zu bewerten. Um einen gemeinsamen Begriffskanon herzustellen, führt dieses Gutachten zunächst die typischen Begrifflichkeiten und Vorgehensweisen bei Befragungen für die Markt- und Meinungsforschung aus. Dann wird das Verfahren von Civey kurz vorgestellt um das spezifische Analyseobjekt zu vermitteln. Civey kombiniert typischerweise verschiedene Methoden wie Riversampling, Panelerhebungen und damit verbunden probabilistische und nicht-probabilistische Verfahren sowie Quoten und Gewichtungen, wenn es um Aussagen für eine Gesamtpopulation wie bei der Meinungs- und besonders Wahlforschung geht. Mittels der Kriterien wird die Vorgehensweise bei Civey daraufhin Schritt für Schritt beurteilt. Fazit der Beurteilung ist, dass gerade die durchdachte und gut gestaltete Methodenkombination aus probabilistischen und nicht-probabilistischen Verfahren im Regelfall eine hohe Aussagekraft und Repräsentativität auf der Basis von Online-Befragungen und insbesondere einer sehr durchdachten und kontrollierten Systematik seitens Civey erlaubt. Wenn man nach einer Schwäche an dem Verfahren sucht, könnte man anführen, dass die wenigen Nicht-Nutzer des Internets unberücksichtigt bleiben. Dieser Mangel lässt sich aber einfach durch die Kombination mit Offline-Erhebungen beseitigen und etwaige negative Wirkungen werden auch durch die Gewichtungen und Quoten stark minimiert.*

# 1. Einleitung

Angesichts der immer stärkeren Nutzung des Internets von Individuen und Organisationen, sinkenden Responseraten bei traditionellen offline Meinungs-, Wahl- und Marktforschungen sowie der Potentiale durch neue Verfahren, die Elemente der künstlichen Intelligenz nutzen, stellt sich die Frage, wie aussagekräftig Online-Befragungen heutzutage sind und wie sie mittlerweile in Punkto Qualität und Repräsentativität beurteilt werden können.

Dieses Gutachten setzt hier an und beurteilt die Befragungs- und Auswertungssystematik seitens des Instituts Civey spezifisch und vor dem Hintergrund einer generellen Beurteilung von Online-Befragungsformen bzw. ihren gängigen Komponenten auf der Basis umfangreicher Literatur- und Studienanalyse. Wichtig ist dabei vorzuschicken, dass eine pauschale Kritik in Sinne von ‚Online-Befragungen sind nicht repräsentativ‘ eben gerade nicht für spezielle Befragungs- und Auswertungssystematiken von Anbietern von Online-Befragungen gelten können. Online-Befragungen weisen eine hohe Breite auf und sind damit auch mit Varianz ihrer Aussagekraft für bestimmte Populationen verbunden. Online-Befragungen können unterschiedliche multimediale Gestaltungsformen nutzen. Sie können gezielt auf anderen Webseiten integriert oder beworben werden. Sie können in einem Extremfall eines Riversampling Befragte mittels Banner, Pop-up Fenstern auf Webseiten Ergebnisse generieren, ohne sozio-demografische Merkmale der Befragten mit zu erheben, aber dabei meist sehr hohe Responseraten erreichen. Online-Befragungen müssen nicht isoliert ablaufen und sie lassen sich auch mit verschiedensten on- und offline Methoden zur Rekrutierung der Teilnehmer kombinieren, z. B. Einladungen versenden in Form von Briefen, Telefon oder E-Mail. Online-Befragungen bieten die Möglichkeit mit dem Aufbau und Nutzung von Panels verbunden zu werden. Online-Befragungen erlauben es Riversampling mit der Erhebung von sozio-demografischen Merkmalen zu verbinden und Quoten und Gewichtungen vorzunehmen und dabei klassische propabilistische mit Nicht-propabilistischen Verfahren zu kombinieren.

Verschiedene Kriterien aus der traditionellen Statistik sowie seitens von Berufsverbänden können angelegt werden, um Online-Befragungs-Praktiken mittels verschiedener Kriterien zu beurteilen. Die Aussagekraft und Repräsentativität von Online-Befragungen ist damit differenziert in Abhängigkeit der jeweiligen Kriterien, der Methoden(kombination) und der Gestaltung der Befragung zu evaluieren. Aufgrund der Vor- und Nachteile von Datenerhebungs- und -analyseverfahren in bestimmten Anwendungskontexten, wird dieses Gutachten verschiedene Kriterien zur Beurteilung diskutieren und anwenden. Die Beurteilung richtet sich dabei zum einen generell an die Aussagefähigkeit von Online-Befragungen und deren Datenanalyse sowie speziell an die Praxis im Unternehmen Civey.

Auf der Basis einer hohen Anzahl vorherigen Studien und unterschiedlichen Studien und Kriterien analysiert dieses Gutachten daher die Aussagekraft von gängigen Komponenten und Praktiken von Online-Befragungen und die Systematik bei Civey im Speziellen. Dieses Gutachten beurteilt die typische Vorgehensweise bei Datenerhebungen und -analysen des Unternehmens Civey vor dem Hintergrund der generellen Stärken und Schwächen von Online-Befragungsinstrumenten nicht isoliert, sondern auch im Vergleich zu Face-to-Face, Telefon und E-Mail. Dabei werden die speziellen Routinen bei Civey zur Stärkung der Aussagefähigkeit und Repräsentativität der Aussagen sowie der speziellen Schätzalgorithmen sowie die Nutzung von ersten Ansätzen der künstlichen Intelligenz bei der Datenanalyse herangezogen.

Etwas generalisierend lässt sich vorausschicken, dass eine durchdachte Gestaltung der Datenerhebung und -verarbeitung bei Online-Befragungen, wie sie eben bei Civey vorliegt, eine hohe Aussagekraft und Repräsentativität erzielen kann. Die Aussagekraft und Repräsentativität der Systematik bei Civey

wird vor allem hergestellt durch eine gute audio-visuelle Gestaltung der Befragung (mit Feedbackmethoden wie z. B. einer Landkarte bei der Wahlforschung) in Verbindung mit einer Methodenkombination aus dem Riversampling, der Erhebung von sozio-demografischen Merkmalen, dem Anlegen von Quotierungen und Gewichtungen sowie der Akquise und Nutzung von Panelteilnehmern. Hohe Aussagekraft und Repräsentativität von Civey wird so letztlich aus einem Methodenkanon aus probabilistischen und nicht-probabilistischen Auswahlverfahren sowie entsprechenden Analysemethoden hergestellt. Einzige Schwäche an dem Verfahren von Civey ist, dass die wenigen Nicht-Nutzer des Internets eben nicht direkt online befragt werden können. Hier ist eine Ausweitung der Praktiken um eine Kombination mit brieflicher oder telefonischer Ansprache zu empfehlen, ggf. mit Interviewern, die die Daten dann Online eingeben. Allerdings wird die zunehmende Durchdringung des Internets immer diese Einschränkung bzw. Aktivitäten dagegen erfordern bis ganz überflüssig machen.

Überraschend ist, dass das Unternehmen Civey in der Kommunikation nach außen die Nutzung von Riversampling und nicht-probabilistischen Stichprobenverfahren überbetont und sich damit angreifbarer für Kritik über mangelnde Repräsentativität macht, als die Erhebungs- und Auswertungssystematik es rechtfertigt. Während also wenig Kritik an der Erhebungs- und Auswertungssystematik bei Civey zu äußern ist, kann diese eher an der Außenkommunikation von Civey geführt werden. Mehr Klarheit im Sinne der Nutzung von probabilistischen und nicht-probabilistischen Auswahlverfahren ist zu empfehlen.

Das Gutachten gliedert sich dabei in fünf Hauptteile. Begonnen wird im Anschluss an die Einleitung mit dem Hintergrund der Entwicklung von Online Datenerhebungen. Weiter geht es mit der Beschreibung der gängigen Verfahren zur Befragung, die bereits dann schon kurz beurteilt werden. Nach der Kurzvorstellung von Civey werden verschiedene Perspektiven und Kriterien der Beurteilung dargestellt und daran Online-Erhebungen generell und bei Civey im Speziellen beurteilt. Die Beurteilung zieht bereits vorliegende Analysen und Vergleichsstudien zur Beurteilung und Diskussion heran. Der letzte Teil umfasst eine Gesamtbeurteilung und zieht ein Fazit. Es schließt sich ein umfassendes Literaturverzeichnis an.

## **2. Hintergrund**

Die vergangene Dekade weist eine starke und ständig zunehmende und immer vielfältigere Nutzung des Internets inklusive der digitalisierten Kommunikation, meist mit Unterstützung von automatisierten Systemen und jüngst der zunehmenden Nutzung von digitalen Plattformen auf (Baker et al., 2010; Callegaro, Villar, Krosnick, & Yeager, 2014; Duffy, Smith, Terhanian, & Bremer, 2005; Struminskaya, Kaczmirek, Schaurer, & Bandilla, 2014). Gleichzeitig findet auch die Kommunikation zwischen Individuen sowie die zwischen Unternehmen untereinander und mit ihren Stakeholdern immer stärker über digitale Medien statt und bedient sich der Vorteile, die audio-visuelle Austauschformate erlauben (Chu & Manchanda, 2016; Reischauer & Mair, 2018). Briefliche und selbst telefonische Kommunikation werden dadurch tendenziell verdrängt.

Die letzten Jahre sind daher für die Meinungs-, Markt- und Wahlforschung gekennzeichnet durch zum einen ein stärkeres Ansteigen der Nutzung von Web-basierten Anwendungen und Plattformen seitens der Unternehmen sowie seitens der Bevölkerung (Baker et al., 2010; Callegaro et al., 2014; Duffy et al., 2005; Struminskaya et al., 2014). Zum anderen lässt sich eine schlechtere Erreichbarkeit und Rückgang der Bereitschaft von Individuen an persönlichen und Telefonerhebungen teilzunehmen konstatieren (Chang & Krosnick, 2009; Duffy et al., 2005). Andere Mediennutzung und Erreichbarkeit

stellen daher Probleme, aber auch Chancen für die Datengewinnung und -analyse dar (Duffy et al., 2005). Um angesichts der veränderten Bedingungen eine hohe Aussagegüte von Befragungen statistischen Auswertungen herzustellen, geht die Meinungs-, Markt- und Wahlforschung neue Wege zur Datenerhebung und -auswertung (Chang & Krosnick, 2009).

Neue Potenziale auf der Seite der Datenerhebung ergeben sich durch Online-Erhebungen (Chang & Krosnick, 2009), wie etwa Befragungen auf Plattformen, Portalen oder Webformen oder durch automatisierte Texterhebungen (meist Computer aided text analysis: CATA). Diese sind in den USA schon seit längerer Zeit im Einsatz. In Deutschland zählt das Unternehmen Civey gewissermaßen zu den Vorreitern von Online-Befragungen. Im Hinblick auf die Datenanalyse erlaubt gerade die Künstliche Intelligenz neue durch sie unterstützte Berechnungen. So verwundert es auch nicht, dass jüngst Online-Befragungen, CATA (Fisher, 2018; Schumann & Klein, 2015; Serfling, Kastl, & Sengül, 2018) und Big Data Analysen (Kraus, 2013; Wamba et al., 2017) intensiv und auch kontrovers in der Unternehmenspraxis, Wissenschaft und Bevölkerung diskutiert werden.

Die Unternehmenspraxis und die Meinungs-, Markt- und Wahlforschung wie auch die wissenschaftliche Forschung hat sich seit der Jahrtausendwende mit der Entwicklung von neuen elektronisch basierten Befragungsformen beschäftigt (Andrews, Nonnecke, & Preece, 2007; Buttler & Christian, 2000; Duffy et al., 2005; Evans & Mathur, 2005; Fouladi, McCarthy, & Moller, 2002; Hofmann, Steinmeyer, & Paul, 2001; Ilieva, Baron, & Healey, 2002; Schaefer & Dillman, 1998; Wirtz, Schmidt-Holz, & Beaujean, 2004; Wright, 2005) wobei zu betonen ist, dass Studien seit ca. 2013 von einer immer stärkeren Repräsentativität der Aussagen von Online-Befragungen für die Bevölkerung ausgehen. In Deutschland nahm die Online-Forschung bereits 2009 einen Anteil von 31% ein (Treiblmaier, 2011). Aus der Vielzahl der Möglichkeiten haben sich eine Reihe von Befragungsinstrumenten herauskristallisiert, die zur Datenerhebung und -analyse für die Meinungs-, Marktforschung sowie für weitere Prognoseinstrumentarien von Organisationen eingesetzt werden.

Wissenschaftliche Studien haben traditionell elektronische Befragungen zunächst unterschieden in E-Mail basierte (Fricker, 2008; Schaefer & Dillman, 1998; Yun & Trumbo, 2000) und online-basierte Verfahren unterschieden (Evans & Mathur, 2005; Fouladi et al., 2002; Hofmann et al., 2001; Wright, 2005). Jüngst werden noch Texterkennungsverfahren und erste Verfahren zur Beurteilung von Nutzungszeiten und anderen Daten mittels Big Data und der Erkennung und Extrapolation von Muster eingesetzt (McKenny, Aguinis, Short, & Anglin, 2018; Niculescu & Thorsteinsson, 2011). Die zuletzt genannten Texterkennungsverfahren sind grundverschieden zu den zuerst genannten Erhebungsverfahren, weil sie nicht auf aktiver Beantwortung von Personen beruhen, sondern vorliegende Texte und Narrative untersuchen (Bartel & Garud, 2009; McKenny et al., 2018). Anstelle von Primärerhebungen analysieren sie vorliegende Daten und Informationen und werden daher den Sekundäranalysen zugeordnet. Die Texterkennungsverfahren werden in dieser Beurteilung nicht weiter berücksichtigt.

E-Mail basierte Erhebungen werden auch als asynchrones Verfahren eingestuft (Fricker, 2008; Ocker & Yaverbaum, 1999): eine zu befragende Person erhält (wie bei einem Brief) eine E-Mail mit der Bitte den angehängten Fragebogen oder das Fragebogen Dokument hinter dem Link auszufüllen. Das Ausfüllen erfolgt nach der Ansprache, kann auch schrittweise erfolgen und erzeugt typischerweise ein Dokument, das danach in eine Datenbank importiert wird. Dagegen werden Online-Befragungen direkt und ggf. dynamisch auf Webseiten, Foren, Plattformen u. ä. implementiert (Callegaro et al., 2014; Szolnoki & Hoffmann, 2013). E-Mail Befragungen pushen demzufolge die Fragen zu der Person während Online-Befragungen tendenziell eine Pull-Funktion bei dem interessierten Nutzer im Internet nutzen.

Die beiden elektronischen Befragungsverfahren greifen auf unterschiedliche Datenbanktechnologien zurück. E-Mail basierte Fragebögen nutzen im Regelfall zuvor erhobene E-Mail Adressen, die um Verifikationen und Links ergänzt werden und die zeitverzögert beantwortet werden. Bei Online-Befragungen antworten die Personen im Regelfall direkt im Internet und ihre Antworten werden zeitgleich verifiziert und gespeichert (Andrews, Nonnecke, & Preece, 2003). Dabei lässt sich aber auch mit der Verlinkung von E-Mail Ansprache zu Online Fragen eine Verbindung von beiden Verfahren herstellen. Zunächst wird die E-Mail Ansprache verwendet, dann werden die Daten synchron bei der Beantwortung in eine Datenbank gespeichert. Dies bietet auch den Vorteil bestimmte Bevölkerungsgruppen über Vorselektionskriterien wie z. B. Demografiemerkmale direkt anzusprechen und dann die Vorteile von Online-Verfahren und deren gestalterischen Potentialen zu nutzen. Darüber hinaus ermöglicht die Nutzung von (ggf. zuvor aufgebauten) Panels bei Online-Erhebungen eine Ansprache von Befragten nach Quoten hinsichtlich soziodemografischer Kriterien. Es liegt somit eine Kombination von nicht-probabilistischen und probabilistischen Ansätzen vor, die die Vorteile beider Auswahlverfahren nutzt.

Online-Befragungen werden besonders dann hohe Responseraten erzielen, wenn die Befragung sinnvoll in Webseiten integriert wird, die Befragungsseite bzw. das Fenster so gestaltet ist, dass sie Interesse hervorruft und die Fragen bedeutsam oder attraktiv sind. Bei Online-Befragungen lassen sich dazu viele verschiedene gestalterische Ansätze, Spiele und Gewinnmöglichkeiten nutzen (Wu, Chen, & Cho, 2013). Die Einbindung von Bannern oder pop-up Fenstern auf Webseiten soll die Personen einladen teilzunehmen. Interessierte Personen klicken dann direkt auf die Befragung, beantworten die Fragen, sodass Daten zeitgleich gespeichert werden (Callegaro et al., 2014).

Viele Markt-, Meinungs- und Wahlforschungsinstitute nutzen für ihre Online-Befragungen den Riversampling-Ansatz. Im engeren Sinne spricht man von Riversampling, wenn Befragte mittels Banner, Anzeigen, Werbung, Offerten und anderer Einladungsformen auf Webseiten motiviert werden an einer Befragung teilzunehmen. Wenn sie reagieren, werden ihnen zunächst Vorauswahlfragen präsentiert und sie danach zu bestimmten Befragungen weitergeleitet. In der Reinform des Riversampling ist dem Erhebenden nur bekannt, auf welchem Weg die Teilnehmer zur Befragung geleitet werden jedoch nicht wer antwortet, so dass keine weiteren Charakteristika über den Befragten bekannt sind oder erhoben werden. Daher erntet diese Extremformen eines Riversampling Ansatzes viel Kritik im Hinblick auf die Repräsentativität der Stichprobe, wenn um Aussagen über die Gesamtpopulation oder gar eine nationale Bevölkerung geht. Gerade in der Meinungsforschung und politischen Forschung ist daher die Extremform eines Riversampling Ansatzes, in der keine soziodemografischen Merkmale erhoben werden mit hohen Aussageproblemen für nationale Grundgesamtheiten verbunden. Meist werden aber diese Riversampling Ansätze mit der Frage nach soziodemographischen Charakteristika der Befragten kombiniert. In der Extremform eines Riversampling, das dann nicht-probabilistische Verfahren einsetzen muss, wissen die Erhebenden nicht wer antwortet und kontaktieren den Befragten auch nicht absichtlich erneut. In der Datenanalyse kommen dann Gewichtungen und Quoten zum Einsatz. Hierdurch können Institute dann eine sehr viel bessere Repräsentativität herstellen.

Andere Institute wie z. B. Civey nutzen den Riversampling Ansatz zum Aufbau und Einsatz von Online-Panels. Der Riversampling Ansatz wird zunächst für die Rekrutierung eingesetzt. Die Befragten werden in diesem Fall nach weiteren z. B. soziodemographischen Charakteristika befragt und dann zufällig für weitere Befragungen ausgewählt, sodass hier eine hohe Überschneidung zu probabilistischen Ansätzen vorliegt. Panelansätze werden generell in der elektronischen Meinungs- und Marktforschung verwendet, um schnell an die Befragung dieser interessierten Personengruppe heranzukommen (Struminskaya et al., 2014). Civey verwendet Riversampling auch um für neue Befragungen innerhalb

des mittlerweile der großen Panels Teilnehmer zu rekrutieren. Die Verbindung von Riversampling, Online-Panel und Quoten/Gewichtungen ermöglicht innerhalb kürzester Zeit eine zufällige und zielgruppenspezifische Auswahl von Teilnehmern für Befragungen zu treffen/durchzuführen und dabei auch repräsentative Antworten für die Meinungs- Wahl- und Marktforschung zu erhalten.

Mit zunehmender Vereinfachung der Datenerhebungs- und -analysemethoden nehmen Zeitbedarf und Kostenbedarf der Online-Befragungen stark ab. Dies kann natürlich dazu führen, dass zu schnell und unreflektiert Online-Befragungen angegangen werden, gerade wenn reines Riversampling integriert wird und keine weiteren sozio-demografischen Merkmale der Befragten bekannt sind. Es bedarf fundierter Beurteilungen von Online-Befragungen und guter Handlungspraxis in diesem Feld. Diese sind vor dem Hintergrund spezifischer Besonderheiten im Bereich der Technologie, Demografie und Rücklaufquoten zu formulieren (Andrews et al., 2003). Allgemeingültige Hinweise aus der Wissenschaft wie Online-Erhebungen gestaltet werden sollten, wie sie verwendet und implementiert werden sollten, sind eher noch rar (Ausnahme z. B. Ansolabehere & Schaffner, 2014; Kennedy et al., May 2016; Struminskaya et al., 2014). Allerdings liegen eine Reihe von Vergleichsstudien zur Aussagegüte von Online-Erhebungen vor (Callegaro et al., 2014; Szolnoki & Hoffmann, 2013). In der Praxis dagegen sind schon viele gestalterisch ansprechende und vielgenutzte Online-Befragungen seitens Unternehmen, Befragungsinstituten sowie wissenschaftlichen Instituten zu finden, die unterschiedliche Praktiken bei der Stichprobenauswahl, bei der Nutzung von Panels, bei der Gewichtung der Daten sowie bei der Datenauswertung nutzen. Allerdings finden sich auch praxisgeprägte Empfehlungen zu Qualitätsstandards (ADM, Mai 2001; ESOMAR, 2015).

Trotz oder gerade aufgrund der verschiedenen Online-Befragungsformate und Methodenkombination stellt sich die Frage wie gut aussagekräftige und repräsentative Datenerhebungen über dieses Medium durchführbar sind. Diese Frage ist schon bereits zu Beginn der Ausweitung des Internets und der Zunahme der Kommunikation von Individuen und Organisationen über das Internet gestellt worden (Schmidt & Computers, 1997). Geringe Kosten, Schnelligkeit, Zielgruppenspezifität, zu vernachlässigbare Interviewer-Bias und die reichhaltigen Möglichkeiten audio-visueller Darstellungen sowie schnelles, teils sogar unmittelbares Feedback an die Befragten gelten als zentrale Vorteile von Datenerhebungen und -analysen über das Internet (Schleyer & Forrest, 2000; Schmidt & Computers, 1997).

Online-Erhebungsinstrumente haben neben ihren Vorteilen jedoch auch Schwachpunkte und Andersartigkeiten (Yun & Trumbo, 2000). Im Vordergrund der Kritik an Online-Befragungen steht die Stichprobenauswahl bzw. deren Repräsentativität für Populationen und Zielgruppen, besonders nationale Grundgesamtheiten (Blasius & Brandt, 2009, 2010; Buttler & Christian, 2000; Duffy et al., 2005). Dies gilt gerade für die Extremform des Riversampling, bei der keine soziodemografischen Merkmale erhoben und keine Gewichtungen vorgenommen werden. In der Tat werden zudem in Online-Stichproben Personen, die keinen Zugang zum Internet haben, vermutlich nicht zufällig fehlen (Callegaro et al., 2014). Dies darf aber nicht dazu führen zu übersehen, dass die Repräsentativität von Online-Erhebungen auch durch bestimmte Korrekturen und Methodenkombinationen (z. B. multimediale Gestaltung der Befragung, ganz gezielte Integration in Foren für Randgruppen, Feedback an den Befragten, Wahl der Foren/Plattformen, Einladung über andere Medien entlang von Bevölkerungsstatistiken, Erhebung, Quotierung und Gewichtung von sozio-demografischen Merkmalen) positiv beeinflusst werden kann. Obendrein könnten die Vorteile von Online-Befragungen (Zielgruppenspezifität, zu vernachlässigbare Interviewer-Bias und die reichhaltigen Möglichkeiten audio-visueller Darstellungen sowie schnelles, teils sogar unmittelbares Feedback an die Befragten) auch etwaige Nachteile des Verfahrens anders ausgleichen und die Aussagekraft stärken. Überdies existieren auch Studien, die eine hohe Repräsentativität von Online-Befragungen belegen, weil sie

Zugang zu sonst schwer zugänglichen Bevölkerungsgruppen erlauben. So bieten sie Zugang zu einer großen Breite an Personen mit Varianz auf den Merkmalen wie bei Geschlecht, sozio-ökonomischer Status, Ort, Alter und kultureller Hintergrund, physischen Einschränkungen usw. (Gosling, Vazire, Srivastava, & John, 2004).

Für Individuen, Organisationen und die Wissenschaft stellt sich daher die Frage, für welche Ziele und Kontextbedingungen welche Form der Online-Befragung sinnvoll sind, bzw. welche Vorgehensweise die Qualität absichert und welche Methodenkombinationen vorteilhaft sind. Im Gegenzug müssen auch die Limitationen und Verbesserungsbedarfe analysiert und bedacht werden. Eine kontinuierliche Weiterentwicklung der Verfahren, Methodenkombination und intensiverer Nutzung des Internets in der Bevölkerung verspricht weitere Potentiale in der Zukunft. Herauszustellen ist bereits hier, dass die Frage nach besseren oder schlechteren Erhebungsmethoden kaum generell zu beantworten ist. Die jeweiligen Ziele, Bedingungen und die Durchdachtheit der Vorgehensweise und Methodenkombination entscheidet über die Aussagekraft von Datenerhebungen, auch von denen im Internet. Die Andersartigkeit von Online-Befragung liegt nicht allein in der elektronischen Präsentation der Fragen und ihrer Übertragung, sondern auch in der Frage der Synchronität und dem Zugang zu bestimmten Bevölkerungs- und Zielgruppen.

Zur Aussagekraft und Vorziehenswürdigkeit einzelner Methoden und Methodenkombinationen für die Datenerhebung und -analyse wurden verschiedenste Vergleiche zwischen den Erhebungs- und Auswertungsformen (z. B. Face to Face vs. Telefon, Face to Face vs. E-Mail, Online vs. Face-to-Face oder Online vs. Telefon) sowie entlang unterschiedlicher Kriterien (z. B. Responseraten, Repräsentativität, Aktualität, soziale Erwünschtheit, kognitive Überlastung) geführt (Eysenbach, 2004; Fan & Yan, 2010; Fricker, 2008; Gilljam et al., 2011; Gosling et al., 2004; Szolnoki & Hoffmann, 2013). Die Auswahl der zu vergleichenden Objekte und die Auswahl der Kriterien sowie der Ziele der Meinungs-, Wahl- und Marktforschung werfen dann allerdings keine deckungsgleichen, sondern unterschiedliche Bilder auf ihre Vorteilhaftigkeit und Konfigurationen von Methodenkombinationen.

Beispielsweise erlauben Online-Befragungen ähnlich wie Telefonbefragungen einfache Befragerorientierte Filterung der Fragen und fast synchrone Reaktion der Softwaresysteme bzw. eben der Interviewer. Damit haben Online-Befragungen Gemeinsamkeiten mit Telefonerhebungen und unterscheiden sich in diesem Punkt gegenüber den sogenannten asynchronen Verfahren wie Brief und E-Mail, bei denen die befragte Person erst später antwortet und die Daten oft erst später in Datenbanken integriert werden. Face-to-Face und telefonische Befragungen erfolgen simultan zwischen Menschen und sind dadurch verschieden von brieflichen, E-Mail oder Online-Erhebungen, die keine menschliche Interaktion erfordern.

