



---

# Wissenschaftliches Gutachten

---

Qualitätsanforderungen an Studien  
zur Ableitung von Regulierungsmaßnahmen:  
Kritische Evaluation des Glücksspiel-Surveys 2021

Katharina Schüller

# Qualitätsanforderungen an Studien zur Ableitung von Regulierungsmaßnahmen: Kritische Evaluation des Glücksspiel-Surveys 2021

Katharina Schüller<sup>1</sup>

Wissenschaftliche Beratung: Prof. Dr. Ralf Münnich<sup>2</sup>

München

September 2023

---

<sup>1</sup>Geschäftsführung STAT-UP GmbH, Vorstandsmitglied Deutsche Statistische Gesellschaft

<sup>2</sup>Professur für Wirtschafts- und Sozialstatistik der Universität Trier

# STAT-UP

## Statistical Consulting & Data Science GmbH

Augustenstr. 5

D - 80333 München

Geschäftsf.: Katharina Schüller

Telefon: +49 (89) 34 077 451

Telefax: +49 (89) 34 077 453

E-Mail: [muenchen@stat-up.com](mailto:muenchen@stat-up.com)

Web: [www.stat-up.com](http://www.stat-up.com)

Rechtlicher Hinweis: Dieses Gutachten wurde von der STAT-UP Statistical Consulting & Data Science GmbH erstellt und von den folgenden Verbänden in Auftrag gegeben: Bundesverband deutscher Spielbanken gegr. 2008 als BupriS e.V. (BupriS), Deutscher Online Casinoverband e.V. (DOCV), Deutscher Sportwettenverband e.V. (DSWV), Die Deutsche Automatenwirtschaft e.V. (DAW). Die aus der Untersuchung abgeleiteten Schlussfolgerungen und Empfehlungen erfolgten nach bestem Wissen von STAT-UP Statistical Consulting & Data Science GmbH sowie nach den anerkannten Regeln der wissenschaftlichen Praxis. Dieses Gutachten dient ausschließlich dem Auftraggeber, es hat keine Schutzwirkung gegenüber Dritten und begründet daher keinerlei Haftung von STAT-UP Statistical Consulting & Data Science GmbH für Ansprüche oder Schäden Dritter gleich aus welchem Rechtsgrund, die aus der Kenntnis oder Nutzung dieses Gutachtens oder daraus resultierenden Handlungen entstehen können.

Hinweis: Zur besseren Lesbarkeit wird in diesem Gutachten das generische Maskulinum verwendet. Die in dieser Arbeit verwendeten Personenbezeichnungen beziehen sich – sofern nicht anders kenntlich gemacht – auf alle Geschlechter.

# Abstract

*Was braucht Politik von der Wissenschaft, um evidenzbasiert zu sein? Das vorliegende Gutachten setzt sich anhand einer kritischen Evaluation des Glücksspiel-Surveys 2021 mit der Frage auseinander, welche Qualitätsanforderungen Studien erfüllen müssen, um Evidenzgrundlagen für regulatorische Maßnahmen liefern zu können, und kommt zum Schluss: Der Glücksspiel-Survey 2021 ist zur Erreichung der Mehrheit seiner multiplen Ziele qua Design nicht geeignet. Seine Ergebnisse bilden aufgrund fragwürdiger Präzision und Intransparenz keine belastbare Entscheidungsgrundlage hinsichtlich der Bewertung und Anpassung gesetzlicher Regelungen.*

*Diese Einschätzung ergibt sich aus drei Kernproblemen: (1) der fehlenden Nennung und Diskussion der gravierenden Nonresponse-Quote im Survey, (2) der gravierenden Verzerrung durch die (teilweise) Nutzung eines Online-Panels und (3) einer unzureichenden Gewichtung der Befragungsergebnisse. Der Umstieg auf ein teilweise nichtprobabilistisches Mixed-Mode-Design führt zu einer Verschlechterung der Datenqualität: Die Nonresponse-Quote ist unannehmbar hoch und aufgrund des Selektionsfehler ist nach wissenschaftlichem Stand eine Überschätzung der Zahl von Menschen mit Glücksspielproblemen anzunehmen. Die Autoren des Glücksspiel-Surveys 2021 leiten zudem kausale Aussagen bezüglich der Effektivität bestehender Regulatorik, der Gefährlichkeit einzelner Spielformen und der Wirkungen des neuen Glücksspielstaatsvertrag aus den von ihnen erhobenen Daten ab, obwohl die Daten aufgrund des dafür ungeeigneten Erhebungsdesigns grundsätzlich nicht geeignet sind, Kausalschlüsse zuzulassen. Des Weiteren senkt der Survey den Schwellenwert für das Vorliegen „problematischen Spielverhaltens“ nach DSM-5 entgegen allen vergleichbaren vorherigen Untersuchungen und unvalidiert von vier auf lediglich ein erfülltes DSM-5-Kriterium ab. Die Gruppe derer, die nur ein bis drei DSM-5-Kriterien erfüllen, bezeichnet der Survey als „riskante Spieler“ (5,7 %) und rechnet sie entgegen allen vergleichbaren vorangegangenen Studien der Gruppe der „Problemspieler“ zu. Erst hierdurch vervielfacht sich der Anteil der Personen in Deutschland zwischen 18 und 70 Jahren, die der Survey als „problematische Spieler“ einstuft, von 2,3 % auf 8,0 % (auf die weithin rezipierte Zahl von rund 5 Mio. Menschen).*

*Weiterhin lassen sich die Ergebnisse des Glücksspiel-Surveys entgegen wissenschaftlicher Qualitätsstandards nicht überprüfen, da die Autoren Mängel des Surveys, die ihnen zumindest teilweise bekannt sind, nicht diskutieren und trotz ausführlich begründeter Anfrage für das vorliegende Gutachten weder Daten noch den Fragebogen zur Verfügung stellten. Die seitens der Survey-Autoren zu verantwortende Intransparenz hinsichtlich gravierender Limitationen und die Fehlschlüsse, die im Survey selbst sowie in der begleitenden Kommunikation gezogen werden, erklären die zahlreichen Beispiele statistischer Fehlrezeption der Survey-Ergebnisse im gesellschaftlichen, politischen und medialen Bereich.*

# Abstract

*What does policy need from science to be evidence-based? Based on a critical evaluation of the Gambling Survey 2021, this report addresses the question of what quality requirements studies must meet in order to provide evidence bases for regulatory measures, and concludes: The Gambling Survey 2021 is not suitable for achieving the majority of its multiple objectives by design. Due to its questionable precision and lack of transparency, its results do not form a reliable basis for decision-making with regard to the evaluation and adaptation of legal regulations.*

*This assessment stems from three fundamental issues: (1) the lack of identifying and discussing the severe non-response rate in the survey, (2) the serious bias due to the (partial) use of an online panel and (3) insufficient weighting of the survey results. Switching to a partially nonprobabilistic mixed-mode design leads to a deterioration in data quality: the non-response rate is unacceptably high and, due to selection bias, scientific evidence suggests an overestimation of the number of people with gambling problems. The authors of the Gambling Survey 2021 also derive causal statements regarding the effectiveness of existing regulations, the riskiness of individual forms of gambling and the effects of the new State Treaty on Gambling from the data they collected, even though the data is fundamentally not suitable to allow causal conclusions due to the inappropriate survey design. Furthermore, contrary to all comparable previous studies and unvalidated, the Survey lowers the threshold for the presence of „problem gambling behavior“ according to DSM-5 from four to only one fulfilled DSM-5 criterion. The group of those who meet only one to three DSM-5 criteria is referred to by the survey as „risky gamblers“ (5.7 %) and, contrary to all comparable previous studies, is classified as „problem gamblers“. Only in this way does the proportion of people in Germany between the ages of 18 and 70 whom the survey classifies as „problem gamblers“ multiply from 2.3 % to 8.0 % (to the widely received figure of around 5 million people).*

*Furthermore, contrary to scientific quality standards, the results of the Gambling Survey cannot be verified, as the authors do not discuss shortcomings of the survey, of which they are at least par-*

*tially aware. Moreover, despite a detailed justified request, the authors did not provide data or the questionnaire for the present report. The lack of transparency with regard to serious limitations, for which the survey authors are responsible, and the erroneous conclusions drawn in the survey itself and in the accompanying communication explain the numerous examples of statistical misreception of the survey results in the social, political and media spheres.*

# Executive summary

Eine evidenzbasierte Antwort auf die Frage, in welchem Ausmaß Politik in verschiedene Märkte regulierend eingreifen sollte, bedarf grundsätzlich einer aussagekräftigen demoskopischen Evaluation, die höchsten wissenschaftlichen Qualitätsstandards entspricht, wie sie beispielsweise in den Richtlinien der Deutschen Forschungsgemeinschaft (DFG) formuliert sind. Eine solche Evaluation beruht insbesondere auf geeigneten, auf Zufall basierenden Stichproben, präzise gemessenen Daten und angemessenen Analysemethoden.

***Der bislang nicht peer-reviewte Glücksspiel-Survey 2021 soll dahingehend untersucht werden, ob er Qualitätsanforderungen an Studien als Grundlage evidenzbasierter Politik erfüllt.*** Zur speziellen Frage, inwiefern die bestehende Glücksspielregulierung in Deutschland zielführend ist bzw. angepasst werden sollte, will der Glücksspiel-Survey 2021 ebendiese Grundlage liefern. Im März 2022 haben Sven Buth, Gerhard Meyer und Jens Kalke ihren Forschungsbericht „Glücksspielteilnahme und glücksspielbezogene Probleme in der Bevölkerung – Ergebnisse des Glücksspiel-Surveys 2021“ vorgelegt (Buth, Meyer und Kalke 2022). Es handelt sich um den ersten von drei Surveys, die im Abstand von jeweils zwei Jahren durchgeführt werden sollen. Projektträger sind das ISD Institut für interdisziplinäre Sucht- und Drogenforschung in Hamburg und die Arbeitseinheit Glücksspielforschung der Universität Bremen; gefördert wird das Projekt vom Deutschen Lotto- und Totoblock (DLTB).

Ziel der Surveys ist die Erhebung einer aktuellen validen Zahlenbasis über das Glücksspielverhalten in Deutschland, über Art und Ausmaß von problematischem Glücksspielverhalten und über Zusammenhänge u.a. zwischen Glücksspiel und Alkoholkonsum, allgemeinen psychischen Belastungen sowie Werbung für Glücksspiel. Damit bezwecken die Autoren einerseits, neue Erkenntnisse über verschiedene Aspekte des Glücksspiels zu gewinnen, andererseits sollen verschiedene gesetzliche Bestimmungen evaluiert werden, um (gegebenenfalls) eine Datenbasis für deren Verbesserung bereitzustellen.

Tatsächlich haben der Survey und die Aufbereitung seiner Ergebnisse durch die Autoren im großen Um-

fang Eingang in aktuelle wissenschaftliche, politische und gesellschaftliche Debatten über das Glücksspiel in Deutschland gefunden. Insbesondere die kontraintuitive Ergebnisinterpretation der Survey-Autoren, wonach 8 % der 18-bis-70-jährigen Bevölkerung „problematisches Spielverhalten“ aufweisen sollen (vgl. Buth, Meyer und Kalke (2022), S. 6; die letzte Erhebung der BZgA 2019 erhob noch einen Bevölkerungsanteil von Problemspielern von lediglich 0,73 %), wurde politisch und medial breit rezipiert; etwa der Beauftragte für Sucht- und Drogenfragen der Bundesregierung und der Senator für Inneres der Freien Hansestadt Bremen leiteten hieraus Forderungen nach einer restriktiveren Glücksspielregulierung ab.

Ein Peer-Review des Surveys, d.h. eine Begutachtung der Qualität der Arbeit durch unabhängige Wissenschaftler, hat bis heute ebenso wenig stattgefunden wie eine kritische Diskussion in der Fachliteratur. Beides gehört zu den Selbstverständlichkeiten des transparenten Wissenschaftsbetriebs, zumal wenn Forschungsergebnisse Einfluss auf politische Entscheidungen nehmen. Gerade dann sind besonders hohe Qualitätsmaßstäbe anzulegen. Die privatwirtschaftlichen Glücksspielanbieter und ihre Branchenverbände veranlasste dieses Fehlen einer wissenschaftlichen Qualitätskontrolle zur Beauftragung der vorliegenden Begutachtung des Surveys durch die Statistikerin Katharina Schüller, Vorstandsmitglied der Deutschen Statistischen Gesellschaft.

Im Folgenden werden die Ergebnisse der kritischen Auseinandersetzung mit dem Survey anhand allgemein anerkannter Qualitätskriterien kurz dargestellt. Dazu gehören die Stichprobenauswahl, der Umgang mit systematischen und unsystematischen Fehlern und die Eignung des verwendeten Messinstruments. Da der Glücksspiel-Survey 2021 unter anderem Daten als Grundlage für staatliche Maßnahmen und deren Bewertung zur Verfügung stellen will, muss er an Qualitätskriterien gemessen werden, die für eine Erhebung mit dem Ziel der Information oder Beratung von staatlichen Akteuren gelten. Die breite Rezeption des Surveys in Exekutive, Rechtsprechung und Medien stützt diese Sichtweise. Demnach müssen Qualität und Präzision der Daten höchste Priorität haben.

**Die Stichprobenauswahl führt zu nicht repräsentativen Daten.** Der Glücksspielsurvey 2021 will im Vergleich zu den vorherigen Surveys der Bundeszentrale für gesundheitliche Aufklärung (BZgA) einen methodischen Neustart durchführen, indem 39 % der Befragten nun aus einem Online-Panel stammen. Es gibt jedoch gewichtige Gründe für die Annahme, dass dieser Methodenwechsel zu einer Verschlechterung der Datenqualität führt:

- Erstens ist das gravierende Problem der Antwortverweigerer (*Nonresponse*) bei telefonischen Befragungen eben gerade nicht, wie von den Autoren suggeriert, auf diese beschränkt, sondern betrifft auch Online-Befragungen, die zusätzlich noch unter einem (*Self-*)*Selection-Error* leiden.

Dies ist in der Umfrageforschung nicht nur hinreichend bekannt und erforscht, sondern zeigt sich hier in besonderem Maße: Die *Nonresponse*-Quote im Glücksspiel-Survey in der Online-Befragung (90 %) ist höher als in der telefonischen Befragung (73 %), was im Survey allerdings ignoriert und nicht diskutiert wird.

- Zweitens zitieren die Autoren stattdessen selektiv Passagen aus der Literatur bezüglich der Vorteile von Online-Befragungen, obwohl dieselbe Literatur eindeutig von Online-Panels in Befragungen abrät, wenn das Ziel die präzise Erfassung von Prävalenzen bzw. allgemein Bevölkerungsanteilen ist.
- Drittens berufen die Autoren sich auf einen revidierten Report zur Herleitung des *Mixed-Mode*-Designs (Telefon- und Online-Befragung mit Selbst-Selektion), dessen Ergebnisse in einem späteren (auf dem Report basierenden) peer-reviewten Artikel explizit gegen die Inklusion eines Online-Panels sprechen.

Der Umstieg auf ein *Mixed-Mode*-Design, das zum Teil *nicht-probabilistisch* ist, hat im Ergebnis keine methodologische Grundlage und führt zu nicht objektiv korrigierbaren Fehlabschätzungen, beispielsweise bei der Anzahl von Menschen, die Glücksspiel spielen und Glücksspielprobleme haben.

***Zufallsbedingte Schwankungen der Ergebnisse werden nicht hinreichend kommuniziert und tendenziell unterschätzt.*** Der Glücksspiel-Survey 2021 muss sich entsprechend der einschlägigen Literatur an der Präzision seiner Ergebnisse messen lassen. Das bedeutet konkret, dass die Schätzungen aus dem Survey – zum Beispiel der Anteil an Menschen mit Glücksspielstörung – so nah wie möglich an den wahren Werten der *Grundgesamtheit* liegen sollten, d.h. die Auswahlgesamtheit (*Sampling-Frame*) muss die Grundgesamtheit vollständig abbilden und die Inklusionswahrscheinlichkeit jeder Einheit der Auswahlgesamtheit muss positiv und bekannt sein. Diese Evaluation wird im vorliegenden Gutachten anhand des *Total-Survey-Error*-Frameworks (Biemer 2010; Lohr 2021) durchgeführt, soweit dies möglich ist, denn die Autoren des Glücksspiel-Surveys 2021 haben die Bereitstellung der dafür notwendigen Materialien – entgegen wissenschaftlichen Qualitätsstandards (Deutsche Forschungsgemeinschaft 2019) – trotz begründeter Anfrage verweigert. Diese mangelnde Transparenz erschwert die Einschätzungen potenzieller Verzerrungen aus verschiedenen Quellen, da über die klassischen Fehlerquellen in Befragungen im Glücksspiel-Survey 2021 sehr wenig bis gar nicht berichtet wird.

Ein wesentlicher Kritikpunkt ist das Fehlen durchgängiger Angaben zum (zufälligen) Stichprobenfehler, der üblicherweise in Form von Konfidenzintervallen angegeben wird. Im Survey wird die Unsi-

cherheit, die mit jeder Stichprobe verbunden ist, in einer für wissenschaftliche Studien unüblich selektiven und überoptimistischen Weise kommuniziert, was das Risiko für Fehlinterpretationen erhöht und offenkundig zu undifferenzierten Aussagen im Survey führte. So erscheinen Unterschiede zwischen Teilgruppen als bedeutsam, obwohl in der überwiegenden Zahl der Fälle reine Zufallsschwankungen nicht auszuschließen sind. Da zudem beim Vergleich von Teilgruppen ein sogenanntes „multiples Testproblem“ vorliegt, müssen die Konfidenzintervalle geeignet adjustiert werden, da andernfalls der zufällige Fehler unterschätzt wird. Zudem erscheint die wechselnde Kategorisierung der Glücksspielarten teilweise willkürlich und erweckt den Eindruck, als sollten Unterschiede herausgestellt werden, die bei einer strikt sachlogischen oder zumindest durchgehend einheitlichen Kategorisierung unauffällig wären.

***Systematische Fehler begünstigen eine Fehlabschätzung der Prävalenz von Glücksspielstörungen und werden nicht hinreichend korrigiert.*** Zum anderen ist aufgrund aktueller wissenschaftlicher Literatur von einem großen (*Self-*)*Selection-Error* in der Online-Stichprobe des Surveys auszugehen, was eine Fehlabschätzung der Prävalenz von Glücksspielproblemen begünstigt. Zusätzlich ist aufgrund der niedrigen Ausschöpfungsquote der Befragung (Telefon: 27 %, Online: 10 %) und des Entfernens von unvollständigen Interviews von einem erheblichen *Nonresponse-Error* auszugehen. Der Umstieg auf ein teilweise nichtprobabilistisches Mixed-Mode-Design führt deshalb zu einer Verschlechterung der Datenqualität: Die Nonresponse-Quote ist unannehmbar hoch und aufgrund des Selektionsfehler ist nach wissenschaftlichem Stand eine Überschätzung der Zahl von Menschen mit Glücksspielproblemen anzunehmen.

Weitere Verzerrungen durch *Measurement-Error* und *Processing-Error* können nicht evaluiert werden, weil die dafür notwendigen Informationen nicht von den Studienautoren zur Verfügung gestellt wurden. Im Ergebnis sind die Resultate des Surveys aufgrund verschiedener methodischer Probleme höchstwahrscheinlich stark verfälscht. Entgegen wissenschaftlichen Standards gehen die Autoren jedoch nicht auf diese Problematiken ein. Folgerichtig wurden die Ergebnisse des Surveys in der wissenschaftlichen Rezeption bisher nur mit großer Zurückhaltung aufgenommen, wie beispielsweise bei Bühringer u. a. (2023) dokumentiert.

**Die verwendeten Korrekturmechanismen sind für den probabilistischen Teil der Befragung unzureichend; für den nicht-probabilistischen Teil ist eine strikt objektive Korrektur qua Design nicht möglich.** Die Autoren des Glücksspiel-Surveys 2021 versuchen die existierenden Verzerrungen in den Daten mit der sehr simplen *Poststratification*-Gewichtung auszugleichen, die dies allerdings fast ausschließlich im telefonisch durchgeführten (*Probability-Sampling*) Teil der Befragung

(und dort auch nur in begrenztem Ausmaß) korrigieren kann.

Angebracht wären stattdessen komplexere Gewichtungsverfahren (Ermittlung von Response-Propensities in Kombination mit weiteren Modellannahmen), die jedoch bei *Nonprobability-Samples* immer auf starken Modellannahmen beruhen, deshalb mindestens in Teilen als subjektiv anzusehen sind und grundsätzlich einer Sensitivitätsanalyse zur Abschätzung des Modellrisikos unterzogen werden sollten. Grundsätzlich hätten im Glücksspiel-Survey 2021 von Beginn an mehr Gewichtungsvariablen (zum Beispiel Arbeitslosigkeit) abgefragt werden müssen, die die Effektivität der Korrekturmechanismen gesteigert hätten. Die *Mode*-Gewichtung im Survey überzeugt aufgrund der Heranziehung überholter Begründungen und genutzter Aggregatdaten ebenso wenig. Zudem bleibt offen, warum die Autoren nur mit dem Umsatz des DLTB und nicht mit anderen Anbietern oder anderen Spielformen vergleichen, etwa auf der Grundlage von Steuerdaten.

***Das gewählte Design als Querschnittsbefragung erlaubt keine Kausalschlüsse; gleichwohl werden diese unzulässigerweise abgeleitet.*** Obwohl zur Erreichung der mit dem Glücksspiel-Survey 2021 angestrebten Ziele eine Längsschnittbefragung erforderlich wäre, wurde lediglich eine Querschnittsbefragung durchgeführt. Damit können grundsätzlich keine kausalen Aussagen getroffen werden<sup>3</sup>, was die Autoren der Survey dennoch tun, beispielsweise mit der Behauptung, „dass es auch unter den Teilnehmer\*innen von Glücksspielen mit geringem Gefährdungspotential einen nennenswerten Anteil von Problemspielenden gibt – auch wenn überwiegend andere Glücksspielformen diese glücksspielbezogenen Probleme **verursacht** haben dürften“ (eigene Hervorhebung). Für diese und ähnliche Aussagen kann allerdings bei einer Querschnittsbefragung prinzipiell keine Evidenz gefunden werden.

Ein Hinweis auf diese Limitation findet sich korrekterweise in früheren (BZgA-)Surveys und wäre nicht zuletzt deshalb bei den Autoren als bekannt vorauszusetzen gewesen. Im Glücksspiel-Survey 2021 wird die mangelnde Aussagekraft hinsichtlich Ursache-Wirkungs-Beziehungen nicht thematisiert. Ein Übergang zu einer Längsschnittbefragung hätte einen wesentlich effektiveren „methodischen Neustart“ dargestellt als der Wechsel zu einer Befragungsmethodik, die nach aktuellem Stand der Wissenschaft eher Probleme schafft als löst. Zudem beschränkt sich der Glücksspiel-Survey auf ein einziges Modell zur Entstehung von Glücksspielproblemen und beleuchtet nicht andere in der Literatur verbreitete Erklärungsansätze.