Schon in den letzten zwei Jahrzehnten hat sich die Forschung vor allem aus Unternehmenssicht mit Kommunikationsprozessen und Feedback über Webseiten von Unternehmen sowie der ein-, zwei- und mehrseitigen Kommunikationsprozessen in virtuellen Communities, Wikis, Online Gruppen, Facebook, Blogs usw. beschäftigt (Albors, Ramos, & Hervas, 2008; Chu & Manchanda, 2016; Faraj, Jarvenpaa, & Majchrzak, 2011; Gruner, Homburg, & Lukas, 2014; Guo, Schwartz, Burstein, & Linger, 2009; Li, Wei, Tayi, & Tan, 2016; Reischauer & Mair, 2018; Sproull, Dutton, & Kiesler, 2007). Fakt ist, dass Unternehmen und Individuen immer stärker über diese digitalen Medien kommunizieren. Gerade Individuen bilden durch diese Mediennutzung und Kommunikation auch ihre eigene Meinung und verbreiten diese (Compagni, Mele, & Ravasi, 2015; Dell'Era & Verganti, 2011; Talke & Hultink, 2010; Van den Bulte, 2004). Gerade in Foren kommunizieren Individuen, die als Meinungsführer agieren. So lassen sich bei Erhebungen in Web-Foren, verglichen mit anderen Kanälen, leichter Meinungsführer identifizieren und befragen (Mathes & Czaplicki, 1993; Noelle-Neumann, Haumann,

& Petersen, 1998). Die erreichbaren Bevölkerungsgruppen können sich jedoch zwischen Online- und Offline-Kontaktierung und in Hinblick auf die Responses unterscheiden. Dies impliziert, dass bestimmte Bevölkerungsgruppen (z. B. Hochbetagte) nur schwer im Internet zu erreichen sind, eine geringere Responserate aufweisen und ggf. unterrepräsentiert sind. Dies kann zu einer Verzerrung der Ergebnisse führen und erfordert ein Gegensteuern durch Methodenkombinationen oder ein Nachsteuern mittels durchdachter Ansprache der zu Befragenden und Quotierung. Durch Voransprache und Quotierung lässt eine Nichtberücksichtigung oder Überrepräsentation jedoch ausgleichen.

Verschiedenste Umfragen und Erhebungen zeigen die immer stärkere Internet Verfügbarkeit und Nutzung von Individuen in verschiedensten Bevölkerungsschichten. 2018 nutzen 90,3 Prozent der deutschen Gesamtbevölkerung das Internet mit einer durchschnittlichen Mediennutzungszeit von 196 Minuten und dabei sind 77,0 Prozent sogar täglich online (ARD/ZDF, 10.10.2018). Die Nutzungsintensität des Internets nimmt insgesamt immer stärker zu und tendiert in Richtung einer immer stärkeren Durchdringung verschiedener Bevölkerungsgruppen. Insbesondere die Nutzungszeit für mediale Internetnutzung, Individualkommunikation und sonstige Internetnutzung steigt und wird als Indikator für die gerichtete Verteilung von Aufmerksamkeit angesehen (ARD/ZDF, 10.10.2018; Koch & Frees, 2017).

Jedoch gibt es immer noch Einschränkungen der Verfügbarkeit des Internets und vor allen Dingen der Übertragungsgeschwindigkeiten – oft im ländlichen Bereich in Deutschland. Demografisch betrachtet, zeichnet sich bereits 2017 ab, dass die Bevölkerung in Deutschland bis 59 Jahren deutlich über 90 Prozent, in den Altersgruppen von 14-29 Jahren zu 100 Prozent vertreten sind, während allerdings nur 74,2 Prozent der Bevölkerung ab 60 Jahre das Internet zumindest selten nutzt (Koch & Frees, 2017). Soziografisch betrachtet diffundiert die Nutzung des Internets auch immer weiter. Während zunächst vor allem stärker gebildete Bevölkerungsgruppen Zugang zum Internet hatten weitet sich die Nutzung immer stärker auf alle Bevölkerungsgruppen aus.

Es liegt auf der Hand, dass es einer differenzierten Begutachtung der Vor- und Nachteile, Einsatzbedingungen und Empfehlungen für Vorgehensweise für online Datenerhebungen und Auswertungen bedarf. Dies gilt ebenfalls für spezifische Datenerhebungen und Datenanalysen von Instituten und Anbietern und ihre Methodenkombinationen.

### 3. Kurzbeschreibungen und -beurteilungen gängiger Erhebungsinstrumente

#### a. Face-to-Face Befragungen

Face-to-Face Befragungen, auch als persönliche Befragung oder persönliches Interview bezeichnet, beschreiben die physische Anwesenheit von Interviewer und befragter Person (Duffy et al., 2005; Ellis & Krosnick, 1999). Vorteile liegen bei Face-to-Face Befragungen vor allem darin, dass sie klar strukturiert, aber auch flexibel und adaptierbar einsetzbar sind und dass der Interviewer direkt auf Fragen eingehen und die befragte Person direkt beobachten kann (Duffy et al., 2005; Ellis & Krosnick, 1999). Darüber hinaus kann der Interviewende auch positive Emotionen und Begeisterung über Mimik, Gestik und Sprachen bewusst und unterschwellig vermitteln, den die Befragten wahrnehmen, so dass sie engagierter sind über die Befragung nachzudenken und alle Fragen zu beantworten (Chang & Krosnick, 2009). Allerdings liegt hierin auch die Problematik von Verzerrungen aufgrund sozialer Erwünschtheit, die im weiteren noch diskutiert wird (Stocké, 2004; Winkler, Kroh, & Spiess, 2006). Die Repräsentativität von Face-to-Face Befragungen gilt allgemein als gut, jedoch finden sich auch Einschränkungen wie z. B. dass Befragte mit Universitätsabschluss unterrepräsentiert sind (Szolnoki & Hoffmann, 2013).

Auch Face-to-Face Befragungen bzw. persönliche Befragungen werden zunehmend von digitaler Technologie unterstützt. Dafür hat sich der Begriff Computer-Assisted Personal Interviews (CAPI) etabliert (Ilieva et al., 2002). Durch die Verbindung mit digitaler Technologie bei der persönlichen Befragung können weitere audio-visuelle Medien zur Veranschaulichung und Erklärung genutzt werden. Zudem können Antworten direkt online von dem Interviewer eingegeben werden, so dass eine synchrone Datenbankintegration erfolgt.

Nachteile von Face-to-Face Befragungen liegen vor allem in den hohen Kosten, der geografischen Gebundenheit, Zeitbedarfen und zudem dem recht hohen Risiko von Interviewer-Verzerrungen, die direkt auf die Datenqualität einwirken (Szolnoki & Hoffmann, 2013). Verschiedene Studien zeigen, dass Face-to-Face Befragungen (vor allem im Vergleich zu Online-Befragungen) Limitationen der Aussagequalität durch Verzerrungen durch den Interviewer insbesondere in Richtung der sozialen Erwünschtheit von Antworten beinhalten (Duffy et al., 2005). Interviewte versuchen bewusst oder unbewusst so zu antworten, wie sie glauben, dass der Interviewer es gutheißt oder wie die öffentliche Meinung es gutheißt würde (Stocké, 2004; Winkler et al., 2006). Verzerrungen sind vor allem bei sensiblen Fragen wahrscheinlich und werden noch stärker, wenn die Interviewte lieber mit „ich weiß nicht“ antworten würden, aber dann Interviewer die Befragten zu vollständiger Beantwortung motivieren (Duffy et al., 2005).

Wie bereits angeführt, sind Vor- und Nachteile je nach Vergleich von Verfahren ungleich. Verschiedene Charakteristika von Methoden wirken komplementär. So wird oft gegen Online-Befragungen das Argument mangelnder Repräsentativität angeführt. Aber bereits 2005 konnte eine Studie eine gute Vergleichbarkeit von Face-to-Face Befragungen und Online-Erhebungen bei Wahlentscheidungen belegen, wenn die Online-Befragung über Bevölkerungsgruppen gewichtet wurde (Duffy et al., 2005). Der Vorteil von geringerer Verzerrung durch soziale Erwünschtheit bei Online-Befragungen gegenüber Face-to-Face oder Telefonbefragungen geht allerdings mit dem gewissen Nachteil einher, dass online öfter die Antwort „ich weiß nicht“ oder „ich bin unsicher“ gegeben wird (Duffy et al., 2005). Die Gründe dafür können in der Unsicherheit des Interviewten liegen, aber auch (siehe oben) in der bei persönlicher oder telefonischer Befragung vorkommenden Verzerrungen der sozialen Erwünschtheit (Duffy et al., 2005). Damit können die Antworten „ich weiß nicht“ oder „ich bin unsicher“ eben genau eine höhere Validität implizieren.

## b. Telefonbefragungen

In den letzten 70 Jahren hat sich das Telefoninterview zu der am stärksten verwendeten Befragungsform entwickelt, allerdings mit jüngst abnehmender Tendenz (Szolnoki & Hoffmann, 2013). Telefonbefragungen umfassen im Regelfall Befragungen durch Anruf von Interviewern auf dem Festnetz oder Mobiltelefon des Befragten. Typischerweise beantwortet der Interviewte dann die Fragen direkt und der Interviewer notiert und gibt die Daten ein. Alternativ vereinbart der Interviewer einen Termin zur Beantwortung der Fragen und kann ggf. Informationsmaterial vorher postalisch oder elektronisch versenden. Denkbar ist auch, dass Befragungen über internetbasierte Telefonie durchgeführt werden oder dass Telefonbefragungen über Initiative von interessierten Personen auf Webseiten – also letztlich Online – angestoßen werden.

Telefonische Befragungen werden auch zunehmend von Technologie unterstützt, wofür typischerweise der Begriff Computer-Assisted Telephone Interviews (CATI) verwendet wird (Ilieva et al., 2002). Die technologische Unterstützung erleichtert den Befragten und Interviewern die Befragung, bricht komplexe Fragebögen herunter und ermöglicht die synchrone Integration der Daten durch den Interviewer in eine Datenbank. Damit steigt im Regelfall die Datenqualität.

Die Vorteile von Telefonbefragungen gegenüber der Face-to-Face-Befragung liegen in der Schnelligkeit, der guten und günstigen geografischen Deckung und Erreichbarkeit von Personen und den generell geringeren Kosten (Chang & Krosnick, 2009; Dillman, Sangster, Tarnai, & Rockwood, 1996; Ellis & Krosnick, 1999). Gegenüber E-Mail Befragungen und den automatisierten Online-Befragungen kann der Interviewende am Telefon Fragen des Interviewten beantworten. Gegenüber schriftlichen oder elektronischen Fragen kann der Interviewende aber auch positive Emotionen und Begeisterung durch die Stimme vermitteln, so dass Befragte engagierter sind über die Befragung nachzudenken, die Fragen möglichst valide zu beantworten und alle Fragen zu beantworten (Chang & Krosnick, 2009). Die persönliche Interaktion bei Telefonbefragungen ist weniger umfassend als bei Face-to-Face Befragungen angelegt, aber ermöglicht durch die Interaktionssituation ein direktes Eingehen auf Fragen sowie Einflussnahme durch die Stimme und Gesprächsatmosphäre.

Studien berichten über geringere Zurückhaltung bei Telefonbefragungen hinsichtlich der Beantwortung von sensiblen Fragen und über geringere Verzerrungen im Sinne eines Interviewer-Biases (Stocké, 2004; Winkler et al., 2006) durch soziale Erwünschtheit als persönliche Befragungen (Szolnoki & Hoffmann, 2013). Auch wenn die Response-Rate bei Face-to-Face Befragungen höher als die bei Telefonbefragungen liegt, sind kaum Unterschiede im Hinblick der Responserate zwischen Face-to-Face und telefonischen Befragungen entlang von demografischen Merkmalen zu konstatieren. Ältere (auch Meta-) Untersuchungen vor 2000 sehen allerdings Limitationen der Repräsentativität von Telefonbefragungen insbesondere, weil besser ausgebildete und wohlhabendere Befragte überrepräsentiert sind (Chang & Krosnick, 2009; Ellis & Krosnick, 1999). Vereinzelt finden Studien, dass bestimmte Konsumentenwünsche überrepräsentiert sind, wie z. B. die Bevorzugung von mehr Süße von Befragten am Telefon in einer Vergleichsstudie von Weinkonsumenten (Szolnoki & Hoffmann, 2013). Wenn mit der Möglichkeit gewichteter Stichproben gearbeitet wird, können gewisse Einschränkungen der Repräsentativität allerdings kompensiert werden (Szolnoki & Hoffmann, 2013). Dabei nutzen Telefonbefragungen typischerweise probabilistische Verfahren und ziehen die Stichproben bzw. verwenden die Antworten entlang von Quoten auf der Basis von z. B. sozio-demografischen Merkmalen der Befragten.

Nachteile von Telefonbefragungen liegen darin, dass bildliches Material kaum zum Einsatz kommt (ggf. durch versendete Materialien und digitale Links) und in den Interviewer-Verzerrungen, die beide direkt auf die Datenqualität einwirken (Szolnoki & Hoffmann, 2013). Allerdings ermöglicht die

internetbasierte Telefonie audio-visuelle Materialien einzubetten und die Befragung durch Internetnutzung zu begleiten. Elektronische Unterstützung und Technologien ermöglichen zudem die Unterstützung von Filtern und die direkte Eingabe der Daten durch den Befragten oder den Interviewer. Hierdurch kommt es dann auch zu einer Verschränkung von Telefon- und Onlinebefragungen, die sich auf die Datenqualität positiv auswirkt.

### c. Brieflich-postalische und E-Mail-Befragungen

Brieflich-postalische Erhebungen versenden bereits gedruckte Fragebögen an die Hausanschrift von Personen. Sie gelten als klassisches Instrument neben der persönlichen Befragung, das aber immer mehr in den Hintergrund gedrängt wird. Brieflich-postalische Erhebungen erfolgen asynchron wie typischerweise auch die Befragung mittels E-Mails. E-Mail-Befragungen integrieren in die E-Mail ein Dokument anhand dessen befragt wird. Die Befragten füllen dann den ggf. ausgedruckten Bogen aus und versenden diesen per E-Mail, Post, Fax oder Scan. Es handelt sich bei brieflich-postalische Erhebungen um ein im Regelfall asynchrones Verfahren, es sei denn die Befragten nehmen direkt an der verlinkten Online-Befragung teil. Rückläufe bei der E-Mail-Befragung sind relativ schnell möglich, z. B. 9,16 Tage gegenüber 14,39 Tagen bei postalischen Erhebungen (Schaefer & Dillman, 1998). Gerade bei kurzen Fragebögen ließ sich ein recht schneller und hoher Rücklauf in Studien zeigen (Schaefer & Dillman, 1998).

Erfahrungswerte und gute Praxis weisen darauf hin, dass z. B. personalisierte Einladungen sowie Erinnerungen bei E-Mail Befragungen die Rücklaufquote und Antwortqualität steigern können (Schaefer & Dillman, 1998). Zunehmend haben brieflich-postalische Erhebungen und E-Mail Befragungen immer niedrigere Responseraten und längere Responnzeiten zu verzeichnen. Ohnehin sind brieflich-postalische Erhebungen kosten- und zeitintensiv sowie fehleranfällig bei der manuellen Transkription in eine Datenbank (Ilieva et al., 2002). Bei mehrfachen Nachfassaktionen steigen die Kosten und Zeitbedarfe weiter an. Um insbesondere die Responseraten zu erhöhen, Kosten und Qualität des Datentransfers zu verbessern enthalten E-Mail Befragungen auch Links zu einer Online-Befragung. In dem Fall verschwinden wiederum die Grenzen von E-Mail und Online-Erhebungen.

### d. Online-Befragungen

Online-Befragungen werden auch als web-basierte Befragungen oder Internet-Befragungen bezeichnet. Von Online-Befragungen wird gesprochen, wenn die befragte Person Fragen direkt im Internet beantwortet und diese Daten zeitgleich - „synchron“ – in Datenbanken abgespeichert werden. Antworten bei Online-Befragungen werden von dem Befragten direkt am Bildschirm gegeben. Die Personen reagieren dabei auf Werbung und die Einbettung der Befragung. Meist leiten einleitende kurze oder wenige Fragen die Befragten auf bestimmte Fragenkomplexe. Filterfragen leiten die Personen in der eigentlichen Befragung und ermöglichen gerade bei komplexeren Fragebögen eine leichte Navigation und damit eine kognitive Erleichterung beim Befragten (Schaefer & Dillman, 1998). Fragen und Antworten können bei Online-Befragungen unmittelbar und ohne die Limitation geografischer Nähe transferiert werden, solange ein Internetzugang besteht (Ilieva et al., 2002).

Online-Befragungen sind entweder direkt auf spezifischen Webseiten, insbesondere in Foren und Plattformen integriert oder werden nach Einladung der befragten Person über einen Link angesteuert (Callegaro et al., 2014; ESOMAR, 2015; Fisher, 2018; Reischauer & Mair, 2018; Ward & Meade, 2018). Gerade wenn es um das Erreichen von bestimmten Zielgruppen oder um größere Audiences

geht, gilt die Entwicklung von Pop-up-Fenstern auf Internetseiten bzw. Browsern, Plattformen und Foren als wichtige Weiterentwicklung (Ilieva et al., 2002). Typisch sind dabei Antworten auf Fragen, die auf Plattformen insbesondere Foren elektronisch platziert werden. Bei der Streuung von Fragen auf Webseiten und die Einladung mittels Banner, pop-up Fenster usw. wird oft ein Riversampling angewendet (Dixon, Smyth, & Chiswell, 1999; Steber, 2018). Neben der Integration der Befragung auf Webseiten wie beim Riversampling kann die Einladung der befragten Person wiederum mittels unterschiedlicher Medien wie per E-Mail, Telefon, Brief erfolgen.

Dabei kann es sich um eine einmalige Befragung oder die Akquisition von Panelmitgliedern handeln. Dann erhält die befragte Person am Ende die Anfrage, ob sie auch an einem Panel teilnehmen will. Online-Befragungen ermöglichen insofern eine Kombination mit Online-Panels. Teilnehmende Personen nehmen teils kontinuierlich an Befragungen bzw. an einem Panel teil (Chang & Krosnick, 2009; Duffy et al., 2005). Entweder werden Personen gleich zu Beginn der Befragung gebeten, an einem Panel teilzunehmen und sich zu registrieren oder sie werden am Ende der Befragung gefragt, ob sie öfter teilnehmen wollen.

Je nachdem ob alle erreichbaren Teilnehmer einer Befragung berücksichtigt werden oder ob eine zufällige Auswahl von Teilnehmern aus der Gesamtpopulation erfolgt, werden sie als nicht-probabilistische oder probabilistische Stichproben bzw. Auswahlverfahren bezeichnet. Jüngst werden diese Verfahren zur Verbindung der jeweiligen Vorteile auch gewissermaßen kombiniert. Hierfür gibt es verschiedenste Varianten wie z. B. eine Zufallsauswahl nach zuvor festgelegter Quote der Zielpopulation.

Der Interviewer-Bias kommt bei Online-Erhebungen nicht nennenswert zum Tragen und ist auch geringer als in Vergleich zu Telefon und besonders zu Face-to-Face-Befragungen, weil Online-Befragungen die Antworten mittels und in ein automatisiertes System erfassen (Chang & Krosnick, 2009). Interviewer-Verzerrungen und sozial erwünschtes Antwortverhalten sind gering, weil die Antwort eben neutral und nicht gegenüber einer Person (meist anonym) dargelegt wird. Bei Online-Befragungen und dem automatischen System für die Antworten fällt dabei vor allem der (unterschwellige) Druck einen guten Eindruck auf den Gesprächspartner zu machen weg. Im Vergleich dazu wirken wie bereits angesprochen Interviewer-Verzerrungen bei Face-to-Face und Telefonerhebungen in verschiedenen und sogar unterschweligen Formen. So können bspw. beim Telefonat selbst Pausen als unangenehm wahrgenommen werden und den Befragten unterschwellig beeinflussen (Chang & Krosnick, 2009). Bei Face-to-Face- und Telefonbefragungen kann der Druck auf die befragte Person und seine Rückkopplung sowie eine Verstärkung durch den Interviewenden zu einer zirkulierenden Dynamik führen, welche selbst beim Interviewenden Unsicherheiten und ein unwohles Gefühl erzeugen kann und die Antworten auf sensible Themen (z. B. zu Einstellungen und Werten) verzerrt (Chang & Krosnick, 2009).

Online-Befragungen erlauben generell die Erhebung zu relativ niedrigen Kosten, weil die fixen Kosten durch die Automatisierung mit zunehmender Stichprobengröße abnehmen. Sie sind relativ schnell, flexibel, interaktiv, adaptiv und erlauben die Nutzung von audio-visuellen Präsentationen ohne dass Interviewer präsent sein müssen (Szolnoki & Hoffmann, 2013). Befragte von Online-Befragungen können die Fragen beantworten, wenn sie Lust und Zeit haben, ohne sich auf den Fragenden einzustellen oder sich auf einen Termin festzulegen. Sie können zwischendrin anhalten und später weitermachen. Dies kann das Interesse aufrecht erhalten und die Aufmerksamkeit steigern. Befragte können Fragen mehrfach durchlesen, so dass kognitive und soziale Anstrengungen sinken (Chang & Krosnick, 2009). Daraus ergeben sich viele praktische Vorteile, aber auch Vorteile im Hinblick auf die

Aussagequalität, die zudem deutlich durch die geringen Interviewer-Verzerrungen durch soziale Erwünschtheit gestärkt wird.

Obendrein können Online-Erhebungen sehr schnell auf aktuelle Themen und Trends reagieren. So lassen sich schnell Antworten bei der Bevölkerung zu gesellschaftlichen und politischen Fragen erzielen. Hier finden vor allem nicht-probabilistische Stichproben, teils in Kombination mit probabilistischen Annahmen in Form von Quoten Einsatz. Wahlrends und Reaktionen auf Wahlen sind damit ein Kerngebiet der Online-Befragungen mit hohen Rücklaufquoten (Ansolabehere & Rivers, 2013; Ansolabehere & Schaffner, 2014).

Im Hinblick auf die Stichprobe bzw. Repräsentativität bieten Online-Befragungen vor allem den Vorteil Menschen zu befragen, die unwillig sind an telefonischen Erhebungen teilzunehmen. Online-Befragungen können die Antwortverweigerung (non-response bias) und damit Verzerrung bei sensiblen Personengruppen reduzieren. Dies erlaubt wiederum eine Verbesserung der Repräsentativität und Aussagequalität. Überdies lassen sich aber bei Online-Befragungen auch einfacher sogenannte Snowball-Verfahren einsetzen. Gerade diese haben niedrigere Repräsentativitätsraten, weil sie meist von Personen einer Gruppe (z. B. bei Studierenden, die den Ball weiter an andere Studierende weiter rollen und so in der Gruppe der Studierenden bleiben) intensiver geteilt werden (Chang & Krosnick, 2009; Ellis & Krosnick, 1999; Szolnoki & Hoffmann, 2013). Der Einsatz von Snowball Verfahren führt so leicht zu einer Überrepräsentation von Gruppen.

Online-Befragungen erlauben den Zugang zu Themen und Individuen, die über anderen Kanäle wie Briefe, Festnetz oder direkte Befragung in Fußgängerzonen schlecht bis gar nicht erreichbar sind. Bis vor ein paar Jahren galten Online-Befragungen dennoch als weniger repräsentativ für die Gesamtpopulation als Face-to-Face, briefliche oder Telefonbefragungen, jedoch gilt diese Limitation seit neuerem als im Wesentlichen entkräftet (Szolnoki & Hoffmann, 2013), weil das Internet mehr und intensiver von allen Bevölkerungsgruppen genutzt wird (ARD/ZDF, 10.10.2018; Koch & Frees, 2017).

Einige Studien berichten, dass in Online-Befragungen ähnlich wie in Telefonbefragungen tendenziell Personen mit höherer formaler Ausbildung überrepräsentiert sind (Szolnoki & Hoffmann, 2013). Interessanterweise kommt aber die Vergleichsstudie von Kennedy et al. (May 2016) zu dem Schluss, dass gerade bei Riversampling und bei nicht-probabilistischen Verfahren dann niedriger formal ausgebildete Personen und solche mit niedrigen sozio-ökonomischen Satus teilnehmen. Dies deutete auf Vorteile von Kombinationen der Erhebungsverfahren im Online-Bereich hin.

Um die Repräsentativität von Online-Stichproben so groß wie möglich zu gestalten, werden Quoten, Gewichtungen und Korrekturen der Stichprobe mittels sozio-demografischer Merkmale eingesetzt (Berzofsky, Welch, Williams, & Biemer, 2008; Berzofsky, Williams, & Biemer, 2009; Cumming, 1990; Dülmer, 2007; Yang & Banamah, 2014). Während es sich bei Quoten um ein nicht-probabilistisches Stichprobenverfahren handelt, mit dem eine Stichprobe hinsichtlich relevant erachteter Merkmale aus der Grundgesamtheit modelliert (nachgebildet) wird, werden Gewichtungen gewöhnlich bei probabilistischen Stichproben eingesetzt, um die Ergebnisse der Stichprobe an die Grundgesamtheit anzugleichen.

Zweifel werden allerdings noch jüngst geäußert, wenn es um Stichproben aus Online-Panels geht, da hier nur Personen ausgewählt werden können, die sich durch Beitritt zum Panel bereit erklären, wiederholt an Befragungen teilzunehmen. Tendenziell wären dabei jüngere Männer und Menschen, die stärker an den befragten Themenstellungen interessiert sind, überrepräsentiert (Duffy et al., 2005). Dies widerspricht wiederum der Studie von Kennedy et al. (May 2016) und deutet auf die Vorteile der Kombinationen von Riversampling und Online-Panel hin.

Teils findet sich der Einwand, dass Panels in ihrer Aussagekraft beschränkt seien, weil die Befragten zu professionell werden. Diese sogenannte „Panel-Konditionierung“ bezeichnet den Effekt, dass die mehrfache Teilnahme bei den Befragten zu Lerneffekten führt und möglicherweise eine Selbstselektion eintritt, durch welche die Teilnehmer des Panels als weniger repräsentativ für die Gesamtbevölkerung anzusehen sind. Neuere Studien zeigen, dass eine „Panel-Konditionierung“ keinen oder nur einen sehr kleinen Einfluss auf die Repräsentativität für die Gesamtbevölkerung hat (Chang & Krosnick, 2009). Es ist sogar ein umgekehrter Effekt zu beobachten. Panelbefragte geben mit längerer Teilnahme reliablere Antworten auf Einstellungsfragen (Chang & Krosnick, 2009). Befragte werden im Zeitablauf und mit Erfahrungsgewinn immer stärker kognitiv involviert und geben daher zuverlässigere Antworten (Chang & Krosnick, 2009). Wie in Offline-Befragungen besteht aber das Problem, dass Befragte nur einmal teilnehmen und daher dann Längsschnitzaussagen schwierig sind. Allerdings tritt das Problem der Konditionierung von Panelteilnehmern nur auf, wenn sie eben mehrfach teilnehmen.

Aus den verschiedenen Argumenten und Befunden zur Repräsentativität von Online-Befragungen kann man schließen, dass Online-Befragungen nicht-probabilistische und probabilistische Verfahren kombinieren sollten und dabei sowohl Riversampling als auch Quoten einsetzen sollten, um eine höhere Repräsentativität zu erreichen. Darüber hinaus erlauben Gewichtungen eine weitere Steigerung der Repräsentativität. Insgesamt erlaubt eine Kombination von verschiedenen Verfahren und ein durchdachtes schrittweises Verfahren, eine sehr hohe Aussagekraft und Repräsentativität zu erzielen. Wenn z. B. ein Teil der zu befragenden Personen vorher mit anderen Medien kontaktiert und mit einem Link zur Befragung geleitet werden, liegt bereits eine gewisse Methodenkombination der Online-Erhebung mit anderen E-Mail oder telefonbasierten Erhebungsformen vor, die eine zusätzliche Steuerung der Auswahl, Motivation der befragten Personen und damit letztlich der Repräsentativität für Zielgruppen und Populationen erlaubt. Dies entspricht dann dem Vorgehen wie bei probabilistischen Verfahren. Die Kombinationen von Online-Befragungen mit Einladungen über Briefe, E-Mails, Telefonanruf oder Unterstützung mittels Telefon, internetbasierter Telefonie (z. B. Skype/Whatsapp/Facetime, Facebook Messenger), sowie in Social Media Gruppen oder auf mobilen Geräten weiten die Erhebungspotentiale und Methodenkombinationen aus. Weitere Vorteile im Hinblick auf Repräsentativität durch Quotenauswahl und Gewichtungen erlauben Online-Befragungen als Panel-Erhebungen (Chang & Krosnick, 2009; Duffy et al., 2005). Panelteilnehmer hinterlegen sozio-demografische Variablen für sich und können dann gezielt oder zufällig auf bestimmte Befragungen aufmerksam gemacht werden. Dabei können Antworten von Panelteilnehmern mit anderen Informationen verbunden werden. Beispielsweise können Panelteilnehmer auch wiederholt auf dieselben Fragen antworten und dies ermöglicht Längsschnittanalysen (z. B. der Einstellungsänderung) oder eine Analyse von Präferenzen innerhalb der Entscheidungsbildung.