***Der Schwellenwerte für problematisches Spielverhalten (nach DSM-5) wird abgesenkt,***

---

<sup>3</sup>Zwar existieren methodische Ansätze (*Rubin Causal Model, RCM*), die unter bestimmten Annahmen auch in Querschnittsuntersuchungen Kausalität durch kontrafaktische Analysen nachzuweisen versuchen. Diese Annahmen sind nach aktuellem Forschungsstand für den Glücksspiel-Survey jedoch nicht plausibel zu begründen.

*dieses Vorgehen jedoch nicht hinreichend validiert.* Im Survey wird das Instrument DSM-5 zur Evaluation von Glücksspielbezogenen Problemen der Befragten verwendet. Dieses Diagnosemanual der Amerikanischen Psychiatrischen Gesellschaft ist das führende Klassifikationssystem zur Definition psychischer Erkrankungen. Die bisher im deutschen Sprachraum unter Verwendung der DSM-Kriterien durchgeführten Untersuchungen gingen von problematischem Glücksspielverhalten (präklinische und klinische Störung) aus, wenn mindestens vier der neun DSM-5-Kriterien bzw. drei der zehn DSM-IV-Kriterien erfüllt waren. Der Glücksspiel-Survey reduziert diesen Schwellenwert für die Annahme problematischen Spielverhaltens entgegen allen vergleichbaren vorherigen Untersuchungen von der Erfüllung von mindestens vier auf ein Kriterium, wodurch sich der Anteil der Personen in Deutschland zwischen 18 und 70 Jahren, die der Survey als problematische Spieler einstuft, von 2,3 % auf 8,0 % (auf rund 5 Mio. Menschen) vervielfacht. Die Gruppe der neu als problematisch eingestuften Spieler, die ein bis drei DSM-5-Kriterien erfüllen, bezeichnen die Survey-Autoren als „riskante Spieler“ (5,7 %). Die aus der Literatur entnommene Kategorie des „riskanten Spielers“ ist derzeit Gegenstand wissenschaftlicher Diskussion und wird von den Survey-Autoren der Gruppe der „problematischen Spielern“ zugeordnet, ohne hinreichenden Belege zur Rechtfertigung eines solch niedrigen Schwellenwertes anzuführen. „Risikoreiches“, „auffälliges“ oder „riskantes“ Spielverhalten bei nur wenigen erfüllten Kriterien des jeweils genutzten Diagnosemanuals wurden zwar auch schon bei früheren Befragungen (u.a. dem BZgA-Glücksspiel-Survey) ermittelt und quantifiziert, jedoch nie zuvor dem Katalog des „problematischen“, d.h. in unterschiedlicher Intensität gestörten oder pathologischen, Spielverhaltens zugeordnet. Die Survey-Autoren begründen ihre gegenteilige Vorgehensweise primär damit, dass die geringen individuellen glücksspielinduzierten Kosten von riskanten Spielern zusammengenommen hohe gesellschaftliche Kosten verursachen würden. Dabei berufen sie sich auf eine Studie, die unter anderem Merkmale als gesamtgesellschaftliche Kosten einstuft, bei denen es sich vielmehr um Opportunitätskosten handelt, die so auch bei zahlreichen anderen Freizeitaktivitäten auftreten.

Um die Auswirkung dieser Neukategorisierung auf die Ergebnisse im Sinne einer Sensitivitätsanalyse zu untersuchen, bräuchte es mindestens eine Darstellung der genauen Verteilung der Befragten innerhalb der Kategorien. Eine solche erfolgt jedoch, anders als in vergleichbaren Studien, nicht. Vor diesem Hintergrund muss die Erweiterung der Gesamtkategorie „problematische Spieler“ um die Unterkategorie der „riskanten Spieler“ zwar nicht abgelehnt werden; sie ist aber zumindest diskussionswürdig und bedarf klarer Validierung, ist es doch erst diese Ausdehnung des Begriffs „problematisches Spielverhalten“, welche die im Nachgang der Survey-Veröffentlichung vielfach öffentlich rezipierte Zahl von über fünf Millionen Personen in Deutschland mit einer Glücksspielproblematik erzeugt.

*Eine Evidenz als Grundlage politischer Maßnahmen kann mangels Transparenz und*

*grundsätzlicher methodischer Einschränkungen nicht aus dem Glücksspiel-Survey 2021 abgeleitet werden.* Die Autoren des Glücksspiel-Surveys haben die Anfrage auf Bereitstellung der Daten, des Feldberichts und des Fragebogens abgelehnt. Schriftlich begründet wurde dies mit einer angeblich fehlenden Transparenz der Gutachtererstellerin. Aus diesem Grund lassen sich Berechnungen im Survey nicht überprüfen und die Datenqualität sowie die Qualität der wissenschaftlichen Arbeit der Autoren nicht vollumfänglich einschätzen. Das Verhalten der Autoren geht nicht konform mit dem Kodex „Leitlinien zur Sicherung guter wissenschaftlicher Praxis“ der Deutschen Forschungsgemeinschaft (DFG) (Deutsche Forschungsgemeinschaft 2019), der explizit die Replizierbarkeit von Forschung durch Herausgabe der Forschungsdaten und relevanten Materialien betont sowie die Bedeutung einer kritischen wissenschaftlichen Diskussion über Forschungsergebnisse hervorhebt.

Ein grundsätzlicheres, nicht ohne erhebliche methodische Änderungen lösbares Problem ergibt sich jedoch aus der Anlage des Studiendesigns (Stichprobenauswahl und Querschnittsdesign) des Surveys. Querschnittsbefragungen können bestenfalls, so sie denn repräsentativ sind und geeignete Befragungsinstrumente einsetzen, Aufschluss über die Verbreitung eines Phänomens geben. Ob das geschätzte Ausmaß der Verbreitung, hier: der Prävalenz von Glücksspieltörungen, regulatorischer Maßnahmen bedarf, kann in einer reinen Prävalenzabschätzung nicht beurteilt werden, da Kausalzusammenhänge nicht abgeleitet werden können.

Solche Kausalzusammenhänge betreffen einerseits die von Glücksspieltörungen möglicherweise *verursachten* individuellen oder gesellschaftlichen Folgen – hier geht es um die Frage, welches Risiko ein „riskantes Spielverhalten“ eigentlich darstellt. So kann eine Querschnittsstudie weder nachweisen, dass sich aus „riskantem Spielverhalten“ ein „problematisches“ oder „gestörtes“ entwickelt, noch dass es zu anderen psychischen oder psychosomatischen Problemen führt, beispielsweise Alkoholmissbrauch, noch dass sich gesellschaftliche Folgekosten daraus ergeben. (Es sei damit nicht gesagt, dass solche Entwicklungen nicht stattfinden, aber sie sind mit dem gewählten Design nicht belegbar.)

Andererseits können mögliche Maßnahmen zur Verringerung der Prävalenz nicht hinsichtlich ihrer Wirksamkeit beurteilt und verglichen werden. Auch hierzu bedürfte es eines Studiendesigns, das solche Kausalschlüsse zulässt, beispielsweise durch den Vergleich mit einer Kontrollgruppe. Kurz: Die deskriptive Analyse eines Zustands, zu der auch rein korrelative Analysen zählen, erlaubt keine Schlüsse über die Effektivität möglicher präventive Maßnahmen zur Änderung dieses Zustands.

*Die Rezeption fragwürdiger Forschungsergebnisse in Politik, Verwaltung und Medien legt nahe, dass es dringend funktionierender Qualitätssicherungsmaßnahmen für Studien mit breiter Öffentlichkeitswirksamkeit und/oder politischem Auftrag bedarf.* Die Autoren

haben sich dazu entschieden, im Glücksspiel-Survey auf eine umfangreiche und korrekte Diskussion der Aussagekraft sowie der Limitationen ihrer Arbeit zu verzichten. Allerdings ist eine Darlegung der Einschränkungen und Implikationen insbesondere bei einer an die Öffentlichkeit gerichteten Publikation essenziell. Die Auswirkungen dieser Intransparenz zeigen sich in zahlreichen Rezeptionen, in denen Ergebnisse aus statistischer Sicht falsch oder nur ungenügend wiedergegeben oder fehlinterpretiert werden sowie unzulässig auf einzelne Bundesländer projiziert werden. Unkritische Rezeptionen der fragwürdigen Survey-Ergebnisse finden sich in zahlreichen Bereichen, etwa in den Medien, der Exekutive und in Gerichtsbeschlüssen. Für derartige Fehlinterpretationen sind der Survey bzw. seine Autoren mindestens mittelbar verantwortlich, da eine eingehende Diskussion und Nennung der Limitationen im Survey diesen ebenso hätten entgegenwirken können wie ein Peer-Review der Survey-Ergebnisse sowie deren kritische, interdisziplinäre Diskussion in der Fachliteratur durch Wissenschaftskollegen.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>19</b>
<b>2</b>	<b>Ratio</b>	<b>22</b>
<b>3</b>	<b>Studiendesign als Determinante von Datenqualität</b>	<b>25</b>
<b>4</b>	<b>Total-Survey-Error-Framework</b>	<b>33</b>
4.1	Sampling-Error	35
4.2	Nonsampling-Error	43
4.2.1	Coverage-Error	43
4.2.2	Unit-Nonresponse	45
4.2.3	Item-Nonresponse	49
4.2.4	(Self-)Selection-Error	54
4.2.5	Measurement-Error	59
4.2.6	Processing-Error	61
<b>5</b>	<b>Adjustierung von Verzerrungen</b>	<b>62</b>
<b>6</b>	<b>Weitere methodische Kriterien</b>	<b>71</b>
6.1	Querschnittsbefragung	71

6.2	Kausalität	76
<b>7</b>	<b>Inhaltliche Evaluation</b>	<b>82</b>
7.1	DSM-5 Kriterien	82
7.2	Kategorisierung nach Spielform	89
7.3	Nullpunktmessung	94
7.4	Evaluierung der Legalisierung des Online-Glücksspiels	97
<b>8</b>	<b>Transparenz</b>	<b>99</b>
<b>9</b>	<b>Evidenzbasierte Politik</b>	<b>105</b>
9.1	Grundprinzipien evidenzbasierter Politik	106
9.2	Negativ-Beispiel: Stimmungsbild Sucht	108
9.3	Negativ-Beispiel: Studie zu Lockdown und Glücksspiel	110
<b>10</b>	<b>Rezeption von Survey und Begleitmaterial</b>	<b>112</b>
10.1	Pressemeldung der Studienautoren	113
10.2	Deutscher Lotto- und Totoblock	115
10.3	Weitere Referenzierung des Surveys durch die Autoren	117
10.4	Jahrbuch Sucht 2023	120
10.5	Exekutive	121
10.5.1	Senator für Inneres der Freien Hansestadt Bremen	121
10.5.2	1. Glücksspielsuchtbericht der Thüringer Landesregierung	121
10.5.3	Beauftragter der Bundesregierung für Sucht- und Drogenfragen	124
10.5.4	Bundesministerium für Gesundheit	125
10.5.5	Landesregierung Nordrhein-Westfalen	126
10.6	Rechtsprechung	126

10.7 Medien	127
<b>11 Fazit</b>	<b>128</b>
<b>Bibliografie</b>	<b>132</b>

# 1 | Einleitung

„Glücksspielteilnahme und glücksspielbezogene Probleme in der [deutschen] Bevölkerung“ – die Autoren<sup>1</sup> des Glücksspiel-Surveys 2021 geben bereits in der Überschrift das Hauptziel ihrer Befragung an, nämlich den Anteil der Glücksspieler in Deutschland und insbesondere solcher mit glücksspielbezogenen Problemen zu messen. Die Ergebnisse dieser Messungen wurden in den Medien, aber auch von einer Vielzahl an staatlichen Akteuren breit rezipiert: So begründete der Innensenator für Inneres der Freien Hansestadt Bremen, Ulrich Mäurer, sein robustes Vorgehen gegen Sportwettbüros im Sommer 2022 mit Bedenken wegen Geldwäsche, nannte in diesem Kontext aber auch explizit die Zahlen des Glücksspiel-Surveys 2021 (ZEIT ONLINE 2022). Ebenso zitierte der Beauftragte der Bundesregierung für Sucht- und Drogenfragen, Burkhard Blienert, den Survey bereits mehrmals öffentlich. Aber auch über die Politik hinaus wurden die Daten des Surveys genutzt, so zitierten und interpretierten beispielsweise der Bayerische Verwaltungsgerichtshof und das Obergerverwaltungsgericht Sachsen-Anhalt die Zahlen des Glücksspiel-Surveys 2021 in gerichtlichen Entscheidungen. Damit hat der Survey eines seiner Ziele, nämlich Daten zur Verbesserung von Maßnahmen des Spieler- und Jugendschutzes bereitzustellen, anscheinend erreicht. Doch datenbasierte Entscheidungen sind nicht per Definition bessere Entscheidungen. Die Qualität datenbasierter Entscheidungen hängt maßgeblich von der Qualität der genutzten Daten ab (Münnich 2023).

Für die Qualität von Umfragedaten hat Eurostat (2020) Qualitätskriterien definiert, von denen grundsätzlich die Präzision der Daten als wichtigstes Kriterium angesehen wird (Lohr 2021). Präzision bedeutet in diesem Kontext, dass der auf Basis einer Stichprobe geschätzte Parameter möglichst nah am wahren Parameter der Population liegt. Im Falle des Glücksspiel-Surveys hat es somit höchste Priorität, dass die Schätzungen – zum Beispiel der Anteil an Menschen mit einer Glücksspielstörung – so nah wie möglich an der Realität, also dem wahren Wert des Anteils an Menschen mit einer Glücksspielstörung in der Bevölkerung, liegen (Baker, Brick u. a. 2013). Andere Erwägungen hinsichtlich der Qualität

---

<sup>1</sup>Dr. Sven Buth, Dr. Jens Kalke (ISD, Hamburg); Prof. Dr. Gerhard Meyer (Universität Bremen)

von Umfragedaten sind im Normalfall nachrangig (Lohr 2021). Natürlich gibt es Ausnahmen, so kann beispielsweise in dringenden Fällen auch die Zeitkomponente Vorrang haben. Überzeugende Argumente, die eine niedrigere Priorisierung der Präzisionskomponente beim Glücksspiel-Survey rechtfertigen könnten, sind allerdings, insbesondere angesichts der Ziele des Surveys, nicht ersichtlich.

Das vorliegende Gutachten gliedert sich wie folgt: Zunächst wird im vorliegenden Gutachten die Eignung des neuen *Mixed-Mode-Designs* (Telefon- und Online-Befragung) für die Gewinnung entscheidungsrelevanter Daten diskutiert (Kapitel 3), welches eine Neuerung im Vergleich zu den vorherigen Glücksspiel-Surveys der Bundeszentrale für gesundheitliche Aufklärung (BZgA) darstellt. Im folgenden Kapitel 4 wird der Glücksspiel-Survey 2021 an der Präzision seiner Angaben gemessen, was anhand des Frameworks des *Total-Survey-Error* (Biemer 2010; Lohr 2021) geschieht. Das Framework untersucht zwei Hauptkomponenten, den *Sampling-Error* (der Fehler, der sich beim Ziehen der Stichprobe ereignet) und den *Nonsampling-Error* (der Fehler, der sich außerhalb der Stichprobenziehung ereignet). Unter den *Nonsampling-Error* fallen weiterhin der *Coverage-Error*, der *Nonresponse-Error*, der *(Self-)Selection-Error*, der *Measurement-Error* und der *Processing-Error*. Befragungen geben allerdings häufig nur den *Sampling-Error* an, beispielsweise durch Konfidenzintervalle. Damit wird lediglich eine von mehreren möglichen Fehlerquellen quantifiziert und kommuniziert, die verschiedenen *Nonsampling-Errors* bleiben – trotz häufig viel größerer Ausmaße und schwerwiegender Konsequenzen (Lohr 2021) – oft unbeachtet.

Weiterhin werden die von den Studienautoren thematisierte Gewichtung (Kapitel 5) und die methodische Eignung des Surveys zum Treffen kausaler Aussagen, insbesondere unter Berücksichtigung des gewählten Querschnittsdesigns, beurteilt (Kapitel 6). Ergänzend werden in diesem Gutachten noch weitere durch die Ziele des Surveys induzierte Aspekte betrachtet: Zum einen wird die Eignung des Surveys als „Nullpunktmessung“ und Evaluation des Glücksspielstaatsvertrags 2021 untersucht, zum anderen werden das gewählte Messinstrument DSM-5, die Kategorisierung der verschiedenen Spielformen und die Methode der Querschnittsbefragung kritisch gewürdigt (Kapitel 7).

Im Anschluss wird die Transparenz des Glücksspiel-Surveys 2021 sowie der begleitenden Wissenschaftskommunikation anhand anerkannter wissenschaftlicher Grundprinzipien (Deutsche Forschungsgemeinschaft 2019) untersucht und bewertet (Kapitel 8). Wie Münnich (2023) klarstellte, ist Transparenz ein entscheidendes Kriterium zur Bewertung der Datenqualität, da nur diese eine Bewertung überhaupt erst ermöglicht. Ausführungen zu den Grundprinzipien evidenzbasierter Politik folgen in Kapitel 9. Anschließend werden die Rezeption des Surveys und die eventuelle Verantwortlichkeit der Studienautoren für etwaige Fehl- und Falschinterpretationen diskutiert (Kapitel 10). Schließlich wird der

Glücksspiel-Survey 2021 auf Basis der zuvor evaluierten Aspekte in seiner Gesamtheit bewertet und Empfehlungen für die Nutzung des Surveys werden diskutiert (Kapitel 11).

## 2 | Ratio

### Executive Summary

*Die Ziele des Surveys lassen sich in zwei Hauptziele unterteilen: Einerseits wollen die Autoren neue Erkenntnisse über verschiedene Aspekte des Glücksspiels und der Glücksspielnutzung in Deutschland gewinnen, andererseits sollen verschiedene glücksspielgesetzliche und regulatorische Bestimmungen evaluiert werden, um (gegebenenfalls) eine Datenbasis für deren Verbesserung – die notwendigerweise durch staatliche Stellen erfolgen würde – bereitzustellen. Aufgrund des regulatorischen Aspekts muss der Glücksspiel-Survey 2021 demnach an höchsten Qualitätskriterien gemessen werden, die für eine Erhebung mit dem Ziel der Information oder Beratung von staatlichen Stellen gelten. Die breite Rezeption des Surveys in Exekutive, Rechtsprechung und Medien stützt diese Sichtweise. Demnach haben die Qualität und Präzision der Daten höchste Priorität.*

Zu Beginn sollen die Ziele des Glücksspiel-Surveys 2021 betrachtet werden, denn die zu erreichenden Ziele determinieren die Eignung des gewählten Studiendesigns und der Analysemethoden. Wie Münnich (2023) unlängst feststellte, ist die *Ratio* eines Surveys ausschlaggebend für seine Bewertung und Anwendung verschiedener Qualitätsmaßstäbe. Soll beispielsweise mit einem Survey ein aktuelles Stimmungsbild in der Bevölkerung erstellt werden, haben Schnelligkeit der Bereitstellung und niedrige Kosten verständlicherweise Vorrang vor dem Ziel maximal akkurater Daten<sup>1</sup> – qualitativ hochwertige und akkurate Daten kosten (sehr) viel Zeit und Geld (Münnich 2023).

Wird ein Survey mit dem Ziel der Information oder Beratung von Exekutive, Legislative oder Rechtsprechung durchgeführt, muss die Datenqualität höchste Priorität und somit Vorrang vor Ressour-

<sup>1</sup>Dass auch in Ausnahmesituationen, wie der Covid-19 Pandemie, Daten von hoher Qualität schnell bereitgestellt werden können, hat das U.S. Census Bureau mit der Household Pulse Survey gezeigt.

cenerwägungen haben<sup>2</sup> (Münnich 2023). Dies setzt zwingend voraus, dass das gewählte Design einer Studie, die Erhebungsmethodik und die Dokumentation eine derartige Bewertung überhaupt ermöglichen. Insbesondere bei den immer häufiger online durchgeführten Befragungen bereitet dies große Schwierigkeiten (Münnich 2023). Mit einer hohen Datenqualität müssen Transparenz und Unabhängigkeit bei der Datenerhebung, -analyse und -kommunikation einhergehen. Denn ohne Transparenz kann die Qualität der Daten nicht von Dritten überprüft werden, wodurch nicht nur Bürger und Medien, sondern auch staatliche Institutionen blind auf den Bereitsteller der Daten vertrauen müssten. Wenn also ein Parlament oder ein Gericht eine Entscheidung (unter anderem) mit (Umfrage-)Daten begründet, muss garantiert werden, dass die jeweilige Entscheidung nicht durch eine schlechte Datenqualität oder bewusste Datenmanipulation beeinflusst wird. Nicht umsonst haben sich die europäischen Statistikämter, deren Daten regelmäßig von der Politik genutzt werden, hohe Qualitäts-, Transparenz- und Unabhängigkeitsstandards auferlegt (Ausschuss für das Europäische Statistische System 2017)<sup>3</sup> Derlei Richtlinien werden für die Wissenschaft in Zeiten vermehrt geforderter evidenz- und datenbasierter Politik immer wichtiger.

Der Glücksspiel-Survey 2021 hat nach Aussage der Studienautoren, folgende Ziele:

1. die epidemiologischen Kenntnisse über die Glücksspielteilnahmen und -probleme verbessern; (Buth, Meyer und Kalke 2022, S. 4);
2. darauf aufbauend Maßnahmen des Spieler- und Jugendschutzes evaluieren (Buth, Meyer und Kalke 2022, S. 4), denn „[a]uf der Grundlage solcher – in zweijährigen Abständen erhobenen – Daten können Maßnahmen des Spieler- und Jugendschutzes evaluiert und gegebenenfalls verbessert werden.“ (Buth, Meyer und Kalke 2022, S. 4). Zudem soll der Survey eine empirische Grundlage für eine Diskussion über solche Maßnahmen schaffen (Buth, Meyer und Kalke 2022, S. 7);
3. einen „soliden Ausgangspunkt“ bereitzustellen, um die Auswirkungen des seit dem 1. Juli 2021 geltenden Glücksspielstaatsvertrags zu evaluieren (Buth, Meyer und Kalke 2022, S. 4);
4. Zusammenhänge zwischen Glücksspielproblemen, psychischer Belastung und Alkoholkonsum zu untersuchen (Buth, Meyer und Kalke 2022, S. 8);

---

<sup>2</sup>Zu den Merkmalen von Datenqualität im Sinne einer akkuraten Beschreibung der Realität durch die Daten zählen Korrektheit („Correctness“), Genauigkeit („Accuracy“), Vollständigkeit („Completeness“), Konsistenz („Consistency“) und Aktualität („Currentness“). Die Frühzeitigkeit der Bereitstellung („Punctuality“) wird als Datenqualität im weiteren Sinne, d.h. im Sinne der Nutzbarkeit („Usability“) der Information verstanden. Wenn im vorliegenden Gutachten von Datenqualität gesprochen wird, ist stets Datenqualität im engeren Sinne der fünf erstgenannten Kriterien gemeint.