Wie oben erläutert nutzen Online-Befragungen unterschiedliche Erhebungs- und Analyseformen und diese auch in spezifischer Kombination. Online-Befragungen haben ihrerseits unterschiedliche Charakteristika und integrieren teils die Vorteile anderer Verfahren (Duffy et al., 2005). Bereits 1998 argumentierten (Schaefer & Dillman, 1998) daher für die Nutzung von Komplementaritäten durch eine multi-mode Datenerhebung mittels E-Mail und postalischem Anschreiben der Teilnehmer an einer Online-Befragung. Zudem lässt sich kaum von der Online-Befragungsform sprechen und diese stellvertretend für andere bewerten. Aus diesem Grund diskutiert die hier vorliegende Analyse verschiedene Komponenten von Online-Befragungen, zieht wissenschaftliche und praxisgeprägte Kriterien heran und beurteilt daran differenziert die typische Vorgehensweise von Civey und den einzelnen typisch gewählten Komponenten.

## 4. Kurzbeschreibung der Erhebungs- und Analysepraxis bei Civey

Civey setzt einen webbasierten Opt-In-Ansatz zur Rekrutierung von Teilnehmern und zur Datenerhebung über das Internet ein (vgl. zur Rekrutierung Birnbaum, 2004). Bei der als Riversampling bezeichneten Methode, werden Personen durch die Platzierung von Pop-up-Fenstern, Verlinkungen oder Banner auf zahlreichen Webseiten auf eine Umfrageeinladung aufmerksam gemacht (Baker et al., May 2013; Kennedy et al., May 2016). Um eine hohe Heterogenität demografischer, einstellungs- und verhaltensrelevanter Faktoren der angesprochenen Personenkreise (Cluster) zu erreichen, werden in der Regel eine Vielzahl unterschiedlicher Webseiten eingebunden (Baker et al., May 2013).

Civey verwendet ein spezifisches Umfrage-Tool (Widget), das nach eigenen Angaben auf 25.000 Webseiten (URLs) aktiv eingebunden ist und monatlich ca. 20 Millionen Aufrufe erzielt (Civey GmbH, 2018). Diese Aufrufe werden nicht durch das primäre Interesse von Personen an Umfragen erreicht, sondern durch die Einbettung des Umfrage-Tools auf viel besuchten und unterschiedlichen Webseiten. Passend zum Inhalt der Webseite wird dem Webseitenbesucher durch das Umfrage-Tool eine erste Frage präsentiert, die gleichermaßen darauf ausgerichtet ist, die Aufmerksamkeit des Webseitenbesuchers zu erreichen und eine möglichst hohe Intention zu schaffen, an der Entscheidung über diese Frage durch die gezielte Auswahl einer oder den Ausschluss mehrerer der vorgegeben Antworten, teilzuhaben.

Bis zu diesem Schritt entspricht das Umfrage-Tool dem typischen Riversampling-Ansatz. Es liegen keinerlei Informationen über den Teilnehmer vor. Bei Auswahl einer Antwort präsentiert das Umfrage-Tool dem Webseitenbesucher den Aufruf "Lassen Sie Ihre Stimme zählen!" und ermöglicht nach Eingabe von Geschlecht, Geburtsjahr und Postleitzahl sowie der Zustimmung zur Datenverarbeitung nach DSGVO die Einsicht in das Befragungsergebnis, weitere Befragungen sowie die Möglichkeit der Registrierung mit einer E-Mail-Adresse.

Eine anschließende Verifizierung umfasst den Nachweis, dass es sich beim Nutzer um eine natürliche Person handelt und die Abfrage weiterer Stammdaten zur späteren Verwendung. Die automatisierte Verhaltens- und Stammdatenverifizierung resultiert in einem "User-Trust-Score", der eine Voraussetzung zum Einschluss in repräsentative Stichproben bei Civey darstellt. Hier handelt es sich um ein typisches Online-Panel-Vorgehen.

Um Repräsentativität zu erreichen durchlaufen die Umfragen bei Civey einen mehrstufigen Prozess, der den Befragungszeitraum und die Definition der Zielpopulation einschließt. Bei den resultierenden Stichproben handelt es sich dem entsprechend nicht um Zufallsstichproben der Grundgesamtheit, sondern um stratifizierte Stichproben anhand von Quoten, innerhalb welcher die Umfragen den Nutzern zufällig, aber ausbalanciert nach Uhrzeit und Webseite, dargeboten werden.

Durch die Verwendung demografischer Quoten versucht Civey sicherzustellen, dass die Auswahl eine (repräsentative) Zusammensetzung aufweist, welche der Zielpopulation entspricht (Serfling et al., 2018). Der zunehmende Einsatz formeller Designs bei der Stichprobenauswahl zielt auf eine entscheidende Verbesserung der Repräsentativität ab (Baker et al., May 2013). Zugrunde gelegt werden bei Civey offizielle Zahlen und Fakten, wie beispielsweise die des Statistischen Bundesamtes.

Coverage-Fehler (die selektive Erreichbarkeit bestimmter Probandengruppen), werden bei den Stichproben der Civey-Umfragen durch Auswahl-Algorithmen in hohem Maß berücksichtigt und soweit notwendig abschließend über weitere soziodemografische und einstellungsrelevante Merkmale nachgewichtet.

Die in Abbildung 1 dargestellte Systematik, verdeutlicht das schrittweise Vorgehen seitens Civey, um die Datenqualität und die Übertragbarkeit der Ergebnisse auf eine spezifische Grundgesamtheit zu stärken.

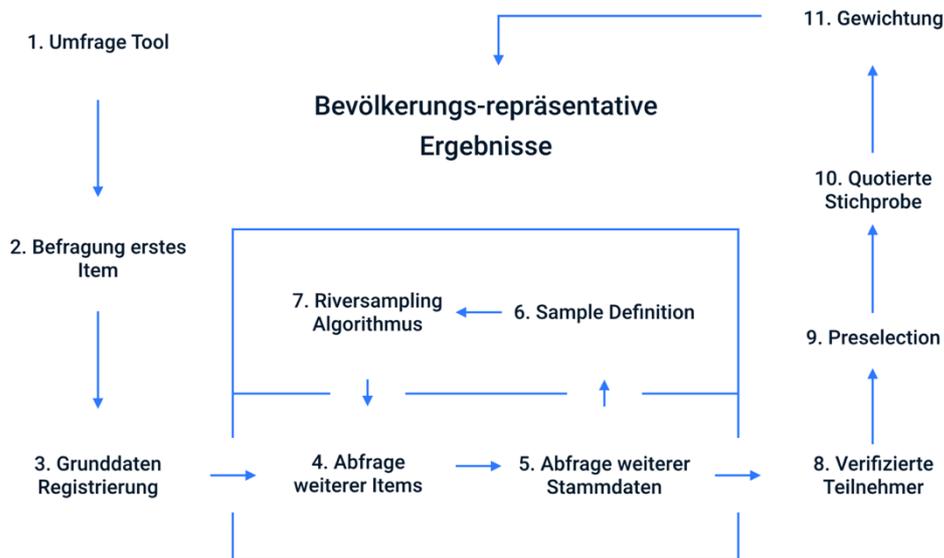


Abbildung 1: Darstellung des Erhebungsverfahrens bei Civey (Civey GmbH, 2018).

1. **Umfrage-Tool:** Das Civey Umfrage-Tool ist (Stand: November 2018 auf 25.000 Webseiten (URLs) aktiv, lt. Angaben von Civey).
2. **Befragung erstes Item:** Das erste Item, das im Tool abgefragt wird, ist passend zum Content der Website gewählt, um eine möglichst hohe Motivation zur Teilnahme sicherzustellen. In der Regel werden die Antworten, die über das erste Item erfasst werden, im weiteren Verlauf aussortiert, da das Antwortverhalten von den Inhalten der Website beeinflusst sein kann.
3. **Grunddaten Registrierung:** Nach der Beantwortung des ersten Items, wird der Teilnehmer aufgefordert, seine Zustimmung zur Datenverarbeitung nach DSGVO zu erteilen und Angaben zu Alter, Wohnort und Geschlecht zu machen.
4. **Abfrage weitere Items:** In der Folge werden beim Nutzer weitere Items abgefragt. Neben dem Erfassen der Antworten erfolgt hier auch eine Verhaltensverifizierung, ob u.a. ob das Clickverhalten des Nutzers dem einer natürlichen Person entspricht.
5. **Abfrage Stammdaten:** Innerhalb der Items werden auch sogenannte Stammdaten abgefragt, die zur weiteren Datenverarbeitung/Gewichtung notwendig sind, wie z.B. Beruf, Bildung, Konsumgewohnheiten u.ä.. Diese Stammdaten müssen zur Verifizierung des Nutzers zu rund 80% vollständig sein und dürfen keine Kontraindikationen (Widersprüche) enthalten.
6. **Sample Definitionen:** Auf Grundlage der Stammdaten erfolgt auch die Definition des Ziel-Samples - also welche Zielgruppen befragt werden sollen (z.B. alle Bayern, 18-29jährige Frauen, Autofahrer etc.).
7. **Randomisierter Riversampling-Algorithmus:** Die Items der Umfragen werden zufällig, aber ausbalanciert nach Zeit und URL innerhalb der Samples abgefragt (Die Schritte 4. bis 7. wiederholen sich solange, bis der Nutzer die Befragung abbricht.)

8. **Verifizierte Teilnehmer:** Nutzer, die die Grunddatenregistrierung, Verhaltens- und Stammdatenverifizierung durchlaufen und bestanden haben, gelten als verifiziert. Dazu muss der Nutzer einen bestimmten Wert auf dem internen User-Trust-Score überschreiten.
9. **Preselection:** Unter allen verifizierten Teilnehmer einer Item-Befragung wird eine Mindestanzahl an Vertretern aus den verschiedenen soziodemographischen Gruppen festgelegt. Diese bestimmt gleichzeitig den Befragungszeitraum. Beispiel: Wenn die angestrebten Stichprobengröße 5.000 Personen ist, müssen in dem in der Preselection festgelegten Befragungszeitraum mindestens 1.500 Frauen das Item beantwortet haben. Dies sorgt im weiteren Verlauf dafür, dass keine übermäßigen Gewichte eingesetzt werden.
10. **Quotierte Stichprobe:** Unter allen Teilnehmer innerhalb der Preselection erfolgt nun die Ziehung einer Quotenstichprobe nach Alter, Parteineigung, Geschlecht, Wohnort und Ort der Teilnahme (über welche Webseite bzw. an welcher Stelle während der Teilnahme).
11. **Gewichtung:** Abschließend wird die Quotenstichprobe über weitere soziodemographische Merkmale nachgewichtet. Sofern gemeinsame Verteilungen vorhanden sind, kommt dabei Poststratifizierung zum Einsatz; falls nur die Randverteilungen vorliegen, kommt Raking zum Einsatz.

## 5. Kriterien-geleitete Beurteilung der typischen Vorgehensweise bei Civey

### a. Vorbemerkungen zur Aussagequalität

Die qualitative Beurteilung von Datenerhebungs- und analyseverfahren stellt einen vielschichtigen Prozess dar, welcher sich nicht auf einen einzelnen allgemein gültigen Bewertungsmaßstab reduzieren lässt (Weichbold, 2009). In Abgrenzung zur wissenschaftlichen Definitionen von "Repräsentativität" zielt die Formulierung im medialen Kontext darauf ab, der Qualität von Fakten und abgeleiteten Argumentationen eine hohe Aussagekraft, Verlässlichkeit sowie Vertrauenswürdigkeit zuzuschreiben (Weichbold, 2009). Die Qualitätssicherung wird in der Praxis durch berufsständische Vereinigungen angeregt. Sie spiegeln den common sense für Best Practices sowie allgemeine (ethische) Vorgehensweisen wider (Weichbold, 2009). Hier sind insbesondere der Rat der Deutschen Markt- und Sozialforschung, der Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM), der Zentralverband der deutschen Werbewirtschaft (ZAW), die World Association for Social, Opinion and Market Research (ESOMAR) sowie die American Association for Public Opinion Research (AAPOR) zu nennen.

In der Wissenschaft existieren schon seit Jahrzehnten Qualitätsmaßstäbe für Datenerhebungen und -analysen. Sie umfassen weitreichende Bewertungskriterien von der Auswahl der Variablen, Personen, Gewichtung, Auswertung nebst Limitationen und Vorgehensweisen zur Sicherstellung Steigerung zur ausreichenden Güte von Reliabilität, Validität, Generalisierbarkeit (Hair, Black, Babin, Anderson, & Tatham, 2010; Schmidt & Hunter, 1998). Elektronisch gestützte und Online-Befragungen (Couper, 2000; Couper, Blair, & Triplett, 1997; Couper, Blair, & Triplett, 1999; Couper, Traugott, & Lamias, 2001) wurden anhand ausgewählter wissenschaftlicher Kriterien untersucht (Birnbaum, 2004; Chen, Chiang, & Storey, 2012; Roy, 2010; Weigold, Weigold, & Russell, 2013; Zinkhan, 2004). Die Ergebnisse sind allerdings auch nicht einhellig wie auch die angewendeten Verfahren und Methodenkombinationen.

Allgemeine Definitionen von Datenqualität fokussieren auf die Eignung der Daten für den angestrebten Verwendungszweck (Strong, Lee, & Wang, 1997). Des weiteren existieren Definitionen von Datenqualität, die sich auf die Relevanz und Aktualität der Daten, die Interpretierbarkeit von Begrifflichkeiten innerhalb der Datenerfassung, die Objektivität und Genauigkeit der Datenanalyse sowie der Kontrollierbarkeit des Datenzugangs differenzieren (Treiblmaier, 2011).

Die in der Wissenschaft seit Jahrzehnten geltenden Maßnahmen zur Bewertung von Datenqualität beruhen auf der Wahrscheinlichkeitstheorie haben sich in ihrer allgemeinen Anwendung als nützlich erwiesen (Baker et al., May 2013). Sie setzen voraus, dass (1) ein Rahmen für die Grundgesamtheit existiert; in der (2) jede Untersuchungseinheit eine positive Auswahlwahrscheinlichkeit aufweist und (3) die Auswahlwahrscheinlichkeit für jede Untersuchungseinheit berechnet werden kann. Diese Standardqualitätsmetriken sind konzipiert, um den Grad zu messen, in dem eine bestimmte Stichprobe diese Annahmen erfüllt oder aufgrund konkreter Einschränkungen (z. B. unvollständige Abdeckung oder non-respondens) verletzt (Baker et al., May 2013).

Jüngst finden sich auch in der Wissenschaft immer mehr Studien die nicht-probabilistische Ansätze und Bayes'sische Statistik nutzen und diese selbst mit multivariaten Verfahren verbinden (Hahn & Doh, 2006; Hayashi & Yuan, 2003; Lee & Shi, 2000; Rupp, Dey, & Zumbo, 2004; Song, Lee, & Hser, 2009; Song, Lu, Hser, & Lee, 2011; Zyphur & Oswald, 2015). Es scheint sich auch hier ein neuer Ansatz zu etablieren, der die alten nicht verdrängt, sondern ergänzt (Song & Lee, 2008; Steel & Kammeyer-Mueller, 2007; Stromeyer, Miller, Sriramachandramurthy, & DeMartino, 2014; Zhang, Lai, Lu, &

Tong, 2013; Zhang, McArdle, Wang, & Hamagami, 2008). Dieses spiegelt sich noch stärker in der praktischen Meinungs-, Wahl- und Marktforschung wider.

Innerhalb der empirischen Forschung wird Datenqualität traditionell vor dem Hintergrund der Messungsgüte dargestellt. Die dabei diskutierten Gütekriterien basieren auf den Axiomen der klassischen Testtheorie und wurden primär für die Verfahren von Eigenschaftsmessungen konstruiert. Sowohl ihre Umsetzung in der Praxis als auch die Übertragbarkeit dieser Kriterien auf Verfahren der kommunikativen Informationsgewinnung (z. B. Umfragen) sind nicht unstrittig (Weichbold, 2009). Hohe Datenqualität ist insofern eine notwendige, aber nicht hinreichende Voraussetzung für Validität, welche in der klassischen Testtheorie als "Gültigkeit" verstanden wird (Calder, Phillips, & Tybout, 1982). Obwohl sich viele Argumente für Online-Erhebungen finden, bezweifeln Kritiker von Online-Erhebungen im engeren Sinne die Glaubwürdigkeit solcher Daten (Dülmer, 2007; Heerwegh & Loosveldt, 2008; Malhotra, 2008; Sparrow, 2007). Die Analysemethoden und damit verbundene Methodengültigkeit unterscheidet nicht zwingend zwischen Online-Erhebungen und anderen Erhebungsformen. Methoden sind jedoch generell auf eine Wahrscheinlichkeitstheorie (und die damit verbundenen Inferenzmethoden) angewiesen und es existieren unterschiedliche Paradigmen sowie damit verbundenen Verfahren. Eine bereits langanhaltende Kontroverse resultiert aus unterschiedlichen wissenschaftstheoretischen Ansichten und führt zu verschiedenen Definitionen des Wahrscheinlichkeitsbegriffs. Während die Wahrscheinlichkeit im frequentistischen Paradigma auf die Wahrscheinlichkeit beobachteter Daten verweist, bezieht sich die bayessche Wahrscheinlichkeit auf die Wahrscheinlichkeit von Parametern, die für Forscher von Interesse sind.

Im traditionellen frequentistischen Paradigma werden Populationsparameter als unbekannte, aber feste Größen angesehen und quantitative Forscher stützten sich in der Vergangenheit auf Schätz- und Inferenzmethoden, die alle nicht-bayesschen Ursprungs sind (z. B. Fisher, 1922, 1971). Auf Grundlage von Stichprobenschätzungen werden demnach Konfidenzintervalle berechnet, innerhalb derer diese Populationsparameter vermutet werden (Rupp et al., 2004). Davon abweichend werden Populationsparameter im bayesschen Paradigma als Zufallsvariablen behandelt, die einer bestimmten Verteilung folgen. Vorkenntnisse bzw. Annahmen über die mögliche Form dieser Verteilung können so spezifiziert werden, das daraus Quasi-Meta-Analysen resultieren (Zyphur & Oswald, 2015; Zyphur, Oswald, & Rupp, 2015). Ein Ziel der bayesschen Inferenz besteht in der Analyse von Verteilungen, um die Wahrscheinlichkeit interessierender Populationsparameter zu schätzen und glaubwürdige Intervalle (d. H. Bayessche Konfidenzintervalle) zu erhalten (Rupp et al., 2004; Zyphur & Oswald, 2015; Zyphur et al., 2015). Obwohl auch außerhalb der Naturwissenschaften zahlreiche Anwendungen auf dem Bayes-Theorem basieren und die bayessche sowie die frequentistische Statistik häufig zu sehr ähnlichen Ergebnissen führen, konkurrieren beide Anschauungen sehr intensiv.

Die hier vorliegende Beurteilung zieht sowohl die in der Wissenschaft üblichen Kriterien sowie die in der Praxis üblichen sowie durch die oben genannten Institutionen empfohlenen Kriterien und Vorgehensweisen heran.

#### b. Rücklauf bzw. Response Rate

Response-Raten oder auch als Rücklaufquoten bezeichnete Bewertungen beschreiben wie viel Prozent der eingeladenen Personen antworten (Alan, 1998; Cook, Heath, & Thompson, 2000). Geringe Rücklaufquoten zeigen im Grunde, dass die Thematik für Befragte und damit letztlich für einen Teil der Population unzutreffend, nicht interessant ist und/oder nicht ausreichend gut vermittelt wurde.

Geringe Rücklaufquoten verweisen damit auf Schwächen im Hinblick der Repräsentativität (Nulty, 2008). Wenn nur wenige Personen antworten, weil wie Fragen zu komplex oder unverständlich waren, dann können die Antwortenden eher weniger stellvertretend für eine Population sein. Schlechte Rückläufe können auch darauf hindeuten, dass dieses Thema nicht ausreichend relevant für die Gesamtpopulation ist bzw. die Relevanz nur für bestimmte Bevölkerungsgruppen gilt. Geringe Rücklaufquoten verzerren darüber hinaus allgemein die Aussagefähigkeit, weil nicht davon auszugehen ist, dass die Antworten gleich der Bevölkerungsverteilung sind (Alan, 1998; Cook et al., 2000; Fan & Yan, 2010; Kaplowitz, Hadlock, & Levine, 2004; Manfreda, Berzelak, Vehovar, Bosnjak, & Haas, 2008; Nulty, 2008).

Viele der wissenschaftlichen Evaluationen zu Response-Raten (Chang & Krosnick, 2009; Cook et al., 2000; DeMars, 2012; Fan & Yan, 2010; Foster Thompson & Surface, 2007) haben sich mit dem Vergleich von brieflichen mittels ausgedruckten Fragebögen, („paper and pencil“-)Erhebungen, E-Mail-Erhebungen und teilweise auch Online-Erhebungen beschäftigt. Einige Studien zeigen, dass E-Mail basierte Befragungen geringere Rücklaufquoten haben (Bachmann, Elfrink, & Vazzana, 1996; Sproull & Kiesler, 1986; Sproull, 1986), andere Studien geben an, dass kaum signifikante Unterschiede vorliegen (King & Miles, 1995; Tse, 1998), oder sogar, dass die Rücklaufquote höher ausfällt (Schaefer & Dillman, 1998; Yun & Trumbo, 2000). Gerade ältere Studien zeigen, dass die Responseraten für E-Mails niedriger als die von brieflichen bzw. Paper-and-pencil-Erhebungen ausfallen (Anderson & Gansneder, 1995; Kittleson, 1995).

Verschiedene Begründungen erläutern die niedrigen Rückläufe bei E-Mail Antworten. Erstens können E-Mails einfach verschoben oder erstmal unberücksichtigt bleiben. Zweitens können E-Mails im E-Mail-Programm sichtbar sein und erscheinen meist nicht automatisch auf dem Bildschirm. Drittens ist gerade bei E-Mails dem Empfänger bewusst, dass die Antworten eben nicht anonym sind. Genau genommen handelt es sich bei E-Mail-Adressen um personenbezogene Daten im Sinne der Datenschutz-Grundverordnung und die Beantwortung von E-Mails zeigt typischerweise immer den Antwortenden und ist damit nicht anonym. Dies kann zur Zurückhaltung bei den Antworten führen. Diese Nachteile liegen bei der Online-Befragung nicht vor.

Andere Studien zeigen, dass bei guter Passung der Befragten zur Fragestellung sowie hoher Kohäsion sogar Responseraten bei elektronischer Befragung von 70% erzielt werden können (Anderson & Gansneder, 1995; Bachmann et al., 1996; Dillman, 2007; Kiesler & Sproull, 1986; Parker, 1992; Sax, Gilmartin, & Bryant, 2003; Schaefer & Dillman, 1998; Yun & Trumbo, 2000). Bei reinen Online-Erhebungen, z. B. bei SCIENCEnet-Abonnenten, wurden Responseraten bis zu 76% erzielt (Walsh, Kiesler, Sproull, & Hesse, 1992). Ältere Studien zeigen, dass die Rücklaufquote weiterhin ansteigt, wenn eine vorherige Einladung zur Online-Befragung wie per E-Mail erfolgt (Kaplowitz et al., 2004). Die hohen Rücklaufquoten werden jedoch oft nicht von allen Befragungen erreicht. Es werden auch Responseraten von nur 19% (Schuldt & Totten, 1994), bis 43% (Couper et al., 1997; Couper et al., 1999) oder 21% Swoboda, Mühlberger, Weitkunat, & Schneeweiß, 1997) portiert.

Dabei muss klärend darauf hingewiesen werden, dass nicht nur eine Form von Rücklaufquote bei Online-Befragungen existiert, sondern verschiedene in Abhängigkeit welche Kontaktform gewählt wurde und ob es sich um Filterfragen oder die Gesamtbefragung handelt (Eysenbach, 2004).

Eng mit der Responserate ist die Abbruchquote verbunden. Ein typisches Problem jeglicher Erhebungen sind unvollständige Antworten. Sehr oft übersehen Befragte einzelne Fragen oder überspringen Fragen, die sie erst später beantworten wollen. Dann aber vergessen sie wieder zurück zu gehen und diese Fragen zu beantworten. Darüber hinaus überspringen Befragte einzelne Fragen, weil sie hier keine Antwort geben wollen oder können (Sax et al., 2003). Ferner treten unvollständige Beantwortungen

auf, weil Befragte nicht sonderlich an den Themen der Befragung interessiert sind, im Laufe der Fragen die Lust oder die Geduld zu antworten verlieren (Schmidt & Computers, 1997). Abbruch schränkt letztlich die Aussagekraft ein. Zu lange Fragebögen sind hier besonders gefährdet, hohe Abbruchquoten zu erzeugen. Hohe Attraktivität der Gestaltung, klare Fragen und die Initiierung von Interesse durch gezielte Ansprache in Foren, Plattformen usw. wie in Online-Erhebungen gebräuchlich können z. B. die Abbruchquote verringern.

Der Befragte von Face-to-Face und Telefonbefragungen kann auf Nachfragen und Klärung erhalten. Face-to-Face und Telefonbefragungen haben so quasi Klarstellungen durch den Interviewer eingebaut. Abbruch aufgrund von Ungewissheit und Nichtverstehen sind daher bei Face-to-Face- und Telefonbefragungen in gewissem Maße auffangbar. Freundlichkeit des Interviewers, eine angenehme Stimme und eine motivierende Gesprächssituation im Face-to-Face-Interview und im Telefonat kann den Befragten bei der vollständigen Beantwortung stützen und Abbrüche sowie „ich weiß nicht“ Antworten reduzieren (Duffy et al., 2005). Diese persönliche und menschliche Unterstützung sowie Motivierung ist bei Online-Befragungen stark eingeschränkt. Online-Befragungen können aber Hilfenfenster (Beebe, Mika, Harrison, Anderson, & Fulkerson, 1997), audio-visuelle Sequenzen, Bots zur „intelligenten“ Erklärung oder persönliche Helpdesks per Telefon und Chats einbauen. Zudem kann eine gute Zielgruppenpassung und attraktive Gestaltung die Abbruchgründe von vornherein reduzieren. Insgesamt erfordert die Reduzierung der Abbruchquote ein durchdachtes umfangreiches Erhebungsprogramm bei Online-Befragungen und damit auch höhere fixe Kosten für die Systemumwelt. Möglichkeiten dazu werden nachstehend ausgeführt.