<sup>3</sup>Das Europäische Statistische System ist ein Netzwerk von Behörden, das offizielle Statistiken auf EU-Ebene erstellt. Es besteht aus Eurostat (dem statistischen Amt der Europäischen Union) und den nationalen statistischen Instituten und anderen Behörden.

5. die Bekanntheit und Akzeptanz von einzelnen Maßnahmen des Jugend- und Spielerschutzes erheben (Buth, Meyer und Kalke 2022, S. 8);
6. den Einfluss von glücksspielbezogener Werbung (bzw. deren Wahrnehmung) auf das Spielverhalten zu untersuchen (Buth, Meyer und Kalke 2022, S. 8).

Die Ziele des Surveys lassen sich damit unter zwei Hauptziele subsumieren: Einerseits wollen die Autoren neue Erkenntnisse über verschiedene Aspekte des Glücksspiels und der Glücksspielteilnahme in Deutschland gewinnen, andererseits sollen verschiedene glücksspielgesetzliche und -regulatorische Bestimmungen evaluiert werden, um (gegebenenfalls) eine Datenbasis für deren Verbesserung bereitzustellen. Den obigen Ausführungen folgend, muss der Glücksspiel-Survey 2021 an den hohen Qualitätskriterien gemessen werden, die für eine Erhebung mit dem Ziel der Information oder Beratung von staatlichen Akteuren gelten, da der Survey unter anderem Daten als Grundlage für Gesetzgebungsprozesse zur Verfügung stellen will (Münnich 2023). Die breite Rezeption des Survey in Exekutive, Rechtsprechung und Medien (siehe Kapitel 10 Rezeption) stützt diese Sichtweise. Damit haben die Qualität und Präzision der Daten höchste Priorität. Entsprechend wird der Glücksspiel-Survey 2021 im vorliegenden Gutachten auch an diesen hohen Qualitätsmaßstäben gemessen.

### 3 | Studiendesign als Determinante von Datenqualität

#### Executive Summary

*Der Glückspiel-Survey 2021 will einen methodischen Neustart durchführen, indem ergänzend zur telefonischen Befragung aus einem Online-Panel rekrutiert wird (Mixed-Mode-Design mit nichtprobabilistischer Komponente). Die Begründung überzeugt jedoch nicht:*

- Erstens ist das Nonresponse-Problem nicht, wie von den Autoren suggeriert, auf Telefon-Umfragen beschränkt, sondern betrifft auch Online-Umfragen. Im Survey ist die Nonresponse-Quote der nichtprobabilistischen Online-Umfrage, die zusätzlich unter (Self-)Selection-Error leidet, mit 90 % deutlich höher als die der Telefon-Umfrage (73 %.)*
- Zweitens zitieren die Autoren selektiv Passagen aus der Literatur bezüglich der Vorteile von Online-Befragungen, obwohl dieselbe Literatur eindeutig von Online-Umfragen in Befragungen mit dem Ziel der präzisen Erfassung von Bevölkerungsanteilen abrät.*
- Drittens berufen sich die Survey-Autoren auf eine veraltete Publikation zur Herleitung des Mixed-Mode-Designs, deren Ergebnisse in einer späteren, diesmal peer-reviewten Version explizit gegen die Inklusion eines Online-Panels sprechen. Deren Zitierung an anderer Stelle des Surveys legt nahe, dass die Autoren die neuen Erkenntnisse kannten, sie sich gleichwohl für die Rezeption eines überholten Forschungsstands entschieden haben.*

*Durch den Umstieg sind nach wissenschaftlichem Stand eine Verschlechterung der Datenqualität und eine Überschätzung der Zahl von Menschen mit Glücksspielproblemen anzunehmen.*

Es ist weithin bekannt, dass belastbare und aussagekräftige Daten nur aus sorgfältig geplanten und qualitätsgesicherten Erhebungen gewonnen werden können. Das Studiendesign ist deshalb die erste und wesentliche Determinante von Datenqualität und maßgeblich dafür, ob eine Studie überhaupt zur Schaffung von Evidenz für politische Entscheidungen beitragen kann – oder lediglich „billige Schlagzeilen“ oder gar unseriösen „Datenschrott“ produziert, wie es der Sozialforscher Rainer Schnell erst kürzlich in der Wochenzeitung *DIE ZEIT* ausdrückte (Schnell 2023).

Bevor die im Glücksspiel-Survey 2021 verwendeten Methoden genauer evaluiert werden, ist es deshalb angebracht, sich zuerst mit dem Grunddesign des Surveys, dem *Mixed-Mode-Design*, und dessen Begründung auseinanderzusetzen. Ein *Mixed-Mode Design* bedeutet im Falle des Glücksspiel-Surveys 2021, dass die Befragten teils telefonisch (via *Probability-Sampling*) und teils online (via *Nonprobability-Sampling*) befragt wurden<sup>1</sup>. Beim *Probability-Sampling* hat jedes Mitglied einer Population eine bekannte, positive Wahrscheinlichkeit, befragt zu werden, was eine theoretisch solide Verallgemeinerung der Stichprobe auf die Population ermöglicht (Lohr 2021). Beim *Nonprobability-Sampling* haben nicht alle Mitglieder der Population eine bekannte und/oder positive Wahrscheinlichkeit, befragt zu werden, da beispielsweise bei einer Online-Befragung Menschen ohne Internetanschluss keine Chance haben, an der Befragung teilzunehmen (Lohr 2021).

Die Autoren des Surveys verweisen explizit darauf, dass es sich beim Glücksspiel-Survey 2021, „um keine einfache Fortschreibung des bisherigen Monitorings der BZgA [handelt]. Mit den Veränderungen bei der Erhebungsmethodik (kombinierte Telefon- und Online Befragung) und den -instrumenten (DSM-5) ist vielmehr ein methodischer Neustart verbunden, bei dem die oben genannten Zielsetzungen explizit Berücksichtigung fanden.“ (Buth, Meyer und Kalke 2022, S. 8 f.). Die früheren Surveys der BZgA wurden ab 2013 mit dem *Dual-Frame-Design* durchgeführt, es wurden also (via *Probability-Sampling*) Festnetz- und Mobiltelefon-Anschlüsse angerufen.

Die Studienautoren beschäftigen sich zur Herleitung des „methodischen Neustarts“ zuerst mit den sinkenden Antwortraten bei Telefonbefragungen. Sie nennen als eine Erklärung „die in der Öffentlichkeit geführte Debatte um unerlaubte Werbeanrufe“ (Buth, Meyer und Kalke 2022, S. 11) und verweisen als Quelle auf die Publikation *Junges altes Sachsen-Anhalt: ein Bundesland im Blick der Sozialforschung* von 2011 (Buchwald 2011). Weiterhin sei problematisch, dass „in der Bevölkerung zwischen Direktmarketing und Markt- bzw. Sozialforschung kaum unterschieden wird bzw. nicht unterschieden werden kann. Im Regelfall wird den Interviewer\*innen auch keine Zeit eingeräumt, den Zweck der Studie zu erläutern bzw. die Personen von der Seriosität des Anliegens oder des Institutes zu überzeugen.“

<sup>1</sup>Grundsätzlich können auch Online-Befragungen probabilistisch durchgeführt werden, was die Qualität der Ergebnisse im Allgemeinen erhöht.

(Buth, Meyer und Kalke 2022, S. 11), die Autoren nennen hierzu aber keine Referenz. Die sinkende telefonische Erreichbarkeit betreffe zudem insbesondere „*junge, technikaffine Menschen, die statt eines Festnetztelefons vorrangig oder gar ausschließlich ihr Smartphone zur Kommunikation nutzen*“ (Buth, Meyer und Kalke 2022, S. 11). Dem werde mit dem *Dual-Frame-Ansatz*, also der Inklusion von Mobilfunknummern in die Stichprobe, zwar entgegengewirkt, doch diese Maßnahme führe „*nur bedingt zum Erfolg [...], da Anrufe unbekannter Herkunft von Smartphonenu\*innen häufig mit Hilfe spezifischer Apps per se geblockt werden (Dutwin et al., 2018; Marken, 2018)*“ (Buth, Meyer und Kalke 2022, S. 11).

Als Beleg für diese Aussage führen sie zwei Quellen an, die beide auf einen Report der *American Association for Public Opinion Research* zu *Spam Flagging* und *Call Blocking* verweisen, der dieses Phänomen allerdings nur für die USA und nicht für Deutschland untersucht (Dutwin, Blum u. a. 2018). Abschließend schlussfolgern die Autoren, dass „*die aufgezeigten Probleme in den zurückliegenden Jahren zu stetig abnehmenden Antwortraten [führen]*“ (Buth, Meyer und Kalke 2022, S. 11) und verweisen dabei auf den Report von Dutwin und Lavrakas (2016), der diesbezüglich Daten von Befragungsunternehmen aus den USA untersucht. Aus diesen Entwicklungen ergäben sich höhere Kosten für die Befragungsinstitute und das Risiko für eine „*Erhöhung des sogenannten Selektionsfehlers – gemeint ist damit die systematische Über- bzw. Unterrekutierung spezifischer Teile der Bevölkerung*“<sup>2</sup> (Buth, Meyer und Kalke 2022, S. 11) würde steigen. Als Beleg für das höhere Risiko eines Coverage-Fehlers führen die Autoren des Surveys das Forschungspapier *Solving the Nonresponse Problem With Sample Matching?* von Bethlehem (2016) an. In dem Artikel wird *Sample Matching* als potenzielle Verbesserung im Vergleich zu *Nonprobability Surveys* oder *Probability Surveys* mit niedriger Responserate vorgeschlagen. Abschließend stellen die Autoren des Glücksspiel-Surveys fest, dass bei Telefonbefragungen „*aufgrund des beschriebenen Coverage-Fehlers die Qualität der generierten Datensätze [sinkt], da die daraus abgeleiteten Befunde oftmals nur noch eingeschränkt auf die Bevölkerung übertragbar sind.*“ (Buth, Meyer und Kalke 2022, S. 12).

Die Analyse der Autoren des Glücksspiel-Surveys ist zwar dem Grunde nach nicht falsch, im Ergebnis allerdings jedoch zu undifferenziert: Der *Nonresponse-Error*, also eine systematische Verzerrung der erhobenen Daten aufgrund eines strukturellen Unterschieds zwischen Antwortenden und Antwortverweigerern, kann ein großes Problem bei telefonischen Befragungen sein, dessen Ausmaß von der individuellen Befragung abhängt. Insbesondere sensible Befragungsthemen erhöhen das Risiko für

<sup>2</sup>In der einschlägigen Literatur wird der hier beschriebene Fehler, der resultiert, wenn die in einer Befragung erreichbare Population nicht der Grundgesamtheit entspricht, üblicherweise als Coverage-Fehler bezeichnet. Der Selektionsfehler umfasst jedoch im Allgemeinen zwei weitere Komponenten: den Nonresponse-Fehler durch ein unterschiedliches Antwortverhalten von Antwortenden und Nichtantwortern, und den Selbstselektions-Fehler (bei Nonprobability-Samples), wenn eine Person selbst entscheiden kann, ob sie an einer Befragung teilnimmt.

einen Nonresponse-Error. Dies gilt ebenso für Online-Befragungen, die allerdings (*ceteris paribus*) mindestens so stark wie telefonische Befragungen von einem *Selection-Error* betroffen sind (Lohr 2021). Die Online-Befragung des Glücksspiel-Surveys hat zum Beispiel eine Nonresponse-Quote von 90 %, während die telefonisch durchgeführte Befragung eine niedrigere (und damit günstigere) *Nonresponse-Quote* von 73 % hat.