Online-Erhebungen können z. B. mit automatischen Filter- und Plausibilitätsfunktionen genaue Ansprachen und gewünschte Charakteristika von Befragten zulassen. Dies beeinflusst die die Vollständigkeit der Antworten und die Antwortqualität positiv. Der interaktive Aspekt der Befragung ermöglicht zudem ein automatisches Überspringen irrelevanter Fragen. Dies ist gerade ein Vorteil gegenüber schriftlicher und E-Mail basierter Befragung, bei denen die Befragten selbst zur nächsten relevanten Frage springen müssen, falls folgende Fragen durch vorhergehende Antworten ein- oder ausgeschlossen werden. Online-Erhebungen können ähnlich elegant wie Telefonerhebungen - gerade CATI-gestützte - irrelevante Fragen auslassen und so die Genauigkeit der Ansprache und die Vollständigkeit der Antworten stärken. Stanton (1998) zeigt ähnlich, dass weniger fehlende Daten bei Online-Befragungen auftreten, weil die Filterung den Befragten kognitiv entlastet. Allerdings bestehen auch Vorbehalte gegenüber der Filterfunktion, weil dann sehr oft unterschiedlich Frage-Designs und Layouts präsentiert werden (Beebe et al., 1997).

Im Sinne der quasi Inputfaktoren für eine hohe Datenqualität und Rücklaufquote nutzt Civey ein breites Repertoire. Civey setzt unterschiedliche technologische Assistenzsysteme für Online-Befragungen ein. Neben der optisch ansprechenden Gestaltung und dem bereits angesprochenen Widget sind das insbesondere die Visualisierung der Teilnehmer auf einer Landkarte und in Live-Streams, die wesentlichen soziodemografischen Merkmale der Teilnehmer transparent darstellen. Dazu kommt ein aktiver Support, der an sieben Tagen die Woche, Teilnehmerfragen per E-Mail und in den Kommentarfunktionen von Sozialen Netzwerken bzw. Chats beantwortet, Außerdem gibt es eine fortlaufend aktualisierte FAQ-Sektion auf der Webseite sowie eine übersichtliche Darstellung des Verfahrens in Text und mittels erklärender Videos. Dies erlaubt umfangreiche Klärung, Support und Motivation der Befragten.

Eher ergebnisorientiert lassen sich die Quoten der Rückläufe und Registrierungen für Panel beurteilen. Bei einem stufenbezogenen Verfahren und insbesondere beim Einsatz von Filterfragen bei Online-Befragungen – gerade bei einem Riversampling Ansatz ergeben sich unterschiedliche Rücklauf- und

Abbruchquoten und mögliche Verzerrungen, z. B. durch gezielte Einladung sehr aktiver Panel-Teilnehmer, beobachten. Batinic & Moser (2005) berichten für Online-Panel Rücklaufquoten von 20 bis 60 Prozent. Wenngleich solche Methoden bei Civey nicht eingesetzt werden, kann Civey in Befragungen sehr positive Gesamt-Rücklaufquoten (62%) und Registrierungsquoten von 32 Prozent (vgl. Tabelle 1) berichten.

Conversion-Rates

	Rücklaufquote	Registrierungsquote
<b>Gesamt</b>	62 %	32 %
<b>2018 (Stand Q3)</b>	56 %	38 %
<b>2017</b>	76 %	27 %
<b>2016 (Q3 und Q4)</b>	21 %	13 %

**Tabelle 1: Rücklauf- und Registrierungsquoten von Umfragen durch Civey (Angaben von Civey, November 2018).**

Civey erzielt nach eigenen Angaben Rücklaufquoten<sup>2</sup> von 55.5 Prozent im Jahr 2018 bis zu 75.9 Prozent im Jahr 2017. Vergleicht man dies mit Rücklaufquoten aus Vergleichsstudien (Ilieva et al., 2002) ist dies sehr hoch und deutet auf eine hohe Aussagekraft hin (Nulty, 2008). Dabei muss darauf hingewiesen werden, dass lediglich 26.7% (2017) bzw. 37.7% (2018) der primär angesprochenen Internetnutzer auch für die repräsentative Befragungen rekrutiert wurden indem sie eine Registrierung durchführten<sup>3</sup>, welche die Voraussetzung für die Verifizierung darstellt. Eine Näherung an die tatsächliche Rücklaufquote beträgt daher wahrscheinlich 20,9 Prozent im Jahr 2018 bzw. 20,2 Prozent im Jahr 2017. Gegenüber 2017 ist in 2018 eine deutliche Zunahme der Rekrutierungen bei einer gleichzeitig sinkenden Rücklaufquote in Befragungen zu verzeichnen.

Insgesamt deutet die hohe Rücklaufquote bei Civey auf ein durchdachtes Vorgehen hin, das Interesse bei den Befragten hervorruft. Dies deutet zudem auch eine hohe Qualität der Antworten hin und auf Vorteile bei der Repräsentativität der Daten für eine Population, auch für die Gesamtbevölkerung wie bei der Wahlforschung.

### c. Kosten und Aktualität

Die Kosten für Datenerhebung und Analyse sind hoch und in den letzten Jahren zudem weiter angestiegen, weil immer mehr Befragte weniger motiviert („befragungsmüde“) sind an schriftlichen oder telefonischen Erhebungen teilzunehmen. Kleinere ggf. zu kleine Stichproben haben natürlich dann auch ggf. begrenzende Rückwirkungen auf ihre Aussagekraft. Aufgrund der hohen Kosten, des hohen Zeitbedarfs, der Erfordernisse von Wiederholungsschleifen bei der Ansprache und insgesamt der Ressentiments der Individuen teilzunehmen, werden deshalb weniger häufig weniger und seltener und damit leider auch weniger aktuell Daten bei Individuen erhoben oder aber auf eine kleinere Stichprobe zurückgegriffen. Dies erweist sich besonders nachteilig bei Fragen zu politischen und gesellschaftlichen Veränderungen und Trends.

<sup>2</sup> Verhältnis von Votes verifizierter Nutzer zu Umfragetool-Einblendungen

<sup>3</sup> Conversion Rate von first Click im Umfragetool zur Registrierung

Gerade Face-to-Face-Befragungen verursachen hohe bzw. deutlich höhere Kosten als Online-Befragungen (Duffy et al., 2005; Nulty, 2008). Watt, Simpson, McKillop, and Nunn (2002) legen eine Analyse von elektronischen Befragungen im Vergleich zu anderen Erhebungsmethoden vor. Sie zeigt, dass elektronische Befragungen nur ein Drittel der Kosten einer Erhebung mittels Paper and Pencil Fragebogens kosten. Dies führt dazu, dass Online-Befragungen tendenziell größere Stichproben, insbesondere bei gleichen Kosten ermöglichen (Nulty, 2008). Zudem können Fragen schneller angegangen werden, so dass höhere Aktualität und Stichprobengröße erzielbar ist. Watt et al. (2002) verdeutlicht mittels verschiedener Grafiken die Kosten von verschiedenen Methoden: online, E-Mail, Postalisch, sowie Computer Aided Telephone Interviewing (CATI) Erhebungen sowie durch CATI gestützte Telefon- Erhebungen. Dabei wird evident, dass die Ergebnisse von Watt et al. (2002) vorherige Analysen stützen und zwar darin, dass E-Mail und Online-Befragungen drastisch Kosten senken, wenn ceteris paribus die Stichprobengröße steigt. Geringere Kosten bei steigender Stichprobengröße ermöglichen wie bereits angeführt höhere Stichproben und steigern damit deren Aussagefähigkeit. Dies verdeutlicht, dass traditionelle Verfahren wie die briefliche Erhebung oft nicht schnell zu ausreichend hohen Rücklaufquoten kommen. Gerade Online-Verfahren, die auf verschiedenen Webseiten gestreut werden (Riversampling) oder solche die mittels Panels arbeiten, kommen hier tendenziell auf deutlich schnellere und höhere Rückläufe. Das Ergebnis des 'WELT-Trend' zur Fragestellung: 'Sollten „Ankerzentren“ eingerichtet werden, in denen sich Asylbewerber aufhalten müssen, bis über ihren Antrag entschieden wurde?' basiert auf einer Stichprobe von 5051 Befragten zwischen dem 3. und 5. Mai 2018 WELT.de, 05.05.2018. (vgl. Abbildung 2)



**Abbildung 2: Verhältnis der festgesetzten Stichprobengröße und der aktiven Teilnehmer als Darstellung in einer laufenden Umfrage von Civey (Angaben von Civey, November 2018).**

Civey führt eine Reihe von Erhebungen mit hoher Schnelligkeit durch und ist in der Lage relativ große Stichproben zu erzeugen, wie die folgende Tabelle 2 aufzeigt.

Neueste 10 Umfragen im Oktober 2018 (Befragte: Bundesbürger über 18 Jahre)

Umfrage	Teilnehmer	Ziel-Stichproben-größe	Tage bis Stichprobe erreicht wurde	Umfrage veröffentlicht
Machen Sie sich Sorgen um Ihre Privatsphäre, wenn Videoüberwachung mit Gesichtserkennung in öffentlichen Verkehrsmitteln eingesetzt wird?	14.044	5.000	0,19	2018-10-30 13:37:00.752
Wofür sollte sich die Bundesregierung umweltpolitisch besonders einsetzen?	9.442	5.000	0,22	2018-10-30 13:07:56.602
Wie haben Ihrer Meinung nach die Regierungen unter Angela Merkel als Kanzlerin Deutschland insgesamt verändert?	15.068	5.000	0,31	2018-10-30 09:13:09.693
Wie bewerten Sie, dass in Deutschland zunehmend Halloween gefeiert wird?	13.476	5.000	0,25	2018-10-29 15:42:58.365
Wie bewerten Sie, dass Angela Merkel nicht erneut für den Parteivorsitz der CDU beim kommenden CDU-Parteitag kandidieren wird?	13.619	5.000	0,32	2018-10-29 14:20:10.722
Sollte Angela Merkel bis zum Ende der Legislatur Kanzlerin bleiben, nachdem sie nicht erneut für den CDU-Parteivorsitz kandidieren wird?	13.166	5.000	0,40	2018-10-29 13:19:28.435
Welche/r Politiker/in sollte am ehesten in der Nachfolge Angela Merkels CDU-Vorsitzende/r werden?	15.925	5.000	0,22	2018-10-29 10:47:04.623
Bedroht der Populismus Ihrer Meinung nach die Demokratie weltweit?	14.579	5.000	0,16	2018-10-29 08:37:35.518
Wie bewerten Sie den Ausgang der hessischen Landtagswahl vom 28.10.2018 insgesamt?	8.611	5.000	1,04	2018-10-29 07:48:49.002
Für welche Partei stellt der Wahlausgang der hessischen Landtagswahl 2018 Ihrer Meinung nach den größten Erfolg dar?	13.362	5.000	0,35	2018-10-29 07:10:28.606
Ist Kanzlerin Angela Merkel durch das Ergebnis der Landtagswahl in Hessen Ihrer Meinung nach gestärkt oder geschwächt?	13.160	5.000	0,35	2018-10-29 07:05:46.879

**Tabelle 2: Verhältnis der Teilnehmer, der festgesetzten Ziel-Stichprobengröße sowie der Laufzeit in aktuellen Umfragen von Civey (Angaben von Civey, November 2018).**

Aktualität ist zentral für die Wahl- und Meinungsforschung. Die Möglichkeit bei Online-Befragungen kurzfristig Aussagen zu relevanten Fragestellungen und Themen (wie sie z. B. im Bereich der Meinungsforschung und in der Wahlforschung) zu generieren, zeigt eine große Stärke von Online-Erhebungen generell und seitens Civey im Speziellen. Gerade briefliche und schriftliche Erhebungen weisen wie bereits angeführt im Vergleich dazu höhere zeitliche Verzögerungsraten auf und sind kaum geeignet für die aktuelle Wahlforschung. Oft liegen die Fragebogen unbeantwortet auf Schreibtischen, bis sie wieder hervorgeholt werden oder gar im Papierkorb landen. Ähnliches gilt für Befragungen in E-Mails, die oft direkt gelöscht, ignoriert oder verschoben werden. Meist muss mittels Nachfass-Aktionen die zu befragende Person erneut motiviert werden an der Befragung teilzunehmen. Eine Ausnahme bilden Panels, bei denen die Teilnehmer meist relativ schnell angesprochen werden und die eine tendenziell schnellere Antwortgeschwindigkeit haben. Aufgrund der gewissen Verzögerung sind die oben genannten Verfahren im Regelfall nicht besonders für aktuelle Meinungsforschung und Wahlforschung geeignet.

Im Gegensatz dazu können telefonische Erhebungen und Online-Erhebungen mit wenig Zeitverzögerung durchgeführt werden, wenn bzw. sobald die Befragten teilnehmen und besonders,

wenn die Teilnehmer bereits an einem Panel registriert sind. Bei Telefonerhebungen bedarf es dabei oft mehrfachen Anrufens und bei Online-Befragungen der mehrfachen Einladung (z. B. mittels Telefon oder E-Mails) oder dem erneuten Vorschlagen in Anzeigenfenstern in Foren und auf Plattformen. Auch dann ist jedoch keine Antwort von jeder angesprochenen Person zu erwarten. Allerdings sind beide Verfahren gut geeignet schnell und relativ flexibel Antworten auf aktuelle Fragen zu liefern. Die Online-Befragung kann durch die Automatisierung und die geringeren flexiblen Kosten hier aber Vorteile in Punkto einer hohen Geschwindigkeit für größere Stichproben ausspielen.

Die geringeren Kosten von Online-Erhebungen bei großen Stichproben verbunden mit der Automatisierung der Ansprache sowie leicht durchführbaren Anpassungen erlauben damit tendenziell einen schnellen Rücklauf und eine hohe Aktualität von Online-Erhebungen. Gerade in Methodenkombination mit Panels oder anderen Ansprachen sind Online-Erhebungen besonders geeignet schnell und kostengünstig Antworten und eine größere Stichprobe zu erreichen.

Vorteile bei Online-Befragungen wie auch bei Civey liegen wie bereits erläutert in deren Schnelligkeit und Adaptionfähigkeit des Erhebungsinstruments. So lassen sich Erhebungen schnell implementieren, wieder ändern und sich sehr aktuell anpassen. Ebenfalls lassen sich schnell gegebenenfalls sogar sofort nach der Beantwortung der Fragen die aktuellen Auswertungen und anderes Feedback an den Befragten platzieren (Yun & Trumbo, 2000). Face-to-Face und Telefonbefragungen können dies auch, erfordern dann aber wiederum eine umfangreiche Schulung von Interviewern, die ggf. etwas länger dauern kann und hohe Kosten verursacht.

Wie schon dargestellt bietet das Widget von Civey eine Möglichkeit zur Befragung innerhalb kurzer Zeit und mit einer sehr großen Teilnahmebereitschaft. Die Darstellung von Ergebnisprognosen in Echtzeit kann die Attraktivität und Motivation zur eigenen Stimmabgabe sowie die Entschlossenheit zu einer Aussage steigern. Dies kann jedoch, begleitet von einer als geringer wahrgenommenen realen Konsequenz der eigenen Stimmabgabe, zu starken Abweichungen von Ergebnissen aus realen und geheimen Wahlen bzw. Abstimmungen führen. Für die o.g. Frage zu Ankerzentren legte Civey eine Zielpopulation anhand soziodemografischer Merkmale und a-priori eine Stichprobengröße von  $N=5000$  fest. A-posteriori wurde die Stichprobengröße erweitert, sodass die Verteilung der Soziodemografie die Annahmen der Zielpopulation besser repräsentiert. Beachtenswert ist dabei, dass die Umfrage in einem Befragungszeitraum von zwei Tagen insgesamt 59237 Mal beantwortet wurde. Diese Zahlen weisen, zuzüglich nicht registrierter und daher nicht aufgeführter Umfrageteilnehmer, auf ein hohes Interesse zur Befragung hin.

Die Ergebnisse zeigt Civey in unterschiedlicher Form auf. Als "Rohdaten" bezeichnet Civey die addierten Antworten von Umfragen während der gesamten Erhebungszeit. Als "repräsentativ" werden von Civey diejenigen Ergebnisse dargestellt, welche nach Zielgruppenbestimmung und Stichprobenabgleich mit verifizierten Nutzern innerhalb a-priori festgelegten Zeitraumes erzielt werden. Dies zeigt deutlich, dass schnell Daten gewonnen wurden und diese auch sensibel verarbeitet werden, um eine möglichst gute Repräsentativität herzustellen.

Die geringen flexiblen Kosten und die Schnelligkeit bei Automatisierung der Ansprache, Methodenkombination sowie die Integration von Panels zeigen starke Vorteile der Systematik bei Civey auf. Die real erreichten hohen und schnellen Rücklaufquoten bei aktuellen Fragen belegen dies empirisch.

#### d. Technische Voraussetzungen, Multi-medialität und Darstellung

Der Einsatz textlicher und grafischer Darstellungen im Internet hat sich in den letzten Jahren durch die Implementierung plattformübergreifender Applikationen und einheitlicher Software- und Hardwaresysteme angeglichen und kann daher umfangreicher eingesetzt werden. Zuvor waren einheitliche und damit vergleichbare Darstellungen von Befragungen im Web nicht immer erreichbar (Bishop, 1997; Buchanan & Smith, 1999) und dies konnte negative Auswirkungen auf die Antwortqualität und die Repräsentativität haben (Cook et al., 2000). Dies galt insbesondere dann, wenn einzelne Befragte oder bestimmte Befragtengruppen aufgrund von Darstellungsformaten nicht antworten oder abbrechen konnten (Sills & Song, 2002). In Online-Befragungen sind Bilder und Textpassagen sowie interaktive und filmische Sequenzen zu finden (Yun & Trumbo, 2000), sie erlauben Mehrfarbigkeit und animierte Darstellung, die als Verbesserung gegenüber E-Mail oder Paper -nd pencil-Befragungen angesehen werden (Bishop, 1997). Gerade gegenüber Telefonbefragungen sind ganz andere textliche oder bildliche Darstellungen möglich (Yun & Trumbo, 2000).

Aus der Einbettung audiovisueller Medien (Bishop, 1997), wie z. B. Grafiken und Filmen in Online-Inhalten, ergibt sich eine gesteigerte Attraktivität, die auch bei der ansprechenden und interessanten Aufbereitung von Online-Erhebungen genutzt werden kann (Yun & Trumbo, 2000). Die Verwendung guter Grafiken verstärkte bereits bei schriftlichen Befragungen mit Fragebögen die Antwortqualität, während bei E-Mail-Befragungen oft auf große Grafiken verzichtet wurde. Online-Befragungen können hier ganz andere Wege gehen und die Befragten mit umfangreichen audio-visuellen Erläuterungen bei der Beantwortung und Motivation weiter zu machen unterstützen. Die optische Attraktivität und die hohen Potentiale für eine gezielte Ansprache von Zielgruppen durch Online-Erhebungen kann zu höherem Interesse und Involvement bei den befragten Personen führen (Kiesler & Sproull, 1982; Kiesler & Sproull, 1986; Martin & Nagao, 1989; Sproull & Kiesler, 1986; Walsh et al., 1992). Dies reduziert die Abbruchquote und erhöht die Rücklaufquote, aber vor allem werden positive Wirkungen auf die Antwortqualität aufgezeigt. Studien erläutern z. B., dass Befragte bei Online-Erhebungen kognitiv stärker eingenommen sind, weniger Ablenkung erfahren und sich mehr auf die Befragung konzentrieren (Kiesler & Sproull, 1986; Sproull & Kiesler, 1986).

Dazu tritt die gute Zielgruppenansprache, wenn Communities, Foren und Plattformen bei Online-Befragungen gezielt angesteuert werden (Bennemann, 1999; Faraj et al., 2011; Fisher, 2018). Befragungen können hier gezielt bestimmte Bevölkerungsgruppen ansprechen und sind in der Lage auch schon schwer erreichbare Personengruppen anzusprechen. Dies kann sehr gut auch mit Panelteilnahmen kombiniert werden, wenn dazu verschiedene Charakteristika der Teilnehmer miterhoben wurden und dann für die Selektion und für die Berechnung mitgenommen werden.

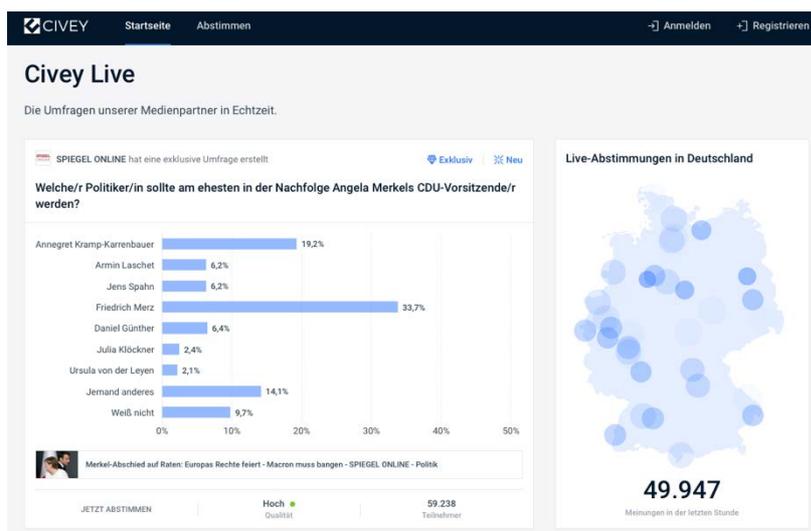
Vorteile der Online-Befragungen liegen also im ansprechenden Design im Allgemeinen, der zielgruppenspezifischen Gestaltung im Speziellen plus der Lokalisierung der Umfrage, auch in den entsprechenden Foren und Plattformen. Sie ermöglichen eine Motivierung an der Befragung teilzunehmen und dies auch bei Personengruppen, die weniger stark das Internet nutzen (Fan & Yan, 2010). Attraktives Design, zielgruppenspezifische Gestaltung und die richtige Lokalisierung ermöglichen zudem eine schnelle Ansprache von Personen, auch den weniger intensiven Internetnutzern. Dies bietet letztlich auch positive Einflüsse auf die Repräsentativität.

Allerdings bedarf es bei vielen Fragestellungen z. B. über Wahleinstellungen zunächst einmal zur Herstellung des Verstehens bei den Befragten keiner besonderen Medien. Auch wenn die Implementierung audio-visueller Instrumente in Online-Befragungen sehr gut möglich ist, mag es gar nicht erforderlich sein oder den Befragten gar überfrachten. Bei einfachen Fragen vermögen Online-Erhebungen daher einen Teil ihrer spezifischen medialen Vorteile nicht gegenüber anderen

Befragungsformen so sehr auszuspielen. Gegen eine mögliche Überfrachtung von Befragten mit audio-visuellen ‚Gimmicks‘ hilft jedoch eine gute Planung und Klärung der Befragungskontexte und -ziele. Die Befragung muss ja nicht audio-visuell überfrachtet werden. Eine Beurteilung von Online-Befragungen erfordert so wieder die Betrachtung ihres Kontextes und ihrer Ziele.

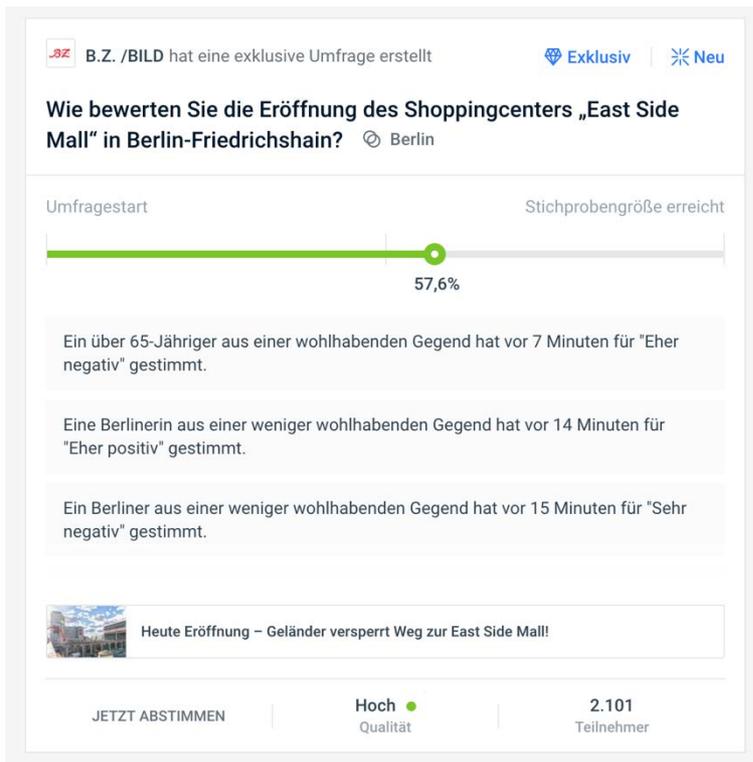
Allerdings erlaubt eine ansprechende und zielgruppenorientierte multi-mediale oder gar nur optisch attraktive Darstellung die Motivation der Personen zur Teilnahme zu stärken. Optisch ansprechende und zielgruppenorientierte Online-Befragungen können daher mit höherer Geschwindigkeit Antworten erzielen und dies auch bei Personengruppen, die sonst unterrepräsentiert wären.

Civey führt eine Reihe unterschiedlicher Studien mit verschiedenem Grad an Komplexität durch und setzt dafür durchdacht unterschiedliche gestalterische und technische Lösungen ein. Neben den audio-visuellen Gestaltungsformen sowie verschiedensten Erklärungen und Interaktionsformaten ist besonders das Einbetten von aktuellen Trends zu nennen. So werden beispielsweise während laufender Umfragen in Echtzeit die Anzahl der abgegebenen Stimmen (sogenannte votes), deren regionale Verteilung sowie vorübergehende Prognosen zum Ausgang der Befragung dargestellt (vgl. Abbildung 3). Dies erhöht das Interesse der Befragten teilzunehmen. Die Auswahl von Portalen, Foren und Plattformen auf denen die weniger internetaffinen Personen erlaubt die Ansprache und die Antworten von sonst unterrepräsentierten Gruppen. Dies alles trägt zu einer hohen Aussagekraft bei und steigert die Repräsentativität.

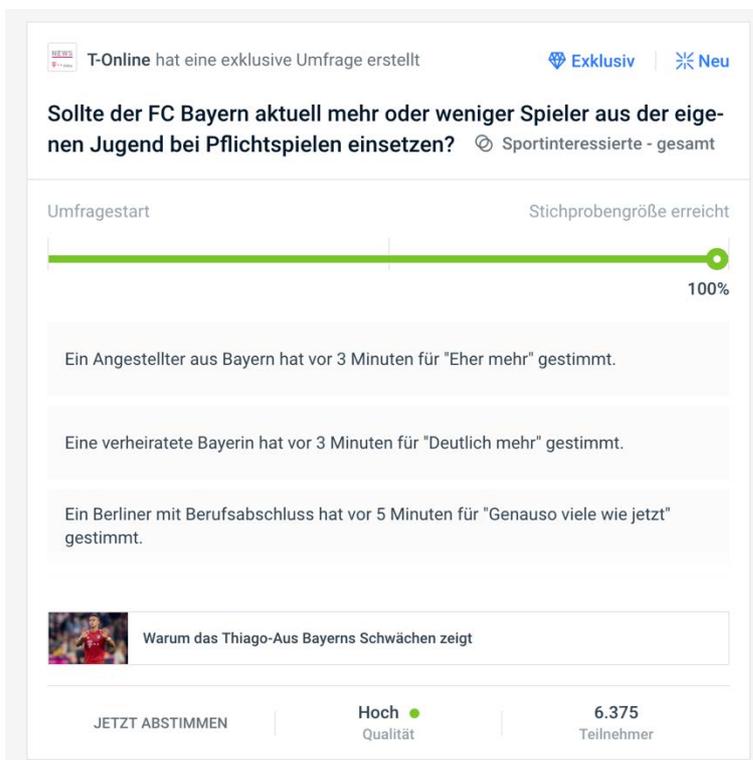


**Abbildung 3: Interaktive Darstellung einer laufenden Umfrage von Civey (Abbildung und Angaben von Civey, November 2018).**

Insgesamt nutzt Civey ein großes Portfolio audio-visueller Gestaltungsformen für die Online-Befragung. Die Elemente werden durchdacht eingesetzt, damit die Befragten nicht überfrachtet werden, aber dennoch ausreichend informierende und motivierende Reize erhalten. Sehr innovativ und wirkungsvoll ist die interaktive Einbettung der laufenden Umfragetrends. Die Wirksamkeit dieser gestalterischen und interaktiven Elemente spiegelt sich letztlich auch in den Werten zur schnellen Response und hohen Responderaten wider.