Weiterhin konstatieren die Autoren, dass die Bedeutung von Online-Befragungen immer mehr gestiegen sei, stellen bezüglich von offenen Befragungen im Internet aber fest: *„Ein gravierendes Problem ist hierbei jedoch die Selbstselektion der Interviewten. Da das Thema solcher Befragungen meist schon bei der Rekrutierung benannt wird, nehmen überproportional viele Themeninteressierte daran teil. Für Studien, die den Anspruch erheben, zumindest annähernd repräsentativ zu sein, ist diese Erhebungsform schon allein aus diesem Grund nicht geeignet.“* (Buth, Meyer und Kalke 2022, S. 12). Dies sei in Online-Panels anders, da die Teilnahme auch *„zufallsbasiert – die Studienthematik wird in diesem Fall bei der Einladung zum Interview nicht bekannt gegeben – erfolgen [kann].“*. Gleichwohl geben sie zu bedenken: *„Trotz der Vorteile der Online-Access-Panels gegenüber Onlinebefragungen mit selbstrekrutierten Teilnehmer\*innen bleibt die Frage, in welchem Maße sich damit die Bevölkerung adäquat abbilden lässt (Baker et al., 2010). Schließlich handelt es sich hierbei um eine Gruppe von Personen mit spezifischen Eigenschaften – Internetanschluss vorhanden, eher onlineaffin und bereit, gegen eine eher geringe Aufwandsentschädigung an thematisch sehr unterschiedlichen Befragungen teilzunehmen. Pauschal wird sich diese Frage nicht beantworten lassen, da anzunehmen ist, dass die Übereinstimmung von Access-Panel und Gesamtbevölkerung je nach untersuchter Thematik variiert“* (Buth, Meyer und Kalke 2022, S. 12).

Bezüglich der diskutierten Befragungen via Online-Panels kann den obigen Aussagen nicht zugestimmt werden. Die Teilnehmer eines *Nonprobability*-Onlinepanels werden eben gerade nicht zufallsbasiert ausgewählt. Da alle Panelteilnehmer sich freiwillig für eine Teilnahme am Panel entschieden haben, wird eine (zufällige) Stichprobe des Panels immer eine Stichprobe von (nicht zufälligen) Freiwilligen bleiben – unabhängig von der Art der Stichprobenziehung (Lohr 2021). Zudem bleibt, anders als von den Autoren behauptet, *„die Frage, in welchem Maße sich damit die Bevölkerung adäquat abbilden lässt (Baker et al., 2010)“*, gerade nicht offen. Baker, Blumberg u. a. (2010) beantworten diese Frage in dem von den Autoren des Surveys herangezogenen Report deutlich: *Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values. There currently is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. Thus, claims of “representativeness“ should be avoided when using these sample sources.* (S. 758). Es bleibt

vielmehr fraglich, worauf sich die Autoren des Glücksspiel-Surveys in dem zitierten Report beziehen, da dieser ausdrücklich davon abrät, *Nonprobability Samples* für „repräsentative“ Inferenzen auf die Gesamtpopulation zu nutzen (Baker, Blumberg u. a. 2010).

Zudem diskutieren die Autoren einerseits Forschungsartikel bezüglich der Nachteile von Online-Studien (Pickering und Blaszczyński 2021) und verweisen andererseits auf Russell u. a. (2022), die „auf die Vorteile solcher Erhebungen [verweisen]“ (Buth, Meyer und Kalke 2022, S. 13). Dabei erwähnen die Autoren des Surveys jedoch nicht, dass Russell u. a. (2022) gemäß den Zielen eines Surveys klar differenzieren: „*online panels do offer important benefits, such as targeting respondents with specific characteristics*“ (S. 109), aber „*online panel data should not be used to generate estimates of population prevalence, with or without weighting, since online panel participants cannot be assumed to be representative of all people in the general population (but instead only ‘some’ people from that larger population).*“ (S. 105). Konkret verweisen die Autoren des Surveys also selektiv auf die in Russell u. a. (2022) genannten Vorteile von Online-Panels, ignorieren allerdings deren im selben Artikel festgestellte mangelnde Eignung, wenn wie im Glücksspiel-Survey 2021 Anteilswerte in der Bevölkerung gemessen werden sollen.

Anschließend stellen die Autoren des Surveys mit Verweis auf Sturgis und Kuha (2022) fest, dass „*die ermittelten Prävalenzwerte für die Spielteilnahme und die glücksspielbezogenen Probleme in den Onlinestudien jeweils deutlich über den Werten der Befragungen mit persönlichem Kontakt lagen*“. Dabei gehen die Autoren nicht auf mögliche Erklärungen für dieses Phänomen ein. Sturgis und Kuha (2022) stellen nämlich im Vergleich von acht Befragungen in Großbritannien fest: „*Our findings show that surveys conducted online produce substantially higher estimates of problem gambling compared with in-person interview surveys. This is because online surveys, whether using probability or non-probability sampling, overrepresent people who are more likely to gamble online and to gamble frequently, relative to the proportions of these groups in the general population*“ (S. 63). Die Information, dass Online-Befragungen aufgrund einer Selektionsverzerrung höhere Glücksspiel(störungs)prävalenzen haben, ist essenziell, um das beschriebene Phänomen zu verstehen. Dies ist für Rezipienten des Glücksspiel-Surveys 2021 nicht möglich, da entsprechende Informationen fehlen. Damit bleibt für den Leser des Glücksspiel-Surveys 2021 zunächst offen, ob persönliche oder Internet-Befragungen den „wahren“ Anteil an Menschen mit Glücksspielstörung wiedergeben – dabei beantworten Sturgis und Kuha (2022) diese Frage klar zugunsten der persönlichen Befragungen.

Abschließend konstatieren die Autoren des Glücksspiel-Surveys mit Verweis auf einen Report von Sturgis (2020): „*Stellt man die dargestellten Defizite der verschiedenen Erhebungsmethoden und de-*

ren Folgen gegenüber, so fällt auf, dass sie jeweils entgegengesetzt wirken (Sturgis, 2020). Bestimmte Bevölkerungsgruppen, die mittels telefonischer Befragungen nur schwer erreicht werden, sind in Onlineerhebungen häufig überproportional vertreten“. Dies schreibt Sturgis (2020) zwar in seinem Bericht (der nicht in einem peer-reviewten Journal veröffentlicht wurde), allerdings ist dieser Report (mutmaßlich) eine Vorab-Version des später in einem peer-reviewten Journal veröffentlichten Artikels (Sturgis und Kuha 2022), der auch von den Autoren des Glücksspiel-Surveys zitiert wird. Dort vergleicht Sturgis acht Surveys – im Vergleich zu zwei Surveys im alten Report – und kommt (wie bereits oben ausgeführt) zu dem Schluss, dass die höheren Prävalenzen in Online-Studien auf eine Selektionsverzerrung zurückzuführen sind (Sturgis und Kuha 2022). Der Unterschied zwischen beiden Reports wird insbesondere bei den festgestellten Fehlerquellen deutlich: Während Sturgis im alten Report noch schrieb, „the 2016 combined health surveys may somewhat underestimate the true prevalence of problem gambling as a result of under-coverage and nonresponse amongst groups with higher rates of problem gambling compared to the general population. Additionally, it seems likely that there may be some degree of downward bias in the combined health survey estimates due to measurement error, with socially desirable responding and the content of the questions preceding the PGSI having a downward effect on the prevalence reports amongst some respondents.“ (S. 20), stellt er im aktuelleren peer-reviewten Artikel fest: „Other potential causes of the differences, including true change in harmful gambling, sampling variability, coverage error, and differential measurement error, seem unlikely to exert a notable influence.“ (S. 68).

Betrachtet man also den Report von 2020 und den veröffentlichten Artikel von 2022, korrigiert letzterer die Grundaussagen des Reports von 2020. Auch die Aussage, dass sich der wahre Wert der Glücksspiel(störungs)prävalenz zwischen der in persönlichen und Online-Befragungen festgestellten Messung befindet, trifft Sturgis im Artikel von 2022 nicht mehr. All diese, für das Design des Glücksspiel-Surveys 2021 und dessen Bewertung essenziellen, Informationen finden sich jedoch nicht im Survey-Bericht wieder. Das ist von nicht zu unterschätzender Bedeutung: Denn auf der überholten, von Sturgis selbst korrigierten These, dass sich der wahre Wert zwischen den Ergebnissen beider Befragungsarten befindet (wenn auch näher an dem Messpunkt der persönlichen Befragung), baut die Begründung der Autoren für eine *Mixed-Mode*-Befragung auf. Berücksichtigt man den aktuelleren (und im Gegensatz zum früheren Report durch ein Peer-Review-Verfahren qualitätsgesicherten) Artikel von Sturgis aus 2022, ist die Entscheidung für eine *Mixed-Mode* Befragung nicht mehr durch den Stand der Wissenschaft begründbar.

Wieso die Autoren sich dazu entschieden haben, den veralteten Report von 2020 und nicht den aktuelleren Artikel von 2022 als Begründung für ihr Design heranzuführen, obwohl ihnen der aktuellere

Artikel offensichtlich (Zitation auf auf S. 12) bekannt war, bleibt offen. Zusammenfassend lässt sich jedenfalls feststellen, dass man bei der Rezeption von Sturgis Artikel aus 2022 zum eindeutigen Ergebnis kommt, dass eine Online-Befragung – wie sie im Glücksspiel-Survey teilweise durchgeführt wurde – kein geeignetes Mittel zur Erfassung der Prävalenz von Glücksspiel und Glücksspieltörung in der Bevölkerung ist<sup>3</sup> (Sturgis und Kuha 2022).

Als weiteres Argument für die Nutzung eines *Mixed-Mode-Designs* verweisen die Autoren des Surveys auf die Wahlforschung, „So umfassen die vom Meinungsforschungsinstitut infratest dimap regelmäßig durchgeführten Befragungen zu den aktuellen Wahlabsichten der Bevölkerung (Sonntagsfrage, ARD DeutschlandTrend) zu zwei Dritteln telefonisch Befragte und zu einem Drittel Teilnehmer\*innen eines Onlinepanels (infratest dimap, 2022).“ (Buth, Meyer und Kalke 2022, S. 13). Allerdings lässt sich dieses Beispiel aus dem Bereich der Wahlprognosen nicht unmittelbar auf den Glücksspielbereich übertragen, da *Nonprobability-Sampling* via Online-Befragungen einzig im Bereich Wahlprognosen bessere Ergebnisse liefert als *Probability-Sampling* (Baker, Brick u. a. 2013). Sturgis und Kuha (2022) lieferten – wie bereits erläutert – sogar Evidenz, die für eine niedrigere Präzision von Online-Befragungen im Glücksspielbereich spricht.