**Abbildung 4: Erstes Beispiel - Live-Stream von Teilnehmern bei einer Umfrage auf einer Plattform für weniger internetaffine Teilnehmer (Abbildung und Angaben von Civey, November 2018).**



**Abbildung 5: Zweites Beispiel - Live-Stream von Teilnehmern bei einer Umfrage auf einer Plattform für weniger internetaffine Teilnehmer (Abbildung und Angaben von Civey, November 2018).**

#### e. Mehrfacherhebung und Adaption

Face-to-Face und Telefonerhebungen erlauben die direkte und kontrollierte Ansprache von Personen. Probleme von Mehrfachantworten sind kaum zu befürchten. Gerade elektronische Befragungen beinhalten aber das Problem von Mehrfachantworten seitens der Befragten bzw. dem automatisierten Nichtzulassen von Antworten (Schmidt, 1997). Typischerweise erlauben Online-Erhebungen allerdings nur eine Antwort von einer IP Adresse und könnten so das Problem der Mehrfachantworten umgehen. Dabei wird aber unterstellt, dass Antworten von einer IP-Adresse auch nur von derselben Person kommen. Tatsächlich aber werden IP-Adresse auch von mehreren Personen genutzt. Antworten von derselben IP-Adresse würden dann nicht mehr berücksichtigt, so dass Vorkehrungen getroffen werden müssen dies zuzulassen. Eine recht etablierte Methode stellt allerdings die Nutzung von Cookies da, die jedem Befragten eine spezifische Identität zuweisen (Yun & Trumbo, 2000).

Civey stellt nach eigenen Angaben sicher, dass Mehrfachantworten nicht berücksichtigt werden. Mit der Durchführung der Registrierung, der Abfrage spezifischer Marker (Geburtsjahr, Geschlecht, Postleitzahl, Bildungsstand und Familienstand; vgl. Civey GmbH, 25.5.2018) sowie der Einverständniserklärung zur Datenverarbeitung setzt Civey ein Browser-Cookie beim Nutzer. Im folgenden Verifizierungsprozess wird überprüft, ob es sich bei dem registrierten Teilnehmer um eine natürliche Person handelt. Die Überprüfung erfolgt mittels der Hidden-Captchas-Technologie, die anhand von Nutzerreaktionen bewertet, mit welcher Wahrscheinlichkeit die Aktionen von einer realen Person ausgeführt werden (Moradi & Keyvanpour, 2015). Während des Verifizierungsprozesses werden weitere spezifische Marker (beispielsweise Bildung, Beruf, Einkommen, Religion oder Familienstand) abgefragt und diese Marker müssen von Teilnehmer bis zu einem gewissen Grad vollständig hinterlegt werden. Aus den Ergebnissen der Hidden Captcha-Analyse, der Widersprüchlichkeit von Angaben und Antworten sowie aus der Vollständigkeit der Informationen bildet Civey einen sogenannten "User-Trust-Index", der zur Steigerung der Stichproben- und Datenqualität eingesetzt wird.

#### f. Zugang zu besonderen Gruppen

Wie bereits angesprochen erlaubt das Internet prinzipiell einen Zugang allen Personen, die technisch über einen Provider, Hub, Wifi-Hotspot usw. sowie über ein stationäres oder mobiles Endgerät angebunden sind. Besonders erlaubt das Internet über eine elektronische Verbindung den Zugang (z. B. Internet als Rekrutierungsinstrument, vgl. Birnbaum, 2004) zu Bevölkerungsgruppen, die über andere Kanäle kaum mehr zu erreichen sind. Interessant für gezielte Erhebungen sind dabei die vielen oft spezialisierten Communities im Internet, die einen direkten Zugang zu Themen und Zielgruppen bieten (Garton, Haythornthwaite, & Wellman, 1997; Garton & Wellman, 1995). Sie können dann auch den Zugang zu schwer erreichbaren Personengruppen erlauben.

Bereits eine größere Anzahl von Forschungen zu Communities zeigen die hohe Bedeutung von Kommunikation, Information, sozialem Austausch und Meinungsbildung im Internet (Brown & Duguid, 1991; Fisher, 2018; Gruner et al., 2014; West & Lakhani, 2008). Befragungen in Diskussionsgruppen und Foren erlauben überdies die Befragung von Individuen, die sehr involviert diese Themen diskutieren (Bagozzi & Dholakia, 2006). Im Kern impliziert dies, dass Online-Erhebungen einen Zugang zu Individuen bieten, die spezifische Interessen Einstellungen und Werten teilen und sich weiter austauschen, wobei sich individuelle und gemeinschaftliche Meinungen bilden. Die Teilnehmer können dabei in anderen Online, aber auch Offline-Communities als Meinungsführer agieren. Unternehmen nutzen bereits seit einiger Zeit den Zugang zu Nutzern für Ideengewinnung und Neuproduktentwicklungen (Olson & Bakke, 2001; von Hippel, 1986). So ist der Zugang über Foren und Communities oft eine kaum zu ersetzende Möglichkeit Daten über und während des Meinungsbildungsprozess (sogar in frühen Phasen) zu erfassen und Daten bei Meinungsführern zu

erheben (Mathes & Czaplicki, 1993; Noelle-Neumann et al., 1998). Online-Erhebungen erlauben insofern Zugang zu Daten und Personengruppen zu einem frühen Zeitpunkt, von Meinungsführern und unter Umständen zu Personen, die vor allem in diesem Medium agieren und kaum mittels Telefon oder Brief ansprechbar sind. Meinungsführer haben dabei auch eine hohe Bedeutung für die Wahlforschung (Mathes & Czaplicki, 1993; Noelle-Neumann et al., 1998c).

Selbst ältere Personen, die das Internet tendenziell weniger nutzen (ARD/ZDF, 10.10.2018; Blasius & Brandt, 2010; Maintz, 2003), sind in Foren und Communities erreichbar (Joshi & Knight, 2015; Martins & Shalley, 2011; Milliken, 1990; Thatcher, 2013). Für wissenschaftliche Studien zeigen verschiedene Autoren die Nutzbarkeit von SeniorNet. Allerdings ist die Durchdringung von Intensivnutzern des Internets bei Älteren noch geringer als in anderen demografischen Gruppen ausgeprägt (ARD/ZDF, 10.10.2018; Blasius & Brandt, 2010).

Interessanterweise erlauben Online-Befragungen einen Zugang zu den sonst schwer erreichbaren stigmatisierten Gruppen. So lassen sich selbst (dauerhaft) erkrankte Personen und Bevölkerungsgruppen mit Behinderungen oder Migranten mittels Web-basierten Erhebungen erreichen. Besonders gut gelingt dies, wenn diese Gruppen in spezifischen Communities agieren. Dabei erlaubt das Internet auch Zugang zu Menschen, die sehr zurückhaltend bei Face-to-Face oder persönlicher Interaktion (auch Telefon) sind (Braithwaite, Waldron, & Finn, 1999). Ganz besonders gilt dies wiederum für stigmatisierte Gruppen und solche, die eine allgemein weniger akzeptierte Meinungen haben. Hier besteht zudem eine enge Verbindung zu dem bei Online-Erhebungen weniger oder kaum vorliegendem Problem der sozialen Erwünschtheit. Online-Befragungen erlauben eine bessere Integration von divergenten und stigmatisierten Meinungen bei Minderheiten. Vor dem Hintergrund der steigenden Diversität und Migration von Bevölkerungsgruppen aus anderen geografischen Regionen oder Religionen wird dies weiter relevant. Zwar liegen bisher wenig Studien über den Mediengebrauch von Personen mit politischen Spezialeinstellungen, Migrations- oder Asylhintergrund vor, dennoch haben andere kulturelle und religiöse Hintergründe Wirkung auf das Antwortverhalten. Beispielsweise diskutiert Muhtaseb and Frey (2008), dass gerade im arabischen Raum und bei Menschen mit diesem ethnisch/kulturellen/religiösen Hintergrund viele Ressentiments gegenüber direkter Kommunikation mit Fremden oder im öffentlichen Raum generell bestehen. Diese Problematik kann mittels automatisierter – nicht persönlicher Ansprache im Internet teilweise umgangen werden. Besonders negativ schlägt sie sonst im Face-to-Face Interview zu.

So können Online-Befragungen schlecht erreichbare Personengruppen ansprechen und gerade darüber eine Repräsentativität auch für Bevölkerungsgruppen erzielen, die sonst unterrepräsentiert sind. Repräsentativität lässt sich insofern im Besonderen herstellen, wenn und weil Online-Befragungen eine große Breite an Personen und Varianz wie Geschlecht, sozio-ökonomischer Status, Ort, Alter und kultureller Hintergrund, physischen Einschränkungen usw. erschließen können (Gosling et al., 2004).

Die Befragungen von Civey nutzen durch die breite und dennoch gezielte Streuung von Befragungen auf verschiedensten Websites den Zugang zu vielen und auch schwer erreichbaren Personengruppen. Dies gilt auch in demografischer Hinsicht. Obwohl 100 Prozent der unter 30-Jährigen Online-Dienste nutzen (Koch & Frees, 2017), sind diese Altersklassen im aktiven Panel von Civey nicht überrepräsentiert. Dies könnte mit einem geringeren Interesse an gesellschaftlichen oder politischen Fragestellungen oder wachsenden Bedenken gegenüber kommerziellen Datensammlungen zusammenhängen. Die prozentuale Häufigkeit der Altersklassen<sup>4</sup>, des Familienstandes<sup>5</sup> und der

---

<sup>4</sup>  $\chi^2=7.14$ ;  $df=4$ ;  $p=0.129$

<sup>5</sup>  $\chi^2=3.73$ ;  $df=2$ ;  $p=0.155$

Herkunft<sup>6</sup> (alte Bundesländer versus neue Bundesländer) zeigen keine bedeutsamen Abweichungen der registrierten Teilnehmer von den im Mikrozensus des Statistischen Bundesamtes berichteten Bevölkerungsdaten. Hingegen weichen die Werte der Geschlechterverteilung<sup>7</sup>, des Bildungsabschlusses<sup>8</sup> und der Berufstätigkeit<sup>9</sup> von den Bevölkerungsdaten ab (eigene Berechnungen).

#### g. Repräsentativität im engeren Sinne

Repräsentativität im Sinne der Übertragbarkeit spezifischer Befragungsergebnisse ist ein zentrales Kriterium für die Qualität statischer Aussagen über Populationen mittels Stichproben. Bereits die vorherigen Abschnitte haben Aspekte und Einflussfaktoren auf die Repräsentativität diskutiert. Hier wird noch einmal vertieft auf die Thematik Repräsentativität im engeren Sinne eingegangen.

Im engeren Sinne ist Repräsentativität nur bei echten Zufallsstichproben aus der Grundgesamtheit (z.B. Zensus) gewährleistet, in denen jede Einheit in der Grundgesamtheit eine gleiche (oder zumindest bekannte) Wahrscheinlichkeit aufweist, ausgewählt zu werden.

Gegen das Kriterium der *methodischen Repräsentativität* wird (strenggenommen) bei Online- und telefonischen Befragungen verstoßen, da nicht die gesamte Grundgesamtheit erreicht wird und somit keine echte Zufallsauswahl stattfindet. Die Rahmenbedingungen des Zentralverbandes der deutschen Werbewirtschaft ZAW e.V. schlagen vor, dass eine Verallgemeinerung von Ergebnissen (z. B. in Hochrechnungen) vorgenommen werden darf, wenn eine *hinreichende Repräsentativität* der Stichprobe für die Grundgesamtheit vorliegt. Hinreichende Repräsentativität wird angenommen, wenn bei einer Stichprobenziehung auf Basis von Auswahlgrundlagen mindestens 85% der Grundgesamtheit abgedeckt werden (ZAW, 2018, S. 14).

Als gängige Vorstellung von Repräsentativität lässt sich auffassen, dass die Teilpopulation so auszuwählen ist, dass aus ihren Ergebnissen so exakt und sicher wie möglich auf die Relationen in der Zielpopulation geschlossen werden kann. Das gilt meist dann, wenn die Teilpopulation ein relativ gutes, wenn auch verkleinertes Abbild der Gesamtpopulation darstellt (Lippe & Kladroba, 2002). Das Problem liegt allerdings darin, dass die interessierenden Merkmale der Gesamtpopulation eben nicht bekannt sind – sonst würde sich die Schätzung ja auch erübrigen. Praxis ist daher auf andere Merkmale auszuweichen und solche zu verwenden, die mit den zu untersuchenden in hoher Korrelation stehen (Backhaus & Blechschmidt, 2009; Backhaus, Blechschmidt, & Eisenbeiß, 2006; Backhaus, Erichson, Plinke, & Weiber, 2003; Lippe & Kladroba, 2002).

Verschiedene Studien haben bereits die methodische Repräsentativität von E-Mail Erhebungen thematisiert und auch gewisse Zweifel geäußert (Dillman, 2007; Dillman et al., 1996; Schaefer & Dillman, 1998; Swoboda et al., 1997; Tse et al., 1994). Die Zweifel sind zum Teil auch nicht ganz unberechtigt, weil nicht alle Personen ähnlich stark E-Mails nutzen oder überhaupt darüber zu erreichen sind, selbst wenn sie über ein E-Mail-Konto verfügen. Insgesamt ist allerdings eine starke Veränderung des Kommunikationsverhaltens im privaten Raum zu konstatieren. Individuen kommunizieren intensiv über Texte und Bilder in Messenger-Diensten. So werden E-Mails oft nur sporadisch und dann nur zur Autorisierung von Messenger Diensten oder für die Autorisierung bei Plattformdiensten genutzt. Selbst im geschäftlichen Raum ergänzen Messenger-Dienste die Kommunikation über das Telefon und die als formaler wahrgenommenen E-Mails. Diese Entwicklung tritt zwar erst seit ein paar Jahren verstärkt

---

<sup>6</sup>  $\chi^2=0.01$ ; df=1; p=0.944

<sup>7</sup>  $\chi^2=6.44$ ; df=1; p=0.011

<sup>8</sup>  $\chi^2=85.56$ ; df=2; p=0.000

<sup>9</sup>  $\chi^2=36.88$ ; df=4; p=0.000

auf. Dennoch wurde in älteren Studien bereits die Erreichbarkeit durch E-Mails diskutiert. Der Hintergrund der Vorbehalte damals war aber ein anderer. Es gab schlichtweg weniger E-Mail-Konten und die Nutzung von elektronischer Kommunikation war noch nicht so etabliert. Tse (1998) pointiert das Problem, in dem E-Mail Befragungen natürlich auf E-Mail Nutzer beschränkt seien. Ein weitergehendes Problem liegt vor, wenn bei E-Mail Befragungen Bevölkerungsgruppen der Mittelschicht überrepräsentiert seien (Mehta & Neale, 2005; Mehta & Sivadas, 1995). Umgekehrt war noch vor 10 Jahren die Internetnutzung noch weniger stark in der Bevölkerung verbreitet damit die Repräsentativität von E-Mail Befragungen für die Gesamtbevölkerung weniger gegeben. Dennoch können die Befragungsergebnisse kontextualisiert werden und durch zusätzliche Informationen in ihrer Repräsentativität gesteigert werden. Ein Indiz dafür ist, dass viele wissenschaftliche Studien E-Mail-Adressen aus News Groups verwendet haben, um Aussagen für die Gesamtpopulation zu treffen (Swoboda et al., 1997). Diese Verwendung findet nicht uneingeschränkte Zustimmung. Durch die zunehmende Internetdurchdringung (ARD/ZDF, 10.10.2018) nehmen diese Begrenzungen aber kontinuierlich ab.

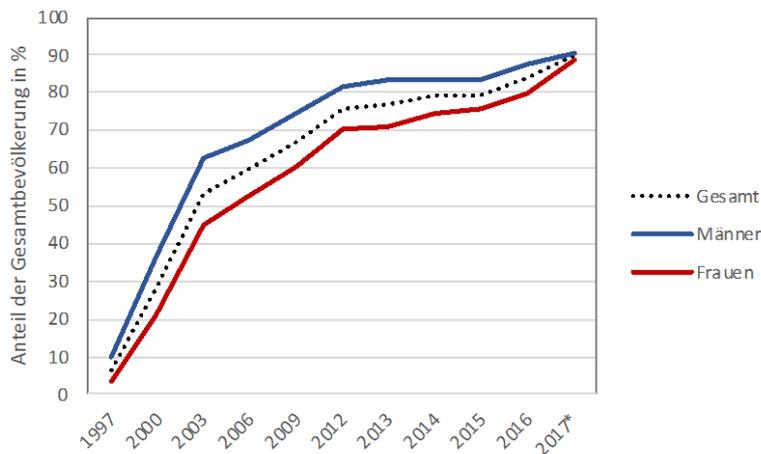
Gerade die Benutzung von E-Mails aus News Groups für Befragungen ist mit Imitationen für die Aussagefähigkeit für Gesamtpopulationen verbunden, gerade wenn sie nicht um Stichproben-Frames oder Quoten korrigiert werden (Bachmann et al., 1996). E-Mail basierten Befragungen wird aber dennoch in verschiedenen Studien eine höhere Antwortrate für viele Bevölkerungsgruppen bescheinigt (Yun & Trumbo, 2000). Swoboda et al. (1997) verwenden für E-Mail basierte Erhebungen eine andere Begründung. Swoboda et al. (1997) zeigen auf, dass 41% der Respondents einer elektronischen Befragung nicht auf eine Telefonbefragung geantwortet hätten. Demzufolge können E-Mail basierte und vor allem Online-Erhebungen nicht nur gute Responseraten generieren, sondern vor allem auch Bevölkerungsgruppen erreichen, die mittels brieflicher oder telefonischer Erhebungen nicht erreichbar sind. Dies verdeutlicht, dass E-Mail Erhebungen nicht allein verwendet werden sollten, wenn es um Aussagen für die Gesamtbevölkerung geht, aber dass sie als sehr gute Ergänzung von andere Erhebungsmethoden anzusehen sind. Diese Argumentation deutet zudem auf die Vorteilhaftigkeit von mehr „multi-node und multi-mode“ Erhebungen hin.

Online-Erhebungen werden zwar auch elektronisch übermittelt wie E-Mails und verlangen den Zugang zu dem Internet. Jedoch sind sie anders angelegt. Zunächst einmal umgehen Online-Erhebungen die Beschränkung auf Kenntnis und Zugang zu E-Mail-Adressen seitens des Befragenden, wenn sie auf Webseiten, Foren, usw. implementiert werden. Da Online-Erhebungen aber naturgemäß über das Internet ablaufen gelten auch hier Fragen und Limitationen des Zugangs zum Internet und zur Nutzungsintensität. Menschen, die keinen Zugang zum Internet haben sind selbstverständlich so kaum nicht erreichbar, auch nicht über das Ausdrucken der E-Mail oder des Links. Darüber hinaus ist die Erreichbarkeit von Menschen, die das Internet wenig nutzen oder nur sehr zurückhaltend stark eingeschränkt. Dies aber nimmt wie erwähnt immer mehr ab (ARD/ZDF, 10.10.2018).

Noch vor 20 Jahren konnte man bei Online-Erhebungen kaum von (hoher) Repräsentativität ausgehen. Schmidt (1997) argumentiert, dass im Web junge Männer mit höherem sozio-ökonomischen und Bildungsstand als Nutzer überrepräsentiert und dass dies insbesondere für die Nutzung von Webforen gilt. Demzufolge sind zumindest vor dem Jahrtausendwechsel noch wenig für die Gesamtbevölkerung repräsentative Aussagen mittels Erhebungen aus Webforen zu erwarten gewesen. Dies gilt zumindest, wenn die Erhebungen allein im Internet bzw. in Webforen abgelaufen sind. Nun aber über 20 Jahre sind diese Zweifel nur noch sehr begrenzt zu begründen. Seit über zwei Dekaden ist die Nutzung des Internets massiv angestiegen und hat sich weit ausgefächert. Im Internet sind unzählige Themen in verschiedensten Formaten vertreten. Somit lassen bei durchdachter Ansprache vielfältige Bevölkerungsgruppen ansprechen. Limitationen der Bewertung der Repräsentativität sind heutzutage

weitgehend zu revidieren. Jedoch kann über gezieltes Ansprechen von Personen mittels Voransprache über E-Mail, Brief bzw. über zielgruppenorientiertes Platzieren in spezifischen Foren, Plattformen sowie dem Design der Befragung auch diese Problematik weitgehend reduziert werden.

Ein Argument für die niedrigere Repräsentativität bei elektronisch übermittelten vor allem Online-Erhebungen kann wie bereits eine demographische Verteilung der Internetnutzer sein, die nicht der Verteilung der Bevölkerungsgruppen entspricht. Ältere Studien zeigen z.B. noch, dass Frauen im Internet unterrepräsentiert sind, aber auch sie zeigen ein starkes Ansteigen der Nutzung des Internets durch Frauen (McPhee & Lieb, 1999) (vgl. Abbildung 5).

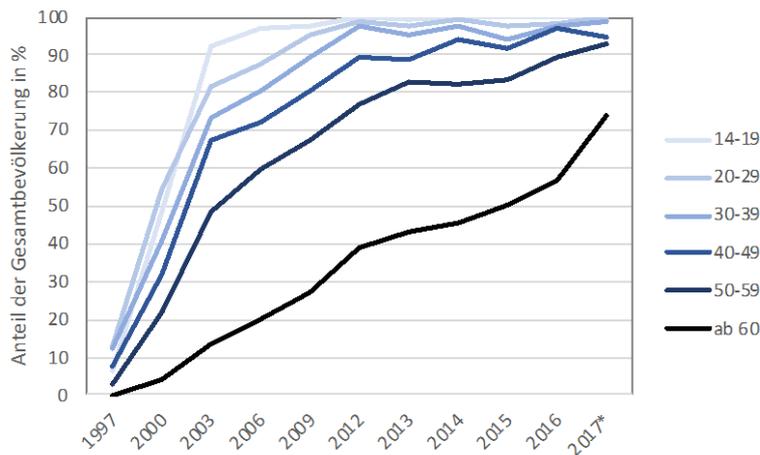


**Abbildung 5: Zeitlicher Verlauf der Internetnutzung nach Geschlecht. Die Daten basieren auf der ARD/ZDF-Onlinestudie 2017 (vgl. Koch & Frees, 2017).**

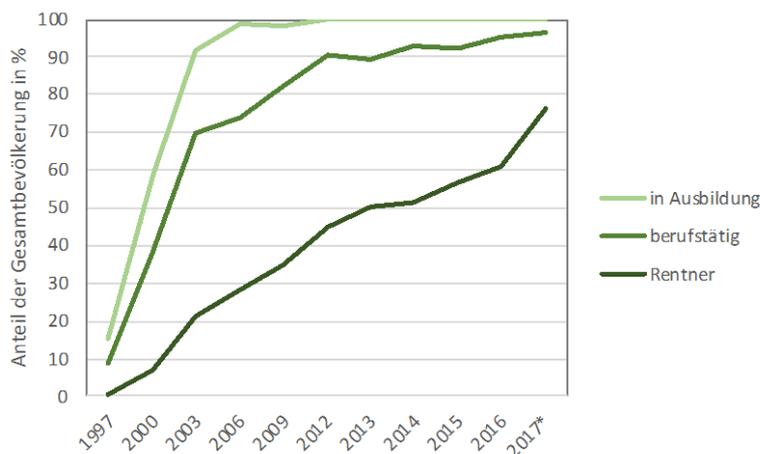
Um die externe Validität zu sichern, können vor und nach der Erhebung Selektionskriterien eingeführt werden, um die Repräsentativität der Stichprobe sicher zu stellen (Treiblmaier, 2011). Dies wird von Online-Verfahren typischerweise durch die einleitenden Fragen und Fragen nach sozio-ökonomischem Status ermöglicht. Ebenfalls kann dies durch die Ansprache von Panelteilnehmern nach den ausgewählten Charakteristika erfolgen. Eine weitere Möglichkeit sind Quoten bzw. Gewichtungen. Sie erlauben eine bessere Vergleichbarkeit und Repräsentativität von Stichproben für eine Population, besonders für die Gesamtbevölkerung eines Landes. Allerdings müssen für eine gute Aussagefähigkeit die Stichproben ausreichend groß sein, um Signifikanzen testen zu können (Duffy et al., 2005).

Diese Normalisierung der Geschlechterverteilung hat eine wichtige Bedeutung für die Repräsentativität von Online-Erhebungen. Welches gilt für die Verteilung von verschiedenen Altersgruppen (vgl. Abbildung 6). Auch hier ist eine Normalisierung der Verteilung zu konstatieren. Media Metrix zeigt, dass die Älteren vermehrt ein Zugang zum Internet haben. Ende 1999 waren 20% der online Population zwischen 45 und 64 Jahre alt mit einem Anstieg von 1,2 % pro Jahr (Media Metrix, 2000). Wenn man dies extrapoliert bedeutet es dies, dass die Quote der Älteren im Internet angestiegen ist. Dynamisch weitergedacht bedeutet es auch, dass die damals zwischen 24 und 44jährigen – nun zwischen 45 und 64 Jahre alt - bereits damals das Internet nutzten und es weiterhin getan haben werden. Nutzungserfahrung und -intensität werden im Normalfall weiter gestiegen sein. Die Internetnutzung wird insofern immer stärker auf alle Altersgruppen verteilt sein. Ein ähnliches Bild zeichnen die Erhebungen des Statistischen Bundesamtes auf, nach welchen im ersten Quartal 2017 die Hälfte der Personen ab 65 Jahren das Internet nutzen und in kommenden Jahren eine noch intensivere Verbreitung

der Internet-Nutzung bei der älteren Bevölkerung und bei Rentnern (vgl. Abbildung 7) zu erwarten ist (Gude, 18.10.2018).



**Abbildung 6: Zeitlicher Verlauf der Internetnutzung nach Altersgruppen. Die Daten basieren auf der ARD/ZDF-Onlinestudie 2017 (vgl. Koch & Frees, 2017).**



**Abbildung 7: Zeitlicher Verlauf der Internetnutzung nach Berufstätigkeit. Die Daten basieren auf der ARD/ZDF-Onlinestudie 2017 (vgl. Koch & Frees, 2017).**

Insgesamt lässt sich feststellen, dass unabhängig von der Erhebungsmethode ein höheres Interesse der befragten Personen positiv auf die Rücklaufquote und Qualität der Antwort wirkt (Chang & Krosnick, 2009). Die Qualität der Antworten wird wiederum auch positiv beeinflusst, wenn keine Verzerrungen der sozialen Erwünschtheit vorliegen sowie Erfahrungen mit Befragungen seitens des Befragten vorhanden sind (Chang & Krosnick, 2009). Darüber hinaus ermöglichen Auswahlfragen wie in Filtern, die Platzierung auf bestimmten interessegeleiteten Onlineforen und -plattformen sowie die gezielte Auswahl der Panelteilnehmer hier deutliche Vorteile

Die von Civey durchgeführten Umfragen weisen explizit darauf hin, wie die Probanden anhand von Parametern der Grundgesamtheit für die Stichproben gewählt bzw. ausgeschlossen werden. Die damit verbundene Transparenz lässt positive Rückschlüsse für die Übertragbarkeit der erhobenen Daten auf die Grundgesamtheit und die damit einhergehende *inhaltliche Repräsentativität* zu.

Zusammenfassend fördert Civey die Repräsentativität ihrer Stichprobenergebnisse durch:

- die randomisierte Auswahl von Teilnehmern aus sehr heterogenen Stichproben durch die Verbreitung auf vielen unterschiedlichen Webseiten,
- Festlegungen spezifischer Quoten (Alter, Parteineigung, Geschlecht, Wohnort und Ort der Teilnahme) und
- die Berücksichtigung unterrepräsentierter Gruppen (insbesondere Frauen, und weniger gebildete) bei der Stichprobenziehung,
- die ausschließliche Berücksichtigung von verifizierten Teilnehmern,
- eine Nachgewichtung der Stichprobengröße zum a-posteriori Abgleich der soziodemografischen a-priori-Verteilungsannahmen der Zielpopulation

Darüber hinaus ist auch anzuführen, dass Repräsentativität oft einseitig und restriktiv dargestellt wird und die Kritik zur Repräsentativität daher häufig überbewertet wird. Lippe and Kladroba (2002) vertritt "...die bei Statistikern im allgemeinen vorherrschende Meinung,

1. dass „Repräsentativität“ ein aus einer im Falle von Stichprobenuntersuchungen falschen Vorstellung heraus entstandenes Konzept ist, und dass

2. Repräsentativität kein Qualitätsmerkmal für eine Untersuchung ist, sondern in einer widersprüchlichen Beziehung zu den theoretisch fundierten statistischen Qualitätskriterien steht.“ (Lippe & Kladroba, 2002, S. 3).

#### h. Soziale Erwünschtheit und Verzerrungen

Datenqualität und Verzerrungen „Biases“ sind seit langem ein zentrales Thema bei empirischen Studien der Praxis und der Wissenschaft. Dabei werden verschiedenste Verzerrungen – oft aus dem Englischen dann eingedeutschten Begriff der „Biases“ diskutiert und man z. B. von Interviewerbias spricht (Bukszar & Connolly, 1988; Doty & Glick, 1998; Ernst, 2003; Meade, Watson, & Kroustalis, 2007; Sax et al., 2003; Scholderer, Grunert, & Brunsø, 2005). Verzerrungen gehen z. B. mit Charakteristika bzw. Einflussgrößen einher, die nicht in die statistische Berechnung aufgenommen wurden, die aber dennoch die Aussagen verändern (Cheung & Chan, 2002; Lu & Thomas, 2008; Lyness & Brumit Kropf, 2007). Verzerrungen können damit z. B. in der Nichtberücksichtigung von Variablen liegen (omitted variable bias). Das könnte z. B. der Einfluss einer guten Laune eines Interviewers sein, die unterschwellige Beeinflussung des Interviewers, die unterschiedlichen Antwortverhalten bei Nutzung unterschiedlicher Erhebungsinstrumente oder veränderte sozio-ökonomische Bedingungen bei verschiedenen Zeitpunkten. Eine wesentliche Verzerrung bei Befragungen liegt mit dem Problem der sozialen Erwünschtheit vor (Ernst, 2003). Diese Verzerrung tritt auf, wenn die befragte Person ungern die eigene, ggf. unpopuläre Meinung wider geben will und sie stattdessen abschwächt oder abändert (Stocké, 2004; Winkler et al., 2006). Verzerrungen der sozialen Erwünschtheit treten besonders auf bei Fragen zu Einstellungen, Werten und Meinungen auf und wenn eine persönliche Interaktionssituation gegeben ist. Die zu befragende Person nähert sich dann eher einer allgemeinen Einschätzung an und verschweigt bzw. verzerrt die eigene. Diese Verzerrung nimmt ebenfalls bei Anonymität des Befragten ab, wie sie typischerweise bei Online-Befragungen vorherrscht (Stanton, 1998).

Standardisierte und automatisierte Bedingungen im Erhebungsrahmen und höhere Erwartungen an Anonymität verringern den Einfluss von Versuchsleitern, steigern die Objektivität der Antworten (Stanton, 1998) und reduzieren Verzerrungen durch soziale erwünschtes Antwortverhalten.

Selbstselektion, Dropout-Raten, sowie Mehrfacheingaben können trotzdem nicht vollkommen ausgeschlossen werden (Treiblmaier, 2011). Aufgrund der Entpersonalisierung und Automatisierung bei Online-Befragungen ist theoretisch von geringeren Verzerrungen der sozialen Erwünschtheit auszugehen. Wenn darüber hinaus Anonymität der Datenverwendung angenommen wird, ist eine Verzerrung der sozialen Erwünschtheit kaum zu erwarten. Studien zeigen darüber hinaus, dass automatisierte Online-Erhebungen besonders niedrige Verzerrungen sozialer Erwünschtheit aufweisen, wenn und weil es sich um geschlossene Fragen handelt (Kiesler & Sproull, 1982; Sproull et al., 2007; Sproull & Kiesler, 1986).

Die Online-Erhebung bei Civey ist insofern kaum dem Problem der sozialen Erwünschtheit ausgesetzt. Darüber hinaus folgt man bei Civey explizit dem Primat der Neutralität. Texte werden kritisch hin auf eine mögliche unterschwellige Beeinflussung geprüft. Fragen und Farben werden sensibel eingesetzt. Die Umfragen durchlaufen vor der Veröffentlichung festgelegte Prozesse und Qualitätskontrollen. Unter anderem berät hier ein Gremium mehrerer Experten, ob mögliche unterschwellige Beeinflussungen in der Befragung vorhanden sind.

#### i. Verbindung mit Online-Panels

Die weltweiten Ausgaben für Online-Forschung sowie panel-basierte Online-Forschung ist in den letzten zehn Jahren stetig gewachsen (Callegaro et al., 2014; Chang & Krosnick, 2009; Duffy et al., 2005). Das primäre Ziel von Online-Panels ist es, das potenzielle Einverständnis der Nutzer zur Rekrutierung in (weiteren) Befragungen einzusetzen (Callegaro et al., 2014 ISO, 2009). Am häufigsten vertreten sind Bevölkerungspanels (Chang & Krosnick, 2009), welche in der Regel sehr groß sind und die Vielfältigkeit der Allgemeinbevölkerung inkl. Menschen in schwer erreichbaren Subpopulationen einzuschließen versuchen (Callegaro et al., 2014). Sie dienen als Rahmen zur Ziehung spezifischer Stichproben, die darauf abzielen, klar definierte Subpopulationen (z. B. wahlberechtigte Bevölkerung) zu befragen. Online-Panels wurden allerdings vor 2013 von Deutschen noch nicht intensiv genutzt (Szolnoki & Hoffmann, 2013).

In ihrer aktuellen Fassung von 2001 weisen die Qualitätssicherungsstandards des Arbeitskreise Deutscher Markt- und Sozialforschung auf, dass: "Repräsentativität [...] bei Online-Befragungen in der Regel nur hinsichtlich der Internetnutzer insgesamt oder spezieller Gruppen von Internetnutzern bzw. der Nutzer bestimmter Websites als Zielgruppe der Untersuchung erreicht werden [kann, wenn] auf der Grundlage einer vorherigen Offline-Auswahl bzw. Offline-Rekrutierung mittels geeigneter Screening Techniken durchgeführt [wird]" (ADM, Mai 2001: 2). Unter Anbetracht der Entwicklungen ist hier eine Aktualisierung zu erwarten.

Einschränkend wird darauf verwiesen, dass "Generalisierbare Aussagen über andere Zielgruppen [...] auf der Grundlage von Online-Befragungen [...] möglich [sind], wenn die Online-Penetration dieser Gruppen nachweislich sehr weit fortgeschritten ist (ähnlich der Ausstattung privater Haushalte mit Telefon) oder wenn die offline für die Befragung ausgewählten Personen mit einem E-Mail- bzw. Internet-Zugang ausgestattet werden, falls sie über keinen verfügen"(ADM, Mai 2001: 2). Einige Institute nutzen zur Ansprache Banner, Suchmaschinen oder Social Media (Riversampling), teilweise ohne sozio-demografische Merkmale zu erheben.

Andere Institute versuchen die Teilnehmer darüber hinaus zu motivieren, sich auf der Webseite des Instituts zu registrieren und dabei demografischen oder persönlichen Angaben zu hinterlegen bzw. an einem Online-Panel teilzunehmen. Diesen Ansatz verfolgt auch Civey. Auch gebräuchlich ist die

Rekrutierung von Teilnehmern für das Panel auf den Seiten von Partnerfirmen (Chang & Krosnick, 2009).

Die Nutzung von Telefon, Post, und E-Mail ist gebräuchlich, um schwer erreichbare demografische Gruppen zu rekrutieren. Typischerweise werden Teilnehmer eher unspezifisch nach der Teilnahme an einer Befragung gebeten und nicht gleich direkt auf ein konkretes Thema hin angesprochen. Je nach Verteilung der gerade ablaufenden Befragungen werden die Teilnehmer an der Befragung dann weiter zu anderen Fragen verlinkt. Die Auswahl folgt typischerweise entlang der Charakteristika des jeweiligen Befragten sowie anhand der Quotierung, welche Gruppen einer bestimmten Befragung gerade am stärksten benötigt werden.

Anreize an die Teilnehmer von Befragungen sind verschieden angelegt. Denkbar sind direkte finanzielle Gegenleistungen an die Teilnehmer (Manfreda et al., 2008). Dies ist aber wenig gebräuchlich. Dafür finden aber indirekte Verfahren oft Anwendung, wie mittels Gewinnspielen oder Punkten. Punkte kann der Befragte sammeln. Sie erlauben dem Paneteilnehmer dann diese später für Konsumgüter, Meilenprogramme, für Spenden oder auch für Geld einzutauschen. Der Erwerb der Punkte kann durch vollständiges Ausfüllen, aber auch bereits bei der Beantwortung von Filterfragen erfolgen.

Unterschiedlich sind die Praktiken wie Institute Stichproben aus dem Panel ziehen. Einige Verfahren sind denen von probabilistischen Stichproben sehr ähnlich (Pasek, 2016). Subgruppen werden dann gemäß ihrer Verteilung an der Gesamtpopulation gezogen. Gewichtungen folgen typischerweise nach Quoten auf den Merkmalskombinationen Alter, Geschlecht und Region (Pasek, 2016). Derartige Gewichtungen werden jedoch nicht von allen Instituten gleichermaßen durchgeführt, z. B. iteratives proportionales angleichen oder mittels regressionsbasierter Verfahren für maßgeschneiderte Befragungen, und die Prognosen unterscheiden sich diesbezüglich auch hinsichtlich ihrer Präzision. Wie die Merkmalskombinationen gewichtet sind und wie sie Befragten kategorisiert werden ist auch wieder unterschiedlich. Zum Beispiel kann ein Institut männlich-weiblich in die Altersgruppen 18-29, 30-49, 50-64, und dann 65 und älter einordnen. Anderes können direkt Alter und Geschlecht in den Quoten verbinden und nach Kategorien suchen wie männlich 18-34, weiblich 18-34, männlich 35-54, weiblich 35-54, männlich 55 und älter, und dann weiblich 55 und älter.

Gewichtungen können aber auch sehr viel ausgereifere Stichprobenauswahlverfahren einsetzen, die elaborierter als Quoten angelegt sind. Zum Beispiel kann jeder Teilnehmer am Panel einen individuellen Score zugeteilt bekommen, der die Wahrscheinlichkeit in dem bestimmten Referenzpanel zu sein ausdrückt. Dieser Score kann direkt das demografische Profil berücksichtigen. Quoten berechnen sich dann aufgrund von z. B. Quantilen des Wahrscheinlichkeitscores anstelle der direkten Charakteristika des Befragten. Des Weiteren können Unterstichproben von einer größeren Wahrscheinlichkeitsstichprobe gezogen werden und dann Vergleiche zwischen diesen und den eigenen Stichproben gezogen werden. Sobald der Fit zwischen den Stichproben ausreichend ist, wird dann mit der eigenen Stichprobe weitergeschätzt. Insofern wird eine größere Anzahl von Charakteristika der Personen in den Selektionsprozess einbezogen als es Quoten erlauben. Bei sehr schlecht erreichbaren Personengruppen könnten auch Teilnehmer von verschiedenen Instituten angesprochen werden.

Einige Institute erlauben zudem die Option „Riversample“-Fälle zusammen mit der Stichprobe des Panels zur Bildung einer finalen Stichprobe. Der Begriff Riversampling wird verwendet, wenn das jeweilige Institut Internetnutzer einlädt die Befragung auf einer in eine Webseite integrierte Anzeigen zu beantworten ohne dass sie an dem Panel teilnehmen. Im engeren Sinne eines Riversampling-Verfahrens werden keine weiteren Merkmale als die für die eigentliche Befragung erforderlichen, damit auch keine sozio-demografischen Merkmale, erhoben. Das Riversampling wird so für eine Anmischung

von anderen Erhebungen in Punkto Stichprobengröße verwendet. Wenn allerdings wie im engeren Sinne eines Riversampling keine sozio-demografischen Merkmale erhoben werden, liegt eine Verwaschung der Repräsentativität vor, wenn Aussagen über eine Gesamtbevölkerung getroffen werden sollen.

Civey bezieht repräsentative Umfrageergebnisse ausschließlich auf die im Online-Panel registrierten und verifizierten Teilnehmer. Die Panel-Teilnehmer werden mit dem Riversampling-Ansatz und einem speziellen Umfrage-Tool (Widget) rekrutiert.

Auch innerhalb des Panels verwendet Civey das Riversampling. Zusätzlich wird dabei immer auch versucht, weitere Charakteristika der Befragten mit zu erheben, um Quoten und Gewichtungen anlegen zu können.

Das Umfragetool, welches Civey zur Rekrutierung der Panelteilnehmer nutzt ist nach eigenen Angaben auf 25.000 Webseiten (URLs) eingebettet. Gezählt werden dabei nur aktive Widgets, also Widgets in welche mindestens einmal im Monat ein Nutzer ein Item beantwortet hat. Webseiten können jederzeit wieder aktiv werden, sobald ein Nutzer teilnimmt. Nutzer, die teilnehmen wollen, müssen sich zunächst registrieren und durchlaufen einen Verifizierungsprozess. Erst nachdem dieser abgeschlossen ist, können sie Teilnehmer am Panel werden. Die Panelteilnehmer besuchen zu einem späteren Zeitpunkt dann eine der Webseiten auf der das Tool von Civey eingebettet ist erneut bzw. eine neue Webseite auf der das Tool mittlerweile neu eingebettet wurde. Dann entscheidet der Riversampling Algorithmus, welche Umfrage ihnen angezeigt wird. Der Nutzer wird über ein Cookie bzw. E-Mail-Login wiedererkannt, so dass Identifikation erlaubt wird und Mehrfachantworten ausgeschlossen werden. Hieran zeigt sich ein sehr durchdachtes Vorgehen, das probabilistische und nicht-probabilistische Ansätze sehr sinnvoll kombiniert. Die Erfassung von weiteren Merkmalen erlaubt Quoten und Gewichtungen und verbindet diese mit einer Art Zufallsauswahl.

#### j. Probabilistische und nicht-probabilistische Stichproben

Die hohen Kosten, niedrige Rückläufe und die Nicht-Erreichbarkeit von bestimmten Personengruppen hat dazu geführt, dass immer mehr Verfahren auf nicht-probabilistischen Stichproben angewendet werden (Skowronek & Duerr, 2009). Diese werden häufig mit Panels (siehe hierzu die anderen Abschnitte dieses Gutachtens) kombiniert. Der Mangel eines umfassenden Stichprobenrahmens für das Internet und seine Nutzer hat zu der anhaltenden Ansicht geführt, dass es unmöglich sein, eine nationale Stichprobe zu ziehen, in der Jeder die gleiche Chance hat ausgewählt zu werden (Kennedy et al., May 2016).

Während die AAPOR noch vor zehn Jahren Zufalls- oder Wahrscheinlichkeitsstichproben zum Standard wissenschaftlicher Anerkennung voraussetzte, wurde diese Sichtweise 2013 infolge einer Task Force (Baker et al., May 2013) in der jüngeren Vergangenheit an die aktuellen Entwicklungen angepasst (ESOMAR, 2015). Dies bedeutet, dass Erhebungen mit nicht-probabilistischen Stichproben zunehmen und ihre Akzeptanz steigt. Neben der wachsenden Verfügbarkeit von Online-Zugängen sowie der steigenden Präsenz von Online-Angeboten für unterschiedliche Bevölkerungsschichten wurden auch alternative Methoden der Datenanalyse weiterentwickelt. Die zunehmende Akzeptanz von nicht-probabilistischen Stichproben zählt möglicherweise zu den zukunftsorientiertesten und gleichzeitig umstrittensten Veränderungen der letzten fünf Jahre (Kennedy et al., May 2016).

In den USA werden nicht-probabilistische Stichprobenverfahren bereits intensiv eingesetzt, gerade weil bestimmte Bevölkerungsgruppen wie Schwarze und Hispanos sonst nur bedingt erfasst werden können (Kennedy et al., May 2016). Die Güte nicht-probabilistischer Stichproben ist bisher sehr

unterschiedlich. Kennedy et al. (May 2016) führte daher einen Vergleich von verschiedenen Verfahren durch.

In "nonprobability-based samples, where respondents opt-in or volunteer to take the survey [...] the probabilities of selection are unknown [and] modeling, weighting or other adjustments are necessary to project from the sample to the larger population." (AAPOR, 2014).

Nicht-probabilistische Stichproben verwenden unterschiedliche Verfahren um Teilnehmer – auch wiederkehrende Panelteilnehmer - an der Befragung zu rekrutieren. Dabei gibt es geradezu professionalisierte Teilnehmer (Hillygus, Jackson, & Young, 2014). Die Entwicklung der Technologien und statistischen Techniken zur Verbesserung ihrer Repräsentativität werden nicht-probabilistische Stichproben wie bereits angeführt in jüngerer Vergangenheit nicht mehr kategorisch ausgeschlossen (Ansolabehere & Rivers, 2013; Ansolabehere & Schaffner, 2014). Obwohl diese Stichproben überwiegend nicht über die erforderlichen statistischen Eigenschaften verfügen, um die beabsichtigte Zielpopulation darzustellen (ESOMAR, 2015), werden sie unter Berücksichtigung demografischer Quoten bei der Auswahl (in erster Linie Alter und Geschlecht, Ellis & Krosnick, 1999) sowie nach zusätzlichen Anpassungen zur Verbesserung der Genauigkeit akzeptiert, wenn sie die ISO-26362-Anforderungen erfüllen und den Leitlinien der AAPOR (AAPOR, April 2016) folgen (ESOMAR, 2015).

"Die Auswahl der Teilnehmer [...] muss entweder mittels eines Zufallsverfahren erfolgen [...] oder anderen mit dem Forschungsgegenstand nachweislich korrelierenden Quoten, deren Verteilung in der Grundgesamtheit bekannt ist." (Rat der Deutschen Markt- und Sozialforschung e.V., 2007: 2)

Nicht-probabilistische Stichproben erfordern hohe Reflektion und Transparenz hinsichtlich der Abläufe zur Rekrutierung für das Panel, der Selektionsprozesse für die Teilnahme an spezifischen Umfragen (Problem der Selbstselektion) sowie der notwendigen und vorgenommenen nachträglichen Datenmodifizierungen (AAPOR, 2014).

Die Studie von Kennedy et al. (May 2016) vergleicht verschiedene Online-Erhebungsverfahren mit nicht-probabilistischem Ansatz im Bereich der politischen Meinungsforschung in den USA. Dabei gelangen sie zunächst zu dem Ergebnis, dass sich die verschiedenen kommerziellen Institute stark in der Rekrutierung von Teilnehmern, der Stichprobenauswahl und der Felderhebung unterscheiden. Diese Unterschiede manifestieren sich auf verschiedene Weise sowohl in der Datenqualität als auch im Einsatz von Gewichtungsverfahren. Genauere Ergebnisse werden demnach erzielt, wenn umfangreiche Anpassungen bei der Stichprobenauswahl und -gewichtung sowie längere Erhebungszeiten vorliegen. Die vorausgehende und/oder erhebungsbegleitende Abstimmung von Populationsmarkern könnte dazu beigetragen haben, dass die Genauigkeit der Ergebnisse aus einer nicht-probabilistischen Stichprobe alle anderen Erhebungen (inklusive einer probabilistischen Stichprobe) übertraf. Diese Erhebung weist die umfangreichsten Anpassungen bei der Stichprobenauswahl auf und schließt Gewichtungen mit Variablen ein, die im allgemeinen als Ergebnisvariablen angesehen werden (politische Interessen, Ideologie und Internetnutzung). Dies legt nahe, dass die Güte der Ergebnisse davon abhängt, wie gut die Stichprobe vor und während der Felderhebung an vielfältige Populationsmarker angepasst wird.

Ferner wird berichtet, dass tendenziell eher Personen an Befragungen teilnehmen, die generell als Freiwillige agieren. Dies gilt allerdings für alle Befragungsinstrumente – ob schriftlich, telefonisch oder elektronisch. So erzeugt die freiwillige Hilfsbereitschaft der Befragten über die letzten 12 Monate hinweg die größten Verzerrungen. Zudem wird festgestellt, dass Personen mit höherem politischen und gesellschaftlichen Interesse in allen Stichproben stärker vertreten sind, als es anzunehmen war. Auch

dies gilt allerdings nicht nur als Limitation für nicht-probabilistische Stichproben, sondern für alle Offline- und Online-Befragungsinstrumente.

Bezogen auf die Lebensstile und Hobbies der Befragten erbrachten die nicht-probabilistischen Stichproben ähnliche Ergebnisse. Befragt wurden verschiedene Interessen und Aktivitäten (z. B. Bibel lesen, Gartenarbeit, Work-Out, News über Prominente und Klatsch etc.). Die Ergebnisse sind über die verschiedenen Erhebungen weitgehend konsistent. Kennedy et al. (May 2016) berichtet ferner, dass die Schätzergebnisse von online erhobenen nicht-probabilistischen Stichproben sehr genau sind (z. B. ob der Befragte einen Führerschein hat oder wie lange er dort lebt) und dass Verzerrungen eben nicht in eine Richtung auftreten.

Tendenziell – nicht konsistent – weisen nicht-probabilistischen Stichproben darauf hin, dass Befragte mit bestimmten sozio-ökonomischem Status überrepräsentiert sind. In den USA nehmen tendenziell mehr Personen an Befragungen teil, die kinderlos sind, alleine leben und Empfänger von Sozialleistungen sind bzw. ein niedriges Einkommen haben (Kennedy et al., May 2016). Interessanterweise widerspricht dies anderen Studien, die tendenziell mehr Befragte von Online-Befragungen aus höheren sozio-ökonomischen Schichten berichten. Dies deutet darauf hin, dass Online-Befragungen sowohl Befragte aus niedrigen oder höheren sozio-ökonomischen Schichten integrieren können, sie diese aber auch über- oder unterrepräsentieren können.

Zusammenfassend kann dargestellt werden, dass Online-Befragungen versuchen das Potential zu nutzen, Menschen aus verschiedensten Schichten und Bevölkerungsgruppen mit probabilistischen und nicht-probabilistischen Methoden zu erfassen. Dies erhärtet wiederum die Aussage, dass Online-Befragungen mit nicht-probabilistischen Stichproben das Potential für Repräsentativität und Aussagekraft haben, wenn die Vorgehensweise durchdacht, zielgerichtet und mit sinnvollen Gewichtungen verbunden ist. Weiterhin lässt sich konstatieren, dass die Unterschiede zwischen probabilistischen und nicht-probabilistischen Stichproben konzeptionell eindeutig sein mögen, aber die praktische Realität stärker durch das Problem der Nichtbeantwortung geprägt ist. Wenn bspw. 90 Prozent der Personen innerhalb einer probabilistischen Stichprobe nicht antworten, ist zwar die Auswahlwahrscheinlichkeit weiterhin bekannt, aber dennoch nicht die individuellen Antwortwahrscheinlichkeit (Dülmer, 2007; Kennedy et al., May 2016). Daher ist es in Bevölkerungsumfragen äußerst schwierig, die Wahrscheinlichkeit der Antwort mit einem hohen Maß an Genauigkeit zu schätzen und Forscher sind darauf angewiesen, der Unsicherheit einer potenziell möglichen Unrepräsentativität durch methodische Anpassungen und Gewichtungen entgegen zu wirken. Wenn der Forscher die Wahrscheinlichkeiten nicht kennt, bedarf es einer Gewichtung, um die Fehler bei der Repräsentativität zu korrigieren (Mercer, Kreuter, Keeter, & Stuart, 2017).

So finden sich sowohl in probabilistischen wie in nicht-probabilistischen Stichproben zur politischen Meinungsforschung z. B. vor allem Verzerrungen durch politisch interessierte Personen. Gründe liegen darin, dass das Thema Politik und Wahlen bei diesen Personen stärker präsent ist und sie daher auch eher an Befragungen zu derartigen Themen teilnehmen. Allerdings mag auch hier eine Verzerrung durch soziale Erwünschtheit vorliegen, in deren Folge Personen angeben stärker politisch engagiert zu sein, als sie es sind. Ferner sind junge und weniger formal gebildete Menschen in solchen Stichproben unterrepräsentiert, weil sie sich weniger für Politik interessieren. Solche Verzerrungen können durch Gewichtung korrigiert, jedoch nicht gänzlich aufgehoben werden.

Typischerweise verwenden Institute eine Reihe von Kriterien zur Qualitätssicherung (Skowronek & Duerr, 2009). Kriterien und Ablaufschritte können einfach oder komplex angelegt sein und werden in den eigentlichen Befragungsteil und/oder im Bereich der Erhebung der personenbezogenen (objektiven) Kriterien integriert. Ferner wird gemessen, wie schnell der Befragte antwortet und ob die

Antworten zu homogen erfolgt (z. B. durchgehende Bevorzugung von Mittel- oder Randpositionen). Übergreifende Evaluationen können sich auf die Häufigkeit der Beantwortung seitens eines Befragten beziehen und ob die Befragung auf Eigeninitiative oder auf Einladung erfolgt ist. Einige Institute validieren die IP Adresse oder nutzen einen digitalen Fingerabdruck um zu verhindern, dass eine Person mehrere Konten auf einem Panel hat. Dies umgeht dann das Risiko von Mehrfachberücksichtigung.

Die mögliche eingeschränkte Zufallsauswahl in nicht-probabilistischen Stichproben resultiert primär daraus, dass die Grundgesamtheit in der Stichprobe mehr oder weniger gut repräsentiert wird. Obwohl selbst potenzielle strukturelle Unterschiede zwischen Zielpopulation und Stichprobe keinen Rückschluss auf mangelnde Repräsentativität zulassen, solange es sich dabei nicht um Faktoren handelt, die eine Verhaltens- oder Einstellungsrelevanz mit der Fragestellung aufweisen, kann die inhaltlichen Übertragbarkeit von Stichproben durch die präzise Definition einer Zielpopulation, die Berücksichtigung spezifischer Quoten der Zielpopulation bei der Stichprobenselektion, den Nachweis struktureller Gleichheit (hinsichtlich der verhaltens- und einstellungsrelevanten Faktoren) sowie dem Einsatz spezifischer Gewichtungen verbessert werden. Weitere Einschränkungen der Repräsentativität, die durch inhaltliche und methodische Ursachen hervorgerufen werden, sind aber kein spezifisches Problem der Online-Forschung.