Die Begründung zur Nutzung eines *Mixed-Mode-Designs* schließen die Autoren folgendermaßen ab: „Mit Blick auf die oben beschriebenen Defizite rein telefonischer Befragungen und den Vorteilen, die sich durch eine ergänzende Online-Erhebung ergeben, ist auch im Rahmen des Glücksspiel-Surveys 2021 ein solcher *Mixed-Mode-Ansatz* gewählt worden. Die Erweiterung des bisherigen ausschließlich telefonischen Erhebungsdesigns um eine Onlinebefragung wird nach unserer Auffassung dazu beitragen, die Qualität der erhobenen Daten und somit die Vergleichbarkeit der Stichprobe mit der Gesamtbevölkerung evident zu verbessern.“ (Buth, Meyer und Kalke 2022, S. 13). Diese Schlussfolgerung ist, wie bereits ausführlich dargelegt, nach dem aktuellen Stand der Forschung nicht gerechtfertigt. Die Defizite rein telefonischer Befragungen treffen ebenso (oder noch stärker) auf *Nonprobability-Sampling* Online-Panels zu. Dies lässt sich beispielsweise an den *Nonresponse-Quoten* der Glücksspiel-Surveys 2021 zeigen: Wurden telefonische Befragungen von den Autoren insbesondere aufgrund ihrer hohen *Nonresponse-Quoten* kritisiert, so hat die Online-Befragung des Surveys eine noch viel höhere (und insgesamt sehr hohe) *Nonresponse-Quote* von 90 %. Diesbezüglich lässt der Glücksspiel-Survey auch eine kritische Selbstreflexion vermissen: Die hohe *Nonresponse-Quote* bei Telefonbefragungen ist der Hauptgrund für die Erweiterung um eine Online-Befragung, schlussendlich hat die Online-Befragung jedoch eine noch höhere *Nonresponse-Quote* von 90 % und die telefonische Befragung eine geringe-

<sup>3</sup>Unabhängig davon diskutieren die Autoren nicht, ob und inwieweit Sturgis 2020 Ergebnisse aus Großbritannien inhaltlich uneingeschränkt auf Deutschland übertragbar sind.

---

re Nonresponse-Quote von 73 %. Eine kritische Reflexion der eigenen Annahmen und Argumente in Anbetracht dieser widersprüchlichen Ergebnisse findet im Survey nicht statt.

Mehr noch, die von den Autoren angestrebte Qualitätsverbesserung der Daten durch die Nutzung eines *Nonprobability-Sampling*-Online-Panels führt nach aktuellem wissenschaftlichen Stand (Sturgis und Kuha 2022; Baker, Brick u. a. 2013; Baker, Blumberg u. a. 2010) sogar zu einer Verschlechterung der Datenqualität. Als Hauptargument für *Nonprobability-Sampling* werden in der Literatur die niedrigeren Kosten angesehen (Lohr 2021), was in Anbetracht der Ziele des Glücksspiel-Surveys allerdings ein nachrangiges Argument sein muss (Münnich 2023). Abgesehen von den ungerechtfertigten Schlussfolgerungen ist zudem problematisch, dass selektiv aus der Literatur zitiert wurde, was dem fachfremden Leser die Bildung einer eigenen fundierten Meinung unmöglich macht. Schlussendlich ist es aus methodischer Sicht ungenau, dass die Autoren nur den Unterschied zwischen Telefon- und Online-Befragungen diskutieren, obwohl damit im Falle des Glücksspiel-Surveys 2021 ein (teilweiser) Wechsel vom *Probability-Sampling* zum *Nonprobability-Sampling* stattfindet. Die erheblichen Konsequenzen dieses Wechsels hinsichtlich der Datenqualität werden im Survey nicht angemessen diskutiert und erläutert.

Bei der Durchführung des Surveys wurden 40 % der Stichprobe durch eine Online-Befragung generiert, die später auf einen Anteil von 33 % gewichtet wurden. Um die Auswirkungen der Verzerrung der Online-Stichprobe auf die Gesamtstichprobe zu erfassen, sind Informationen über die Verteilungen der wichtigsten Variablen in den jeweiligen Befragungsarten sowie eine Analyse möglicher Verzerrungen durch die Befragungsart notwendig. Da die Autoren des Glücksspiel-Surveys diese Informationen dem Leser nicht transparent zur Verfügung stellen, kann über das Ausmaß der Verzerrung nur spekuliert werden. Dadurch kann die Datenqualität nicht – wie es bei den selbst gesetzten Zielen unerlässlich wäre – evaluiert werden, was bereits zu einer starken Minderung der Qualität der Daten führt (Münnich 2023). Festzustellen ist, dass die Begründung der Autoren für ein *Mixed-Mode*-Design weder überzeugt noch dem aktuellsten Stand der Forschung entspricht und der Glücksspiel-Survey ohne das *Nonprobability-Sampling* bei der Online-Befragung präzisere Daten geliefert hätte (Lohr 2021; Baker, Blumberg u. a. 2010; Baker, Brick u. a. 2013).

## 4 | Total-Survey-Error-Framework

### Executive Summary

*Der Glücksspiel-Survey 2021 wird entsprechend der einschlägigen Literatur an der Präzision seiner Ergebnisse gemessen. Das bedeutet konkret, dass Schätzungen aus dem Survey – z.B. der Anteil der Menschen mit Glücksspielstörung – so nah wie möglich an den wahren Anteilen der Grundgesamtheit liegen sollten. Die Evaluation des Surveys wird in diesem Gutachten anhand des Total-Survey-Error-Frameworks durchgeführt. Dies ist allerdings nur sehr eingeschränkt möglich, da die Autoren des Glücksspiel-Surveys 2021 die dafür benötigten Materialien – entgegen wissenschaftlichen Qualitätsstandards – den Gutachtern nicht zur Verfügung gestellt haben. Diese mangelnde Transparenz erschwert die Abschätzungen potenzieller Verzerrungen aus verschiedenen Quellen, da über die klassischen Fehlerquellen in Befragungen im Glücksspiel-Survey 2021 sehr wenig bis gar nicht berichtet wird. Immerhin reichen die Angaben aus, um mehrere potenziell schwerwiegende Probleme bzw. Fehler des Surveys identifizieren zu können: Zum einen wird die Unsicherheit, die mit jeder Stichprobe verbunden ist, nur äußerst selektiv und zu optimistisch kommuniziert, was das Risiko für Fehlinterpretationen erhöht und offenkundig zu undifferenzierten Aussagen im Survey führte. Zum anderen ist aufgrund aktueller wissenschaftlicher Literatur von einem großen (Self-)Selection-Error in der Online-Stichprobe des Surveys auszugehen, was fälschlicherweise zu einer höheren Prävalenz von Glücksspielproblemen führen wird. Zusätzlich ist aufgrund der niedrigen Ausschöpfungsquote der Befragung (Telefon: 27 %, Online: 10 %) und des Entfernens von unvollständigen Interviews von einem erheblichen Nonresponse-Error auszugehen. Ebenso kann eine Verzerrung durch Measurement- und Processing-Error nicht evaluiert werden, weil die entsprechenden Informationen nicht von den Studienautoren zur Verfügung gestellt wurden. Im Ergebnis sind die Ergebnisse des Surveys aufgrund verschiedener methodischer Probleme höchstwahrscheinlich stark verzerrt.*

Um den selbstgesetzten Zielen des Glücksspiel-Surveys gerecht zu werden, ist es erforderlich, dass der Survey höchste Qualitätsmaßstäbe erfüllt, da die Ergebnisse insbesondere der Information und der Entscheidungsunterstützung staatlicher Stellen hinsichtlich legislativer Maßnahmen (zum Beispiel regulatorische Eingriffe in den freien Markt) beraten werden sollen (Münnich 2023). Für die Qualität von Befragungsdaten hat Eurostat (2020) Qualitätskriterien definiert, von denen grundsätzlich die Genauigkeit bzw. Präzision der Daten als wichtigstes Qualitätskriterium angesehen wird (Lohr 2021). Präzision bedeutet in diesem Kontext, dass der auf Basis einer Stichprobe geschätzte Populationsparameter  $\hat{p}$  möglichst nahe am wahren Parameter der Population  $p$  liegt, d.h. geringe Varianz besitzt und unverzerrt ist. Am Beispiel des Glücksspiel-Surveys hat es somit höchste Priorität, dass die Schätzungen – zum Beispiel der Anteil an Menschen mit Glücksspielstörung – so nahe wie möglich am wahren Wert des Anteils an Menschen mit Glücksspielstörung in der Bevölkerung ist (Baker, Brick u. a. 2013). Andere Erwägungen sind im Normalfall nachrangig (Lohr 2021). Natürlich gibt es Ausnahmen, so kann beispielsweise in dringenden Fällen auch die Zeitkomponente Vorrang haben. Argumente, die eine niedrigere Priorisierung des Präzisionskriteriums beim Glücksspiel-Survey rechtfertigen könnten, sind allerdings, insbesondere angesichts der Ziele des Surveys, nicht ersichtlich.

Folglich wird der Glücksspiel-Survey 2021 nachfolgend an der Präzision seiner Angaben gemessen, was in diesem Gutachten anhand des Frameworks des *Total-Survey-Error* (Biemer 2010; Lohr 2021) geschieht. Das Framework ist in Abbildung 4.1 dargestellt. Es untersucht zwei Hauptkomponenten, den *Sampling-Error* (der zufällige Fehler, der sich beim Ziehen der Stichprobe ereignet) und den *Nonsampling-Error* (der nichtzufällige Fehler, der sich außerhalb der Stichprobenziehung ereignet). Unter den *Nonsampling-Error* fallen weiterhin der *Coverage-Error*, der *Nonresponse-Error* auf der Ebene von Individuen bzw. auf der Ebene einzelner Fragen, der *(Self-)Selection-Error* (bei Nonprobability-Sampling), der *Measurement-Error* und der *Processing-Error*. Befragungen geben häufig nur den *Sampling-Error* an, beispielsweise durch Konfidenzintervalle. Damit wird allerdings nur eine mögliche Fehlerquelle quantifiziert und kommuniziert, die verschiedenen *Nonsampling-Errors* bleiben – trotz häufig viel schwerwiegenderer Konsequenzen (Lohr 2021) – häufig unbeachtet, so auch im Glücksspiel-Survey 2021. In diesem Gutachten wird der Survey hingegen nachfolgend hinsichtlich der einzelnen genannten Fehlerquellen an den seinen Zielen entsprechenden Qualitätsmaßstäben gemessen und abschließend der Einfluss des *Nonsampling-Errors* in stark vereinfachten, annahmenbasierten Szenarien dargestellt<sup>1</sup>.

<sup>1</sup>Es sei bereits an dieser Stelle darauf hingewiesen, dass die im vorliegenden Gutachten geschätzten Konfidenzintervalle lediglich illustrativen Zwecken dienen, da hierfür eine genaue Kenntnis des Sampling-Frames und der Auswahlwahrscheinlichkeiten nötig wäre. Die Darstellung soll deshalb lediglich einen Eindruck davon geben, mit welchen Schätzunsicherheiten bei der Ermittlung der Prävalenzen mindestens zu rechnen ist.