Übergreifend beurteilt scheinen die Ergebnisse von Online-Befragungen mit nicht-probabilistischen Stichproben aussagekräftiger, wenn diese Stichproben ausgefeilte Stichprobenziehungsroutinen und Gewichtungsschemata in Anlehnung an probabilistische Verfahren verfolgen. Bereits jetzt haben sich viele Institute die Online-Befragungen einsetzen mit der Verbesserung der Repräsentativität von Online erhobenen nicht-probabilistischen Stichproben beschäftigt. Einige Studien meinen schon zum jetzigen Forschungsstand, dass teure probabilistische Stichproben gar nicht mehr erforderlich sind, um zuverlässige Aussagen über die Population zu treffen (Ansolabehere & Rivers, 2013; Ansolabehere & Schaffner, 2014; Kennedy et al., May 2016). Zudem sind gute Ergebnisse zu erwarten, wenn mehr Erfahrung seitens des Instituts vorliegt und wenn die Befragungen über längere Zeiträume durchgeführt werden (Kennedy et al., May 2016).

Civey führt seit zweieinhalb Jahren Online-Befragungen durch. Bisher sind fast 4.000 Einzel-Umfragen durchgeführt worden. Die meisten davon über längere Zeiträume (mehrere Monate bis zwischenzeitlich zwei Jahre). Auf jeder Umfrage werden täglich neue Stimmen erhoben, mehrfach pro Tag Stichproben gezogen und das Gewichtungsverfahren angewandt. Je nach Notwendigkeit werden pro Tag bis zu 500.000 Neuberechnungen der Stichproben und Gewichtungen vorgenommen. Dies belegt hohe Expertise und Erfolg mit den eingesetzten Verfahren. Darüber hinaus ist der bereits mehrfach angeführte durchdachte Ablauf bei Erhebungen als Vorteil zu anzuführen.

Nach eigenen Angaben nutzt Civey nicht-probabilistische Stichproben. Allerdings werden verschiedene Verfahren (Quoten, Gewichtungen, bayessche Schätzung) in sinnvollen Schritten kombiniert (siehe Abbildung 1). Damit nutzt Civey die Vorteile der verschiedenen Verfahren und analysiert die Panelteilnehmer quoten- und gewichtungsorientiert. Damit lassen sich positive Wirkungen für die Aussagekraft der Ergebnisse und Steigerungen der Repräsentativität erzielen, selbst wenn diese aus verschiedenen Blickwinkeln heraus beurteilt wird.

In der Kommunikation mit der Öffentlichkeit wird der nicht-probabilistische Ansatz der Stichproben dargestellt und diskutiert. Dies ist grundlegend zu begrüßen. Unvertrautheit und Unverständnis der damit verbundenen Analyseschritte können aber unter Umständen zu einer (möglicherweise temporären) negativeren Beurteilung durch die Öffentlichkeit führen.

## 6. Gesamtbeurteilung

Die aktuellen technologischen und gesellschaftlichen Entwicklungen haben zu einer zunehmenden Nutzung von Online-Befragungen und ihrer methodischen Weiterentwicklung sowie weiteren Ausfächerung geführt. Dabei stellt sich die Frage nach der Aussagekraft und auch der Repräsentativität von Online-Befragungen im Generellen wie im Speziellen. Die Breite der Anwendungsfelder und Methoden von Online-Befragungen ist groß. Dennoch lassen sich Kriterien aus der Wissenschaft und guten Praxis für die Beurteilung von Online-Befragungen anlegen. Allerdings kann es keine abschließende und komplette Beurteilung von Verfahren von Online-Befragungen geben, weil eben die Methodenvielfalt und ihre Einsatzfelder groß sind.

Dieses Gutachten beschäftigte sich mit der typischen Vorgehensweise bei Online-Befragungen seitens des Unternehmens Civey. Es geht in dem Gutachten letztlich um die Evaluation von Aussagekraft und Repräsentativität von Online-Befragungen und speziell denen von Civey in einem umfassenderen Verständnis. Die Beurteilung der Systematik bei Civey erfolgt hier vor dem Hintergrund bestehender wissenschaftlicher und praxisgeprägter Studien, Vergleiche und Vorgehensweisen bei Online-Befragungen. Daher finden allgemeinere Beurteilungen von Online-Befragungen und ihren Komponenten Eingang in die hier vorliegende Beurteilung wie eben die spezifische Systematik bei Civey. Um die Aussagekraft und Güte der Online-Befragungen bei Civey aus verschiedenen Blickwinkeln und wissenschaftlich analytisch beurteilen zu können, wurden verschiedene Kriterien angelegt und Aspekte berücksichtigt.

Die umfangreichen Abwägungen und Analysen im Hauptteil dieses Gutachtens lassen insgesamt den Schluss zu, dass Civey eine sehr ausgereifte und durchdachte Systematik bei Online-Befragungen für die Markt-, Meinungs- und Wahlforschung verwendet, die eine hohe Aussagegüte erwarten lässt. In den letzten Jahren wurden seitens Civey bereits viele Befragungen durchgeführt, die Erfahrungswissen gebracht haben, dass für Anpassungen und Verbesserungen eingesetzt werden konnte. Die Anzahl der durchgeführten Erhebungen und die hohe Anzahl der Panelteilnehmer unterstreicht dies. Insgesamt betonen (Vergleichs-)Studien, dass Erfahrung und durchdachtes Vorgehen bei Online-Befragungen stark positiv auf deren Qualität wirken. Dies ist bei Civey gegeben und stärkt generell die Aussagekraft und Güte der Befragungen.

Online-Befragungen können Vorteile im Hinblick auf ihre Aktualität, Schnelligkeit und relativer Kostengünstigkeit gegenüber anderen Verfahren ausspielen, was zudem mit der einfacheren Möglichkeit von aktuellen und großen Stichproben einhergeht, allerdings auch gewisse Gefahren der zu unüberlegten Durchführung birgt. Dies erfordert generell ein hohes Qualitätsbewusstsein bei Online-Befragungen und dem Festhalten an Qualitätsstandards. Die geringeren Kosten von Online-Befragungen gegenüber brieflichen und telefonischen Befragungen ermöglichen Instituten wie Civey umfangreiche und mehrfache Ansprache von Personen. Im Resultat führt dies zu einer Steigerung der Kontakte und der Rückläufer. Die recht hohen Resonanzraten sowie vielen Befragungen seitens Civey belegen dies. Durch die meist einfachere Rekrutierung von Teilnehmern erzielen Online-Befragungen große Stichproben, was sich positiv auf die Reliabilität und statistische Validität auswirkt (Dülmer, 2007; Treiblmaier, 2011; Walsh et al., 1992). Civey kann diese Vorteile nutzen und zeigt dies durch große Stichproben bei aktuellen Fragen sowie in einem großen Online Panel. Aktualität ist zentral für die Wahl- und Meinungsforschung. Hier kann Civey dann die Vorteile im besonderen Maße ausspielen.

Online-Befragungen haben zudem vielfältige Möglichkeiten attraktiver und zielgruppenorientierter audio-visuellen Gestaltung der Befragung und der Lenkung der Befragten durch technologische und persönliche (Hotline) Unterstützung. Zudem tritt die Möglichkeit einer sehr gezielten Zielgruppenansprache durch Implementierung der Befragungen in Foren, Plattformen und

Communities. Eine breite Streuung der Ansprache in verschiedenen Foren, Plattformen und Communities ermöglicht es möglichst viele Zielgruppen und Bevölkerungsgruppen anzusprechen. Die Platzierung in Foren, Plattformen und Communities erlaubt mannigfaltige Zielgruppen, die sich auch durch andere Charakteristika auszeichnen (Alter, Geschlecht, Bildungsstand usw.) anzusprechen. Darüber hinaus gewinnt die Nutzung von Plattformen immer mehr Popularität, so dass von einer steigenden Repräsentativität im Vergleich mit anderen Erhebungsmethoden auszugehen ist. Überlegte Quoten und intelligente Platzierung der Befragung in Foren, Plattformen und Communities für noch unterrepräsentierte Bevölkerungsgruppen kann eine weitere Aussagefähigkeit und Generalisierbarkeit für Populationen bis hin zu nationalen Populationen ermöglichen. Civey spielt umfassend die Klaviatur der Ansprache-Möglichkeiten aus. Als besonders attraktiv erweist sich dabei die Trend-Online-Landkarte innerhalb der Wahlforschung. Durch den Einsatz interaktiver Elemente kann Civey die Aufmerksamkeit angesprochener Zielgruppen steigern und die Teilnahmebereitschaft für konkrete Fragestellungen erhöhen. Dies wirkt sich positiv auf Rücklaufquoten und Datenqualität aus. Die beschleunigte Durchführung von Befragungen, die einfache Adaption und Auswertung von Befragungen sowie die Vermeidung unvollständiger oder fehlerhafter Eingaben durch den Einsatz von Plausibilitätsprüfungen während der Dateneingabe erlauben es Civey aussagekräftige Daten mit hoher Relevanz und Aktualität aus Online-Erhebungen zu beziehen. Im Vergleich zu Offline-Erhebungen lassen durchdachte und qualitätsgesicherte Online-Erhebungen so eine höhere Validität durch heterogene Stichproben, der gesteigerten zeitlichen und örtlichen Flexibilität sowie Attraktivität (Interaktivität und Multimedia) erwarten.

Insgesamt kann die Darstellung und die Leitung der Befragung bei Online-Erhebungen mit integrierter Filterung, Navigation und Plausibilitätsprüfung effizient erfolgen. Die einhergehende Reduzierung der kognitiven Belastungen von Befragten kann die Rücklaufquoten und die Antwortqualität steigern, möglicherweise aber auch die Verarbeitungstiefe von Entscheidungen reduzieren. Civey setzt ein breites Repertoire an Gestaltungsmaßnahmen, Qualitätskriterien und Systematiken ein, um eine hohe Aussagegüte und Repräsentativität der Ergebnisse zu erzielen. Neben der optisch ansprechenden Gestaltung von App und Widget sind insbesondere die Visualisierung der Teilnehmer auf einer Landkarte und in Live-Streams hervorzuheben, die wesentlichen soziodemografischen Merkmale der Teilnehmer transparent darstellen. Dazu kommt ein aktiver Support, der an sieben Tagen die Woche Teilnehmerfragen per E-Mail und in den Kommentarfunktionen von Sozialen Netzwerken bzw. Chats beantwortet. Außerdem gibt es eine fortlaufend aktualisierte FAQ-Sektion auf der Webseite sowie eine übersichtliche Darstellung des Verfahrens in Text und mittels erklärender Videos. Dies erlaubt umfangreiche Klärung, Support und Motivation der Befragten, die sich auf Rücklaufquoten und Datenqualität positiv auswirken sollte. Die Entwicklung neuer Befragungstechnologien, die Weiterentwicklung bestehender Methoden sowie die komplementäre Kombination ihrer spezifischen Stärken und Schwächen weist auf Qualität, Aussagekraft und Repräsentativität von Online-Befragungen wie Civey sie vornimmt, hin.

Wenn die zielgruppenorientierte Integration von Befragungen in spezifische Webseiten zudem mit sozio-demografischen Merkmalen und Wahrscheinlichkeit verbunden werden, lassen sich außerdem Aussagen über Bevölkerungsgruppen und in Richtung besserer Repräsentativität treffen. Civey nutzt diese Möglichkeiten. Verdeutlicht und gestärkt wird dies, weil Civey es bereits geschafft hat, ein großes Panel aufzubauen. Dieses Panel erlaubt die Ansprache und Antworten von Befragten nach Quoten, so dass die Repräsentativität über Bevölkerungsgruppen hinweg gestärkt werden kann. Wenn wie bei Civey weitere Merkmale der Personen bei der Rekrutierung der Panelmitglieder erhoben wurden, können weitere Quoten und Gewichtungen angelegt werden, um Panelmitglieder gezielt und möglichst im Sinne einer Repräsentativität für weitere Online-Befragungen anzusprechen. Dies erlaubt eben dann die Vorteile von Quoten und Gewichtungen vor und nach Befragungen für die Verbesserung der

Aussagen für Zielgruppen und im Sinne einer Repräsentativität für Populationen zu nutzen. Vorteilhaft wirkt dabei eben die gewisse Verbindung von probabilistischen und nicht probabilistischen Verfahren bei Civey, bei dem auch Ideen und Vorteile Riversampling mit der Erfassung weiterer Merkmale auf dem Panel und der Nutzung von Gewichtungen sowie Quoten verbunden werden. Die Verbindung erlaubt die Nachteile bei einem reinen Riversampling zu reduzieren bzw. überkompensieren.

Offensichtlich erlauben Online-Befragungen wie seitens Civey einen weiter wachsenden Zugang zu verschiedenen demografischen Schichten der Bevölkerung, die zudem eine wachsende Übereinstimmung mit verschiedenen Grundgesamtheit der Bevölkerung aufweisen und über andere Kanäle kaum erreicht werden können. Interessant für gezielte Erhebungen sind dabei die vielen oft spezialisierten Communities im Internet, die einen direkten Zugang zu Themen und Zielgruppen bieten (Garton et al., 1997; Garton & Wellman, 1995). Dadurch erreichen Online-Befragungen wie sie auch seitens Civey durchgeführt werden Bevölkerungsgruppen mit einer hohen Varianz in verschiedenen Merkmalen (z. B. Alter, Ort, sozio-ökonomischer Status, kultureller Hintergrund, physischen Einschränkungen) aufweisen und sonst möglicherweise unterrepräsentiert wären (Gosling et al., 2004). Da die Teilnehmer von Online-Befragungen aber nicht vollständige Bevölkerungen von Nationen repräsentieren, liegt ein besonders Interesse in der Generalisierbarkeit bzw. Repräsentativität von Zusammenhängen innerhalb der Stichprobe auf die Population bzw. Grundgesamtheit. Als gängige Vorstellung von Repräsentativität lässt sich auffassen, dass die Teilpopulation so auszuwählen ist, dass aus ihren Ergebnissen so exakt wie und sicher wie möglich auf die Relationen und Wirkungen in der Zielpopulation geschlossen werden kann. Das gilt meist dann, wenn die Teilpopulation ein relativ gutes, wenn auch verkleinertes Abbild der Gesamtpopulation darstellt (Lippe & Kladroba, 2002). Allerdings wird oft Repräsentativität überbetont und der Stichprobenfehler unterschätzt.

Trotz der weitreichenden Aktivitäten zur Verbesserung der Repräsentativität der Aussagen kann Civey zurzeit aber noch nicht ganz dem Vorwurf entgehen, dass bei ihren Online-Befragungen Internet-Nicht Nutzer unberücksichtigt bleiben und damit gewisse Einschränkungen der Repräsentativität vorkommen. Bisher wurden Telefon- und brieflichen Befragungen oft eine hohe Repräsentativität zugesprochen und bis vor ca. 2013 meist die Repräsentativität von Online-Befragungen kritisch gesehen. Begründet wurde dies vor allem damit, dass fast jeder Deutsche einen Telefonanschluss oder eine Adresse hat, aber nicht jeder das Internet nutzt bzw. nicht intensiv nutzt. Der Vorwurf von Limitationen im Hinblick auf die Repräsentativität von Online-Befragungen lässt sich solange nicht entkräften wie nicht zusätzliche Offline-Befragungen (Brief/Telefon) eingesetzt werden, die Entwicklungen in der Zukunft eine komplette Internetnutzung in der Bevölkerung bringen oder sehr durchdacht Personen aus verschiedenen Bevölkerungsgruppen integriert werden. Dies impliziert aber nicht, dass ohne diese zusätzliche Telefonie nicht Repräsentativität erreicht werden kann. Es zeigt sich zudem, dass die Nutzung des Internets und auch der Vielnutzer immer stärker ansteigt und sich dies durch alle Bevölkerungsteile durchzieht. Gleichzeitig nimmt die Anzahl der Festnetzanschlüsse ab und der Anteil der Internetkommunikation, -information, und den Geschäftsbeziehungen über das Internet (inkl. Bestellungen) zu. Angesichts der stärkeren Nutzung des Internets als Informations- und Kommunikationsmedium und der geringeren Nutzung von Festnetzanschlüssen in Deutschland ist die generelle Annahme, dass Telefon- und brieflichen Befragungen repräsentativ sind und Online-Befragungen eher weniger repräsentativ immer weniger haltbar. Dennoch bedeutet dies keineswegs, dass Online-Befragungen generell repräsentativ sind. Umgekehrt bedeutet es auch nicht, dass telefonische und briefliche Befragungen immer repräsentativ sind oder besser im Hinblick auf die Repräsentativität sind als Online-Befragungen. Aber in der Tat ist die Gefahr, dass Online-Befragungen zu schnell, kostengünstig und zu wenig systematisch angelegt werden und damit eben nicht repräsentativ sind, sehr hoch. Dies unterstreicht dann wieder die Anforderung, dass Online-Befragungen sehr durchdacht angelegt sein müssen und dabei Zielgruppen bzw. Bevölkerungsgruppen

sinnvoll integriert. Aufgrund der Systematik, Qualitätssicherung, Einsatz eines Gremiums bei Civey wird bei Civey dieses Problem nicht vorliegen. Die gezielte Nutzung von Websites für weniger Internetaffine Gruppen wie Senioren oder teilweise auch Frauen können wiederum können den Vorwurf enthärten, dass Online-Befragungen weniger Aussagekraft für diese Bevölkerungsgruppen haben. Beispielsweise zeigen die Zahlen von Civey, dass jüngere Männer nicht überrepräsentiert sind. Begründen lässt sich dies eben mit der guten Auswahl von Websites auf denen die Befragungen angesiedelt sind so wie in einer durchdachten Ansprache der Panelteilnehmer. Durch die gezielte Implementierung der Befragungen in Foren, Plattformen und den Sites von Unternehmen ermöglicht eine gezielte Ansprache von entsprechenden Bevölkerungsgruppen. Das Unternehmen Civey nutzt das Potential der steigenden Online-Penetration, in dem das Umfrage-Tool (Widget) auf verschiedensten Webseiten gezielt und durchdacht platziert wird und zur Teilnahme an Abstimmungen einlädt. Dadurch lassen sich sonst schwierig erreichbare Bevölkerungsgruppen ansprechen und an die Verteilungen in der Grundgesamtheit angleichen. Dies gilt auch in demografischer Hinsicht. Obwohl 100 Prozent der unter 30-Jährigen Online-Dienste nutzen (Koch & Frees, 2017), sind diese Altersklassen im aktiven Panel von Civey nicht überrepräsentiert. Dies könnte mit einem geringeren Interesse an gesellschaftlichen oder politischen Fragestellungen oder wachsenden Bedenken gegenüber kommerziellen Datensammlungen zusammenhängen. Zu betonen ist, dass die prozentuale Häufigkeit der Altersklassen, des Familienstandes und der Herkunft (alte Bundesländer versus neue Bundesländer) keine bedeutsamen Abweichungen der registrierten Teilnehmer von den im Mikrozensus des Statistischen Bundesamtes berichteten Bevölkerungsdaten zeigen. Allerdings kann dieses nicht für die Werte der Geschlechterverteilung, des Bildungsabschlusses und der Berufstätigkeit bescheinigt werden.

Um die Repräsentativität der Stichprobe sicher zu stellen, können vor und nach Erhebung Selektionskriterien eingeführt werden (Treiblmaier, 2011). Dies wird von Online-Verfahren typischerweise durch die einleitenden Fragen und Fragen nach sozio-ökonomischem Status ermöglicht. Ebenfalls kann dies durch die Ansprache von Panelteilnehmern nach den ausgewählten Charakteristika erfolgen. Eine weitere Möglichkeit sind Gewichtungen. Die CIVEY-Umfragen erklären, wie die Probanden anhand von Parametern der Grundgesamtheit für die Stichproben gewählt bzw. ausgeschlossen werden. Diese Transparenz steigert die inhaltliche Repräsentativität. Repräsentativität wird zudem gefördert durch die randomisierte Auswahl von Teilnehmern aus sehr heterogenen Stichproben durch die Verbreitung auf vielen unterschiedlichen Webseiten, Anwendung spezifischer Quoten (Alter, Parteineigung, Geschlecht, Wohnort und Ort der Teilnahme), gezielte Akquisition und Integration unterrepräsentierter Gruppen (insbesondere Frauen, und weniger gebildete) bei der Stichprobenziehung auf verschiedensten Webseiten und auf dem Panel. Zudem werden nur verifizierte Teilnehmer integriert und eine Nachgewichtung der Stichprobengröße zum a-posteriori Abgleich der soziodemografischen a-priori-Verteilungsannahmen der Zielpopulation vorgenommen. Quoten bzw. Gewichtungen erlauben eine bessere Vergleichbarkeit und Repräsentativität von Stichproben für eine Population, besonders für die Gesamtbevölkerung eines Landes. Dabei sind bei Civey auch die Stichproben ausreichend groß, um Signifikanzen testen zu können (Duffy et al., 2005). Der Stichprobenfehler wird dadurch geringer. Wie bereits einschränkend vermerkt, erreichen und repräsentieren Online-Erhebungen, wenn sie im Rahmen von Erhebungspanels durchgeführt werden, eher nicht die gesamte Bevölkerungspopulation von Nationen. Analytische Verfahren ermöglichen es jedoch zunehmend Ergebnisse zu generieren, die mit einer spezifischen Wahrscheinlichkeit auf Bevölkerungspopulationen übertragbar sind. Mögliche Verzerrungen werden durch den Ein- bzw. Ausschluss von Selektionsvariablen, die Verwendung von Quoten-Designs und zusätzlicher Gewichtung angepasst (Dülmer, 2007). Online-Befragungen wie seitens Civey nutzen diese Potentiale und können so eine hohe Datenqualität und Repräsentativität erreichen. Allerdings müssen die

Methoden durchdacht kombiniert und fortlaufend an spezifische Systematiken angepasst werden, um die Gültigkeit für nationale Populationen aufzuzeigen. Dies erfolgt bei Civey – soweit eine externe Beurteilung dieses Urteil gestattet.

Gerade elektronische Befragungen beinhalten das Problem von Mehrfachantworten seitens der Befragten bzw. dem automatisierten Nichtzulassen von Antworten (Schmidt, 1997). Civey stellt sicher, dass Mehrfachantworten nicht auftreten. Mit der Durchführung der Registrierung, der Abfrage spezifischer Marker (Geburtsjahr, Geschlecht, Postleitzahl, Bildungsstand und Familienstand; vgl. Civey GmbH, 25.5.2018) sowie der Einverständniserklärung zur Datenverarbeitung setzt Civey ein Browser-Cookie beim Nutzer. Im folgenden Verifizierungsprozess wird überprüft, ob es sich bei dem registrierten Teilnehmer um eine natürliche Person handelt. Die Überprüfung erfolgt mittels der Hidden-Captchas-Technologie, die anhand von Nutzeraktionen bewertet, mit welcher Wahrscheinlichkeit die Aktionen von einer realen Person ausgeführt werden (Moradi & Keyvanpour, 2015). Während des Verifizierungsprozesses werden weitere spezifische Marker (beispielsweise Bildung, Beruf, Einkommen, Religion oder Familienstand) abgefragt und diese Marker müssen vom Teilnehmer bis zu einem gewissen Grad vollständig hinterlegt werden. Aus den Ergebnissen der Hidden-Captcha-Analyse, der Widersprüchlichkeit von Angaben und Antworten sowie aus der Vollständigkeit der Informationen bildet Civey einen sogenannten "User-Trust-Index", der zur Steigerung der Stichproben- und Datenqualität eingesetzt wird. So nutzt Civey bei den Online-Befragungen eine Vielzahl von Aktivitäten zur Steigerung der Qualität und Repräsentativität.

Angesichts der hier noch einmal verdichteten Argumente zeigt sich deutlich, dass das durchdachte Vorgehen von Civey durch sinnvolle Integration der Befragungen auf Websites, die Ansprache von Panelmitgliedern, die Größe des Panels inkl. weiterer Charakteristika der Personen sowie das Anlegen von Quoten und Gewichtungen zu einer möglichst hohen Repräsentativität sowie hohen Aussagekraft der Online-Befragungen führt. Vor dem Hintergrund der aktuellen Entwicklungen lässt sich nicht fundiert argumentieren, dass die Online-Befragungen bei Civey nicht repräsentativ sind. Insgesamt scheint der Vorwurf der mangelnden Aussagekraft und Repräsentativität bei Civey nicht gerechtfertigt.

Die hier geführten Analysen lassen nicht den Schluss zu, ob telefonische Befragungen besser oder schlechter als die Analyse von Civey sind. Die Beurteilung hängt immer davon ab, welches Ziel verfolgt wird und welche Vorgehensweise angelegt wird. Allerdings könnte Civey, wenngleich der potenzielle Nutzenzuwachs mit möglicherweise unverhältnismäßig hohem Aufwand verbunden ist, die Repräsentativität der Aussagen noch erhöhen, wenn das aktuell schon durchdachte und hochqualitative Vorgehen um schriftliche und telefonische Befragungen bei Nicht- oder Geringnutzern des Internets ergänzt werden.

Darüber hinaus lässt sich anführen, dass Repräsentativität im engeren Sinne oft als überbetont gilt. Die Aussagequalität von Befragungen hängt zudem von anderen Aspekten ab. Online-Befragungen zeichnen sich durch fast nicht existente Verzerrungen durch soziale Erwünschtheit aus. Gerade gegenüber persönlichen, aber auch gegenüber telefonischen Befragungen spielen Online-Erhebungen deutliche Vorteile im Bereich der geringeren Verzerrungen aufgrund sozialer Erwünschtheit aus. Aufgrund der Entpersonalisierung und Automatisierung bei Online-Befragungen generell und ebenfalls bei Civey ist von geringeren Verzerrungen der sozialen Erwünschtheit auszugehen. Darüber hinaus folgt man bei Civey explizit dem Primat der Neutralität. Texte und Bilder werden kritisch hin auf eine mögliche unterschwellige Beeinflussung geprüft. Fragen und Farben werden sensibel eingesetzt. Zudem durchlaufen die Umfragen vor der Veröffentlichung festgelegte Prozesse und Qualitätskontrollen. Unter anderem berät hier ein Gremium von mindestens drei (neutralen) Experten, ob unterschwellige Beeinflussungen in der Befragung vorhanden sind. Geringe Verzerrungen durch

soziale Erwünschtheit, Qualitätskontrolle im Allgemeinen plus audio-visueller und interaktiver Gestaltungsmöglichkeiten zeigen eine Stärke von Online-Befragungen wie Civey sie durchführt.

**Fazit:** Die Systematik bei Online-Befragungen seitens des Unternehmens Civey verwendet unterschiedliche durchdachte und attraktiv gestaltete Maßnahmen und Technologien, um kostengünstig, schnell und flexibel Markt-, Meinungs- und Wahlforschung mit gering gradigen Verzerrungen durch soziale Erwünschtheit und Minimierung des Stichprobenfehlers durchzuführen. Die Systematik bei Civey ermöglicht möglichst aussagekräftige und repräsentative Aussagen. Verbesserungen der Repräsentativität vor allem bei Internet-Nichtnutzern oder Wenignutzern des Internets lassen sich durch die Kombination mit schriftlichen und telefonischen Befragungen noch erzielen. Insgesamt ist aber ein Vorwurf von nicht repräsentativen Aussagen der Online-Befragungen nicht haltbar. Generelle Vorurteilungen wie telefonische Befragungen sind besser oder schlechter im Hinblick auf Aussagekraft und Repräsentativität sind ebenfalls nicht haltbar.

## Literaturverzeichnis

- AAPOR. 2014. Best Practices for Survey Research: American Association for Public Opinion Research.
- AAPOR. April 2016. Guidance on Reporting Precision for Nonprobability Samples: American Association for Public Opinion Research.
- ADM. Mai 2001. Standards zur Qualitätssicherung für Online-Befragungen: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute.
- Alan, C. 1998. Comparing the response rate, response speed and response quality of two methods of sending questionnaires: E-mail vs. mail. *International Journal of Market Research*, 40(4): 353.
- Albors, J., Ramos, J. C., & Hervas, J. L. 2008. New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management*, 28(3): 194-202.
- Anderson, S. E., & Gansneder, B. M. 1995. Using electronic mail surveys and computer-monitored data for studying computer-mediated communication systems. *Social Science Computer Review*, 13(1): 33-46.
- Andrews, D., Nonnecke, B., & Preece, J. 2003. Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*, 16(2): 185-210.
- Andrews, D., Nonnecke, B., & Preece, J. 2007. Conducting research on the internet: Online survey design, development and implementation guidelines. *International Journal of Human-Computer Interaction*, 16(2): 185-210.
- Ansolabehere, S., & Rivers, D. 2013. Cooperative Survey Research. *Annual Review of Political Science*, 16(1): 307-329.
- Ansolabehere, S., & Schaffner, B. 2014. Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis*, 22(3): 285-303.
- ARD/ZDF. 10.10.2018. Pressemitteilung: ARD/ZDF-Onlinestudie 2018: Ergebnisse aus der Studienreihe „Medien und ihr Publikum“ (MiP).
- Bachmann, D., Elfrink, J., & Vazzana, G. 1996. Tracking the progress of e-mail vs. snail-mail. *Marketing Research*, 8(2): 30.
- Backhaus, K., & Blechschmidt, B. 2009. Fehlende Werte und Datenqualität. *DBW*, 69(2): 265-285.
- Backhaus, K., Blechschmidt, B., & Eisenbeiß, M. 2006. Der Stichprobeneinfluss bei Kausalanalysen. *Die Betriebswirtschaft*, 66(6): 711-726.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. 2003. *Multivariate Analysemethoden* (10 ed.). Berlin: Springer-Verlag.
- Bagozzi, R. P., & Dholakia, U. M. 2006. Open source software user communities: A study of participation in Linux user groups. *Management Science*, 52(7): 1099-1115.
- Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., & Lavrakas, P. J. 2010. Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly*: 1-71.

- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. May 2013: Report of the AAPOR task force on non-probability sampling. American Association for Public Opinion research.
- Bartel, C. A., & Garud, R. 2009. The Role of Narratives in Sustaining Organizational Innovation. *Organization Science*, 20(1): 107-117.
- Batnic, B., & Moser, K. 2005. Determinanten der Rücklaufquote in Online-Panels. *Zeitschrift für Medienpsychologie*, 17 (N.F.5) 2: 64-74.
- Beebe, T. J., Mika, T., Harrison, P. A., Anderson, R. E., & Fulkerson, J. A. 1997. Computerized school surveys: Design and development issues. *Social Science Computer Review*, 15(2): 159-169.
- Bennemann, S. 1999. Virtual Communities als Instrument des Internet Marketing, *Online&Internet-Marketing: Perspektiven und Erfahrungen aus Deutschland und den US*: 258-278. Stuttgart: Schäffer-Poeschel.
- Berzofsky, M., Welch, B., Williams, R., & Biemer, P. 2008. Using a model-aided sampling paradigm instead of a traditional sampling paradigm in a nationally representative establishment survey. RTI Press Publication No. MR-0004-0802: Research Triangle Park.
- Berzofsky, M., Williams, R., & Biemer, P. 2009. Combining Probability and Non-Probability Sampling Methods: Model-Aided Sampling and the O\*NET Data Collection Program. *Survey Practice*(September 01).
- Birnbaum, M. H. 2004. Human research and data collection via the internet. *Annual Review of Psychology*, 55: 803-832.
- Bishop, I. D. 1997. Testing perceived landscape colour difference using the Internet. *Landscape and Urban Planning*, 37(3): 187-196.
- Blasius, J., & Brandt, M. 2009. Repräsentativität in Online-Befragungen. In M. Weichbold, J. Bacher, & C. Wolf (Eds.), *Umfrageforschung: Herausforderungen und Grenzen*: 157-177. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blasius, J., & Brandt, M. 2010. Representativeness in Online Surveys through Stratified Samples. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 107(1): 5-21.
- Braithwaite, D. O., Waldron, V. R., & Finn, J. 1999. Communication of Social Support in Computer-Mediated Groups for People with Disabilities. *Health Communications*, 11(2): 123-151.
- Brown, J. S., & Duguid, P. 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization science*, 2(1): 40-57.
- Buchanan, T., & Smith, J. L. 1999. Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90(1): 125-144.
- Bukszar, E., & Connolly, T. 1988. Hindsight Bias and Strategic Choice - Some Problems in Learning from Experience. *Academy of Management Journal*, 31(3): 628-641.
- Buttler, G., & Christian, B. 2000. Repräsentativität von Online-Umfragen. In W. Scheffler, & K.-I. Voigt (Eds.), *Entwicklungsperspektiven im Electronic Business: Grundlagen — Strukturen — Anwendungsfelder*: 205-216. Wiesbaden: Gabler Verlag.
- Calder, B., Phillips, L., & Tybout, A. 1982. The concept of external validity. *Journal of Consumer Research*, 9: 240-244.

- Callegaro, M., Villar, A., Krosnick, J., & Yeager, D. 2014. A Critical Review of Studies Investigating the Quality of Data Obtained With Online Panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A Data Quality Perspective*: Wiley.
- Chang, L., & Krosnick, J. A. 2009. National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4): 641-678.
- Chen, H., Chiang, R. H., & Storey, V. C. 2012. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4): 1165-1188.
- Cheung, M. W. L., & Chan, W. 2002. Reducing Uniform Response Bias With Ipsative Measurement in Multiple-Group Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(1): 55-77.
- Chu, J., & Manchanda, P. 2016. Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms. *Marketing Science*, 35(6): 870-893.
- Civey GmbH. 25.5.2018. Datenschutzerklärung.
- Civey GmbH. 2018. Call-Center Institute greifen Civey an.
- Compagni, A., Mele, V., & Ravasi, D. 2015. How Early Implementations Influence Later Adoptions of Innovation: Social Positioning and Skill Reproduction in the Diffusion of Robotic Surgery. *Academy of Management Journal*, 58(1): 242-278.
- Cook, C., Heath, F., & Thompson, R. L. 2000. A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, 60(6): 821-836.
- Couper, M. P. 2000. Web surveys: A review of issues and approaches. *The Public Opinion Quarterly*, 64(4): 464-494.
- Couper, M. P., Blair, J., & Triplett, T. 1997. Comparison of mail and e-mail for a survey of employees in federal statistical agencies, *American Association for Public Opinion Research*. Norfolk, VA.
- Couper, M. P., Blair, J., & Triplett, T. 1999. A comparison of mail and e-mail for a survey of employees in US statistical agencies. *Journal of official statistics*, 15(1): 39.
- Couper, M. P., Traugott, M. W., & Lamias, M. J. 2001. Web survey design and administration. *Public opinion quarterly*, 65(2): 230-253.
- Cumming, R. G. 1990. Is probability sampling always better? A comparison of results from a quota and a probability sample survey. *Community health studies*, XIV(2): 132-137.
- Dell'Era, C., & Verganti, R. 2011. Diffusion Processes of Product Meanings in Design-Intensive Industries: Determinants and Dynamics\*. *Journal of Product Innovation Management*, 28(6): 881-895.
- DeMars, C. E. 2012. A Comparison of Limited-Information and Full-Information Methods in Mplus for Estimating Item Response Theory Parameters for Nonnormal Populations. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(4): 610-632.
- Dillman, D. A. 2007. *Mail and Internet surveys: the tailored design, —2007 update* (2 ed.). Hoboken: John Wiley & sons.

- Dillman, D. A., Sangster, R. L., Tarnai, J., & Rockwood, T. H. 1996. Understanding differences in people's answers to telephone and mail surveys. *New Directions for Evaluation*, 1996(70): 45-61.
- Dixon, W., Smyth, G. K., & Chiswell, B. 1999. Optimized selection of river sampling sites. *Water Research*, 33(4): 971-978.
- Doty, D. H., & Glick, W. H. 1998. Common Methods Bias: Does Common Methods Variance Really Bias Results? *Organizational Research Methods*, 1(4): 374-406.
- Duffy, B., Smith, K., Terhanian, G., & Bremer, J. 2005. Comparing Data from Online and Face-to-face Surveys. *International Journal of Market Research*, 47(6): 615-639.
- Dülmer, H. 2007. Experimental Plans in Factorial Surveys: Random or Quota Design? *Sociological Methods & Research*, 35(3): 382-409.
- Ellis, C. H., & Krosnick, J. A. 1999. Comparing telephone and face-to-face surveys in terms of sample representativeness: A meta-analysis of demographic characteristics. *Ohio State University*.
- Ernst, H. 2003. Ursachen eines Informant Bias und dessen Auswirkung auf die Validität. *Zeitschrift für Betriebswirtschaft*: 1249-1275.
- ESOMAR. 2015. ESOMAR/GRBN Guideline for online sample quality: The World Association for Social, Opinion and Market Research.
- Evans, J. R., & Mathur, A. 2005. The value of online surveys. *Internet Research*, 15(2): 195-219.
- Eysenbach, G. 2004. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of medical Internet research*, 6(3).
- Fan, W., & Yan, Z. 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior*, 26(2): 132-139.
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. 2011. Knowledge Collaboration in Online Communities. *Organization Science*, 22(5): 1224-1239.
- Fisher, G. 2018. Online Communities and Firm Advantages. *Academy of Management Review*, in press(ja): null.
- Fisher, R. A. 1922. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society*, 222: 309-368.
- Fisher, R. A. 1971. *The Design of Experiments*. New York: Hafner Publishing Company.
- Foster Thompson, L., & Surface, E. A. 2007. Employee Surveys Administered Online: Attitudes Toward the Medium, Nonresponse, and Data Representativeness. *Organizational Research Methods*, 10(2): 241-261.
- Fouladi, R. T., McCarthy, C. J., & Moller, N. 2002. Paper-and-Pencil Or Online?: Evaluating Mode Effects on Measures of Emotional Functioning and Attachment. *Assessment*, 9(2): 204-215.
- Fricker, R. D. 2008. Sampling Methods for Web and E-mail Surveys. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods*: SAGE Publications, Ltd.
- Garton, L., Haythornthwaite, C., & Wellman, B. 1997. Studying online social networks. *Journal of computer-mediated communication*, 3(1): JCMC313.

- Garton, L., & Wellman, B. 1995. Social impacts of electronic mail in organizations: A review of the research literature, *Annals of the International Communication Association*, Vol. 18: 434-453.
- Gilljam, M., Granberg, D., Holm, B., Karlsson, D., Persson, M., & Sundell, A. 2011. Comparison of Web and Mail Responses in a Census of Swedish Local Political Representatives. *Survey Practice*, 4(3): 3046.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American psychologist*, 59(2): 93.
- Gruner, R., Homburg, C., & Lukas, B. 2014. Firm-hosted online brand communities and new product success. *Journal of the Academy of Marketing Science*, 42(1): 29-48.
- Gude, J. 18.10.2018. Pressekonferenz „Statistisches Jahrbuch 2018: Aktiv im Alter“. Wiesbaden: Statistisches Bundesamt.
- Guo, T., Schwartz, D. G., Burstein, F., & Linger, H. 2009. Codifying collaborative knowledge: using Wikipedia as a basis for automated ontology learning. *Knowledge Management Research & Practice*, 7(3): 206-217.
- Hahn, E. D., & Doh, J. P. 2006. Using Bayesian methods in strategy research: An extension of Hansen et al. *Strategic Management Journal*, 27(8): 783-798.
- Hair, J. F., Black, B., Babin, B., Anderson, R. E., & Tatham, R. L. 2010. *Multivariate data analysis* (7 ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Hayashi, K., & Yuan, K.-H. 2003. Robust Bayesian Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(4): 525-533.
- Heerwegh, D., & Loosveldt, G. 2008. Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly*, 72: 836-846.
- Hillygus, D. S., Jackson, N., & Young, M. 2014. Professional respondents in non-probability online panels. *Online panel research: A data quality perspective*: 219-237.
- Hofmann, O., Steinmeyer, S., & Paul, M. 2001. Online-Marktforschung — Andere Ergebnisse als Offline? Eine empirische Untersuchung zur Güte internetbasierter Copytests. In A. Theobald, M. Dreyer, & T. Starsetzki (Eds.), *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrungen*: 133-148. Wiesbaden: Gabler Verlag.
- Ilieva, J., Baron, S., & Healey, N. M. 2002. Online surveys in marketing research: Pros and cons. *International Journal of Market Research*, 44(3): 361-376.
- ISO. 2009. ISO 26362 Access panels in market, opinion, and social research: Vocabulary and service requirements. Geneva: International Organization for Standardization.
- Joshi, A., & Knight, A. P. 2015. Who Defers to Whom and Why? Dual Pathways Linking Demographic Differences and Dyadic Deference to Team Effectiveness. *Academy of Management Journal*, 58(1): 59-84.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. 2004. A comparison of web and mail survey response rates. *Public opinion quarterly*, 68(1): 94-101.
- Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., & Gimenez, A. May 2016. Evaluating Online Nonprobability Surveys. Pew Research Center.

- Kiesler, S., & Sproull, L. 1982. Managerial response to changing environments - Perspectives on problem sensing from social cognition. *Administrative Science Quarterly*, 27(4): 548-570.
- Kiesler, S., & Sproull, L. S. 1986. Response effects in the electronic survey. *Public Opinion Quarterly*, 50(3): 402-413.
- King, W. C., & Miles, E. W. 1995. A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of applied psychology*, 80(6): 643.
- Kittleston, M. J. 1995. An assessment of the response rate via the Postal Service and e-mail. *Health Values: The Journal of Health Behavior, Education & Promotion*.
- Koch, W., & Frees, B. 2017. ARD/ZDF-Onlinestudie 2017: Neun von zehn Deutschen online. *Media Perspektiven*(9): 434-446.
- Kraus, H. 2013. Big Data: Einsatzfelder und Herausforderungen: Arbeitspapier der FOM.
- Lee, S.-Y., & Shi, J.-Q. 2000. Bayesian Analysis of Structural Equation Model With Fixed Covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(3): 411-430.
- Li, M., Wei, K.-K., Tayi, G. K., & Tan, C.-H. 2016. The moderating role of information load on online product presentation. *Information & Management*, 53(4): 467-480.
- Lippe, P. v. d., & Kladroba, A. 2002. Repräsentativität von Stichproben. *Marketing ZFP*, 24(2): 139-146.
- Lu, I. R. R., & Thomas, D. R. 2008. Avoiding and Correcting Bias in Score-Based Latent Variable Regression With Discrete Manifest Items. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(3): 462-490.
- Lyness, K. S., & Brumit Kropf, M. 2007. Cultural Values and Potential Nonresponse Bias: A Multilevel Examination of Cross-National Differences in Mail Survey Response Rates, Vol. 10: 210-224.
- Maintz, G. 2003. Leistungsfähigkeit älterer Arbeitnehmer - Abschied vom Defizitmodell. In B. B., S. H., & V. C. (Eds.), *Demographischer Wandel: Herausforderung für die betriebliche Personal- und Gesundheitspolitik. Fehlzeiten-Report, vol 2002.*: 43-55. Berlin, Heidelberg: Springer.
- Malhotra, N. 2008. Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72: 914-934.
- Manfreda, K. L., Berzelak, J., Vehovar, V., Bosnjak, M., & Haas, I. 2008. Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1): 79-104.
- Martin, C. L., & Nagao, D. H. 1989. Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, 74(1): 72.
- Martins, L. L., & Shalley, C. E. 2011. Creativity in Virtual Work: Effects of Demographic Differences. *Small Group Research*, 42(5): 536-561.
- Mathes, R., & Czaplicki, A. 1993. Meinungsführer im Mediensystem: Topdown ‘-und ,Bottom-up ‘-Prozesse. *Publizistik*, 38(2): 153-166.
- McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. 2018. What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, 44(7): 2909-2933.

- McPhee, L., & Lieb, J. 1999. Internet users top 92 million in the United States and Canada: Cupertino, CA: CommerceNet.
- Meade, A. W., Watson, A. M., & Kroustalis, C. M. 2007. Assessing Common Methods Bias in Organizational Research, *22nd Annual Meeting of the Society for Industrial and Organizational Psychology*. New York.
- Mehta, P. D., & Neale, M. C. 2005. People are variables too: multilevel structural equations modeling. *Psychol Methods*, 10(3): 259-284.
- Mehta, R., & Sivadas, E. 1995. Comparing response rates and response content in mail versus electronic mail surveys. *Market Research Society. Journal.*, 37(4): 1-12.
- Mercer, A. W., Kreuter, F., Keeter, S., & Stuart, E. A. 2017. Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference. *Public Opinion Quarterly*, 81(S1): 250-271.
- Milliken, F. J. 1990. Perceiving and interpreting environmental change: An examination of college administrators' interpretation of changing demographics. *Academy of Management Journal*, 33(1): 42-63.
- Moradi, M., & Keyvanpour, M. 2015. CAPTCHA and its Alternatives: A Review. *Security and Communication Networks*, 8(12): 2135-2156.
- Muhtaseb, A., & Frey, L. R. 2008. Arab Americans' Motives for Using the Internet as a Functional Media Alternative and Their Perceptions of U.S. Public Opinion. *Journal of Computer-Mediated Communication*, 13(3): 618-657.
- Niculescu, A., & Thorsteinsson, G. 2011. Enabling Idea Generation through Computer-Assisted Collaborative Learning. *Studies in Informatics and Control*, 20(4): 403-410.
- Noelle-Neumann, E., Haumann, W., & Petersen, T. 1998. Die Wiederentdeckung der Meinungsführer und die Wirkung der persönlichen Kommunikation im Wahlkampf. *Kampa. Meinungsklima und Medienwirkung im Bundestagswahlkampf*: 181-214.
- Nulty, D. D. 2008. The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, 33(3): 301-314.
- Ocker, R. J., & Yaverbaum, G. J. 1999. Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction. *Group Decision and Negotiation*, 8(5): 427-440.
- Olson, E. L., & Bakke, G. 2001. Implementing the lead user method in a high technology firm: A longitudinal study of intentions versus actions. *J Product Innovation Man*, 18(6): 388-395.
- Parker, L. 1992. Collecting data the e-mail way. *Training & Development*, 46(7): 52-55.
- Pasek, J. 2016. When will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence. *International Journal of Public Opinion Research*, 28(2): 269-291.
- Rat der Deutschen Markt- und Sozialforschung e.V. 2007. Richtlinie für Online-Befragungen: Rat der Deutschen Markt- und Sozialforschung e.V.
- Reischauer, G., & Mair, J. 2018. How Organizations Strategically Govern Online Communities: Lessons from the Sharing Economy. *Academy of Management Discoveries*, 4(3): 220-247.

- Roy, S. 2010. So you already have a survey database? – A seven-step methodology for theory building from survey databases: An illustration from incremental innovation generation in buyer-seller relationships. *Journal of Supply Chain Management*, 46(4): 12-24.
- Rupp, A. A., Dey, D. K., & Zumbo, B. D. 2004. To Bayes or Not to Bayes, From Whether to When: Applications of Bayesian Methodology to Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3): 424-451.
- Sax, L. J., Gilmartin, S. K., & Bryant, A. N. 2003. Assessing response rates and nonresponse bias in web and paper surveys. *Research in higher education*, 44(4): 409-432.
- Schaefer, D. R., & Dillman, D. A. 1998. Development of a standard e-mail methodology: Results of an experiment. *Public opinion quarterly*: 378-397.
- Schleyer, T. K., & Forrest, J. L. 2000. Methods for the design and administration of web-based surveys. *Journal of the American Medical Informatics Association*, 7(4): 416-425.
- Schmidt, F. L., & Hunter, J. E. 1998. The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124: 262-274.
- Schmidt, W. C. 1997. World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior research methods, instruments, & computers*, 29(2): 274-279.
- Schmidt, W. C. J. B. R. M., Instruments, & Computers. 1997. World-Wide Web survey research: Benefits, potential problems, and solutions. 29(2): 274-279.
- Scholderer, J., Grunert, K. G., & Brunsø, K. 2005. A procedure for eliminating additive bias from cross-cultural survey data. *Journal of Business Research*, 58(1): 72-78.
- Schuldt, B. A., & Totten, J. W. 1994. Electronic mail vs. mail survey response rates. *Marketing Research*, 6(1): 3-7.
- Schumann, S., & Klein, O. 2015. Substitute or stepping stone? Assessing the impact of low-threshold online collective actions on offline participation. *European Journal of Social Psychology*, 45(3): 308-322.
- Serfling, O., Kastl, J., & Sengül, D. 2018. Der statistische Fehler als Qualitätsindikator bei Civey: Bayesianische Kreditabilitätsintervalle in der Berechnung des statistischen Fehlers bei Onlineumfragen.
- Sills, S. J., & Song, C. 2002. Innovations in survey research: An application of web-based surveys. *Social science computer review*, 20(1): 22-30.
- Skowronek, D., & Duerr, L. 2009. The convenience of nonprobability: Survey strategies for small academic libraries. *College & Research Libraries News*, 70(7): 412-415.
- Song, X.-Y., & Lee, S.-Y. 2008. A Bayesian Approach for Analyzing Hierarchical Data With Missing Outcomes Through Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(2): 272-300.
- Song, X.-Y., Lee, S.-Y., & Hser, Y.-I. 2009. Bayesian Analysis of Multivariate Latent Curve Models With Nonlinear Longitudinal Latent Effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2): 245-266.

- Song, X.-Y., Lu, Z.-H., Hser, Y.-I., & Lee, S.-Y. 2011. A Bayesian Approach for Analyzing Longitudinal Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2): 183-194.
- Sparrow, N. 2007. Quality issues in online research. *Journal of advertising research*, 47: 179–182.
- Sproull, L., Dutton, W., & Kiesler, S. 2007. Introduction to the Special Issue: Online Communities, Vol. 28: 277-281.
- Sproull, L., & Kiesler, S. 1986. Reducing social context cues: Electronic mail in organizational communication. *Management Science*, 32(11): 1492-1512.
- Sproull, L. S. 1986. Using electronic mail for data collection in organizational research. *Academy of management journal*, 29(1): 159-169.
- Stanton, J. M. 1998. An empirical assessment of data collection using the Internet. *Personnel Psychology*, 51(3): 709-725.
- Steber, C. 2018. River Sampling vs. Panel Sampling: What's the Difference? In C. C. f. Research (Ed.), *Market Research Blog*.
- Steel, P. D. G., & Kammeyer-Mueller, J. 2007. Bayesian Variance Estimation for Meta-Analysis: Quantifying Our Uncertainty: 1094428107300339.
- Stocké, V. 2004. Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, 33(4): 303-320.
- Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. 2014. The Prowess and Pitfalls of Bayesian Structural Equation Modeling: Important Considerations for Management Research. *Journal of Management*.
- Strong, D. M., Lee, Y. W., & Wang, R. Y. 1997. Data quality in context. *Communications of the ACM*, 40(5): 103-110.
- Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. 2014. Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research*: 61-85.
- Swoboda, W. J., Mühlberger, N., Weitkunat, R., & Schneeweiß, S. 1997. Internet surveys by direct mailing: An innovative way of collecting data. *Social Science Computer Review*, 15(3): 242-255.
- Szolnoki, G., & Hoffmann, D. 2013. Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy*, 2(2): 57-66.
- Talke, K., & Hultink, E. J. 2010. Managing Diffusion Barriers When Launching New Products. *Journal of Product Innovation Management*, 27(4): 537-553.
- Thatcher, S. M. B. 2013. Moving Beyond a Categorical Approach to Diversity: The Role of Demographic Faultlines. In Q. M. Roberson (Ed.), *The Oxford Handbook of Diversity and Work*: Oxford University Press.
- Treiblmaier, H. 2011. Datenqualität und Validität bei Online-Befragungen. *der markt*, 50(1): 3-18.

- Tse, A., Ching, R., Ding, Y.-B., Fong, R., Yeung, E., & Au, A. 1994. Miscellany: A Comparison of the Effectiveness of Mail and Facsimile as Survey Media on Response Rate, Speed and Quality. *Market Research Society. Journal.*, 36(4): 1-8.
- Tse, A. C. 1998. Comparing Response Rate, Response Speed and Response Quality of Two Methods of Sending Questionnaires: E-mail vs. Mail. *Market Research Society*, 40(4): 1-12.
- Van den Bulte, C. 2004. Multigeneration innovation diffusion and intergeneration time: A cautionary note. *Journal of the Academy of Marketing Science*, 32(3): 357-360.
- von Hippel, E. 1986. Lead Users: A Source of Novel Product Concepts. *Management Science*, 32(7): 791-805.
- Walsh, J. P., Kiesler, S., Sproull, L. S., & Hesse, B. W. 1992. Self-Selected and Randomly Selected Respondents in a Computer Network Survey. *The Public Opinion Quarterly*, 56(2): 241-244.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. 2017. Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70: 356-365.
- Ward, M., & Meade, A. W. 2018. Applying Social Psychology to Prevent Careless Responding during Online Surveys. *Applied Psychology*, 67(2): 231-263.
- Watt, S., Simpson, C., McKillop, C., & Nunn, V. 2002. Electronic course surveys: Does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education*, 27(4): 325-337.
- Weichbold, M. 2009. Zur Bestimmung und Sicherung der „Qualität“ von Umfragen. In M. Weichbold, J. Bacher, & C. Wolf (Eds.), *Umfrageforschung: Herausforderungen und Grenzen*: 553-570. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weigold, A., Weigold, I. K., & Russell, E. J. 2013. Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, 18(1): 53-70.
- WELT.de. 05.05.2018. Mehr als drei Viertel der Deutschen befürworten Ankerzentren.
- West, J., & Lakhani, K. R. 2008. Getting Clear About Communities in Open Innovation. *Industry and Innovation*, 15(2): 223-231.
- Winkler, N., Kroh, M., & Spiess, M. 2006. Entwicklung einer deutschen Kurzsкала zur zweidimensionalen Messung von sozialer Erwünschtheit: DIW Discussion Papers.
- Wirtz, B. W., Schmidt-Holz, R., & Beaujean, R. 2004. *Deutschland Online 2 - Bericht*.
- Wright, K. B. 2005. Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, 10(3): JCMC1034-JCMC1034.
- Wu, C.-C., Chen, Y.-J., & Cho, Y.-J. 2013. Nested Network Effects in Online Free Games with Accessory Selling. *Journal of Interactive Marketing*, 27(3): 158-171.
- Yang, K., & Banamah, A. 2014. Quota Sampling as an Alternative to Probability Sampling? An Experimental Study. *Sociological Research Online*, 19(1): 29.

Yun, G. W., & Trumbo, C. W. 2000. Comparative Response to a Survey Executed by Post, E-mail, & Web Form. *Journal of Computer-Mediated Communication*, 6(1): JCMC613-JCMC613.

ZAW. 2018. ZAW-Rahmenschema für Werbeträgeranalysen: Zentralverband der deutschen Werbewirtschaft ZAW e.V.

Zhang, Z., Lai, K., Lu, Z., & Tong, X. 2013. Bayesian Inference and Application of Robust Growth Curve Models Using Student's t Distribution. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1): 47-78.

Zhang, Z., McArdle, J. J., Wang, L., & Hamagami, F. 2008. A SAS Interface for Bayesian Analysis With WinBUGS. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(4): 705-728.

Zinkhan, G. M. 2004. Accessing academic research through an e-database: Issues of journal quality and knowledge use. *Journal of the Academy of Marketing Science*, 32(4): 369-370.

Zyphur, M. J., & Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41(2): 390-420.

Zyphur, M. J., Oswald, F. L., & Rupp, D. E. 2015. Rendezvous Overdue: Bayes Analysis Meets Organizational Research. *Journal of Management*, 41(2): 387-389.