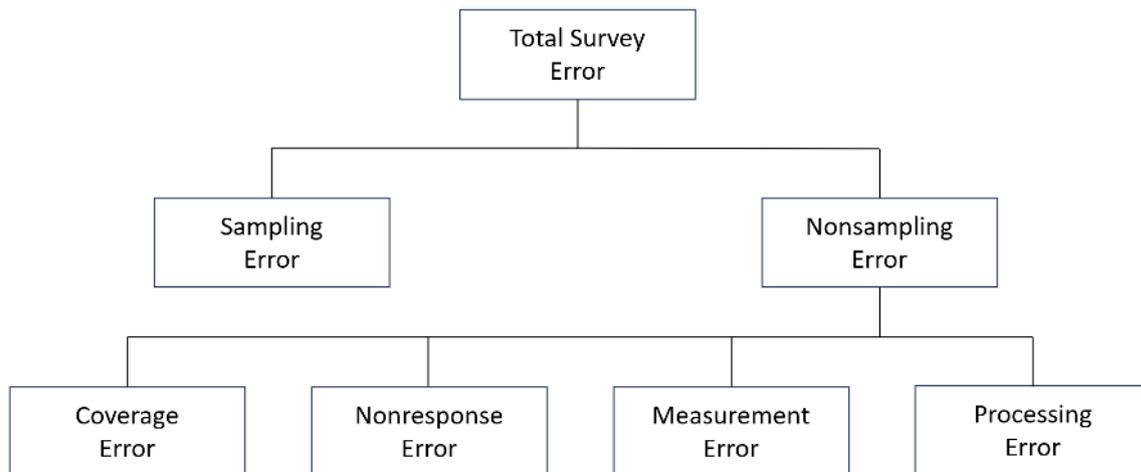


Abbildung 4.1: Komponenten des Total Survey Errors basierend auf Lohr (2021)

## 4.1 Sampling-Error

Der Stichprobenfehler bezieht sich auf den Fehler, der auftritt, weil Befragungen in der Regel auf Stichproben und nicht auf vollständigen Populationen basieren. Zieht man also zwei große zufällige Stichproben mittels *Probability-Sampling* mit der bestmöglichen Methode, werden sich beide Stichproben durch Zufall minimal unterscheiden – diesen Fehler nennt man *Sampling-Error* oder auch Stichprobenfehler. Der Stichprobenfehler wird beim *Probability-Sampling* durch die Standardabweichung des Schätzers approximiert – je größer die Stichprobe, desto kleiner ist in der Regel der Stichprobenfehler (Lohr 2021). In der Praxis bedeutet dies, dass größere (nicht verzerrte!) Stichproben in der Regel präzisere Schätzungen liefern, aber sie sind auch teurer und zeitaufwändiger. Surveys mit dem Ziel präziser Ergebnisse benötigen daher entsprechend große Stichproben. Dies gilt insbesondere, wenn Subpopulationen wie Jugendliche untersucht werden sollen, da bei diesen die Fallzahlen entsprechend kleiner sind.

Ohne Kenntnis des Stichprobenfehlers sind (selbst wenn keine weiteren Fehler vorliegen) keine Aussagen darüber möglich, ob Veränderungen über die Zeit oder Unterschiede zwischen verschiedenen Subgruppen in der Realität tatsächlich vorliegen oder ob es sich bei beobachteten Differenzen um reine Zufallsprodukte handelt.

Häufig wird der Stichprobenfehler mit Konfidenzintervallen angegeben, also einem zufälligen Intervall, das bei bekannter Verteilung der Stichprobe eine vorgegebene Überdeckungswahrscheinlichkeit

für den wahren, aber unbekanntem Parameter besitzt. Das bedeutet, dass bei einem vorgegebenen Konfidenzniveau<sup>2</sup> von 95 % eben 95 % der durch Stichproben ermittelten Konfidenzintervalle den wahren Wert überdecken würden. Diese Intervallschätzung kann mit verschiedenen Berechnungsmethoden vorgenommen werden. Vorzugsweise ist bei Anteilswerten bei simplen Befragungsdesigns das Wilson-Konfidenzintervall zu verwenden (Agresti und Coull 1998, Brown, Cai und DasGupta 2001) – insbesondere in Situationen, in denen der zu schätzende Anteilswert klein oder groß ist, oder wenn die Stichprobengröße klein ist. In solchen Szenarien kann das klassische Konfidenzintervall unangemessen breite (große Konfidenzintervalle) oder schmale (kleine Konfidenzintervalle) Schätzungen liefern. Das Wilson-Konfidenzintervall jedoch minimiert diese Verzerrung und liefert präzisere Intervalle, unter anderem indem es asymmetrische Intervalle berechnet, das heißt die Differenz zwischen der besten Schätzung und dem oberen und unterem Ende des Intervalls unterscheidet sich. Das Wilson-Konfidenzintervall wird mittels folgender Formel berechnet:

$$CI \approx \left[ \frac{\hat{p} + \frac{Z^2}{2n} - Z \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{Z^2}{4n^2}}}{1 + \frac{Z^2}{n}}; \frac{\hat{p} + \frac{Z^2}{2n} + Z \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{Z^2}{4n^2}}}{1 + \frac{Z^2}{n}} \right]. \quad (4.1)$$

Hierbei ist  $\hat{p}$  der Punktschätzer bzw. der Stichprobenanteil,  $Z$  der Z-Wert der Standardnormalverteilung, der vom gewählten Konfidenzniveau abhängt (zum Beispiel 1,96 für ein Konfidenzniveau von 95%), und  $n$  ist die Stichprobengröße.

Für komplexe Survey-Designs hingegen, die beispielsweise verschiedene *Sampling*-Methoden benutzen, ist das Wilson-Intervall hingegen nicht geeignet, vorzugsweise sind hier Methoden wie *Bootstrapping* oder *Taylor Linearization*. Dies gilt allerdings nicht für Stichproben, die mittels *Nonprobability-Sampling* gezogen wurden. Denn beim *Probability-Sampling* hat jedes Mitglied einer Population (theoretisch) eine (theoretisch) bekannte, positive Wahrscheinlichkeit, befragt zu werden (Lohr 2021), was eine Verallgemeinerung der Stichprobe auf die Population ermöglicht<sup>3</sup>. Durch die zugrundeliegende Zufallsauswahl kann der Stichprobenfehler berechnet werden. Dies ist beim *Nonprobability-Sampling* nicht der Fall, denn nicht alle Mitglieder der Population haben eine bekannte und/oder positive Wahrscheinlichkeit, befragt zu werden, wodurch die Varianz der Schätzer unbekannt ist. Daher muss bei solchen Stichproben ein modellbasierter Ansatz mit (starken) Annahmen gewählt werden.

Bei Befragungen sollte der Stichprobenfehler, bestenfalls in Form der niedrigschwelligen Konfidenz-

<sup>2</sup>Es wird im Voraus festgelegt, um anzugeben, in wieviel % der Fälle das aus einer zufälligen Stichprobe berechnete Konfidenzintervall den wahren Wert abdecken würde. Standardmäßig wird ein Niveau von 95 % gewählt.

<sup>3</sup>In telefonischen Befragungen kann beispielsweise eine Person unter mehreren Telefonnummern erreichbar sein, so dass die Inklusionswahrscheinlichkeiten nicht gleich sind. Im Gabler-Häder-Design wird die individuelle Inklusionswahrscheinlichkeit deshalb durch eine entsprechende Frage ermittelt.

zintervalle, immer konsistent und transparent kommuniziert werden. Dies ist unabdingbar, um den Eindruck falscher Genauigkeit zu vermeiden und dem Leser die eigenständige Evaluation der Datenqualität zu ermöglichen. Die Angabe der Unsicherheit ist übrigens nicht auf akademische Sphären beschränkt, sondern findet sich auch in der breiten medialen Kommunikation statistischer Erhebungen wieder, beispielsweise bei der Sonntagsfrage zur Bundestagswahl. Dass die Angabe von Schätzunsicherheiten für das sachgerechte Verständnis statistischer Aussagen unabdingbar ist, zeigen verschiedene Beispiele in Kapitel 10 Rezeption. Weiterhin sollte, sofern Konfidenzintervalle verwendet werden, deren genaue Berechnungsmethode genannt werden. Nur so kann sowohl die Güte des verwendeten Intervalls evaluiert, als auch die Korrektheit der Berechnungen überprüft werden.

### Sampling-Error im Glücksspiel-Survey 2021

Nach dieser generellen Einführung wendet sich das vorliegende Gutachten dem Umgang mit statistischer Unsicherheit im Glücksspiel-Survey 2021 zu. Der Glücksspiel-Survey basiert auf einer großen Stichprobe von 12.303 Interviews und liefert damit auf den ersten Blick eine gute Voraussetzung für Schätzungen mit kleinem Stichprobenfehler. Dies trifft allerdings nur für Schätzungen für die Gesamtbevölkerung zu. Werden Subgruppen – zum Beispiel Jugendliche – analysiert, ist der Stichprobenfehler entsprechend der geringeren Fallzahl sehr groß. Problematisch ist allerdings, dass, obwohl der Survey zu knapp 60 % via *Probability-Sampling* sowie zu ca. 40 % mittels *Nonprobability-Sampling* durchgeführt wurde, die Methode zur Berechnung der Konfidenzintervalle nicht angegeben wird. Während bei der telefonischen Stichprobe der Stichprobenfehler quantifiziert werden kann, ist dies bei der Internetstichprobe nur auf Basis von Modellannahmen möglich (Lohr 2021). Informationen diesbezüglich liefert der Survey allerdings nicht. So bleibt unklar, ob ein der Komplexität des Survey-Designs – immerhin wurden *Probability* und *Nonprobability-Sampling* kombiniert – angemessenes Konfidenzintervall gewählt wurde und falls ja, welche Annahmen getroffen wurden. Aufgrund dieser Intransparenz hätten die Berechnungen der Autoren nicht repliziert werden können, selbst wenn sie ihre Daten für die Erstellung des Gutachtens zur Verfügung gestellt hätten, wie es wissenschaftlichen Standards entsprechen hätte.

Am kritischsten ist jedoch, dass Konfidenzintervalle im Glücksspiel-Survey 2021 nur selten und sehr selektiv dargestellt werden. Welche Faktoren die Studienautoren zur (selektiven) Ermittlung von Konfidenzintervallen heranziehen, ist nicht ersichtlich. Weder in der Zusammenfassung noch in den Grafiken oder im Fazit wird der Stichprobenfehler mitkommuniziert, obwohl Inferenzschlüsse auf die Gesamtbevölkerung vorgenommen werden. Die Selektivität der Darstellung von Konfidenzintervallen gilt für

die tabellarische Darstellung wie auch für den Fließtext: Nur in Tabelle 7 werden Konfidenzintervalle gezeigt. Warum in allen anderen Tabellen keine angegeben werden, bleibt unklar. Ebenso gilt für den Text: Es werden nur drei Konfidenzintervalle erwähnt<sup>4</sup>, ein verschwindend geringer Anteil an der insgesamt sehr hohen Anzahl genannter Anteilswerte. Dies ist in vielerlei Hinsicht problematisch, denn die isolierte Nennung des Stichprobenanteils  $\hat{p}$  vermittelt eine Genauigkeit und Übertragbarkeit auf die Gesamtbevölkerung, die so statistisch nicht haltbar ist. Zudem wird so dem Leser die eigenständige Evaluation der Datenqualität verwehrt. Dieses Vorgehen steht in Kontrast zu den vorhergehenden Erhebungen durch die BZgA – diese berechneten und kommunizierten den Stichprobenfehler konsistent und in einer der Stichprobenkomplexität angemessenen Art und Weise. Warum die Autoren des Glücksspiel-Surveys 2021 diese in der (Umfrage-)Forschung übliche (und sachgerechte) Praxis nicht konsequent weiterführen, ist nicht ersichtlich.

Um die mit den Ergebnissen des Glücksspiel-Surveys 2021 verbundene Unsicherheit und deren Ausmaß einschätzen zu können, sollten als Teil dieses Gutachtens im Survey rein illustrativ einzelne Konfidenzintervalle ermittelt werden. Unter anderem aus diesem Grund wurden die Daten des Surveys bei ihren Autoren angefragt – diese lehnten allerdings die Anfrage nach ausführlicher Korrespondenz und Offenlegung der geplanten Nutzung ab, entgegen wissenschaftlichen Standards (für eine ausführliche Kritik, siehe Kapitel 8). Aus diesem Grund ist es nicht möglich, der Komplexität des Surveydesigns angemessene Konfidenzintervalle zu berechnen. Dennoch soll anhand eines Beispiels gezeigt werden, welche Konsequenzen das (un)bewusste Weglassen der Konfidenzintervalle hat. Aufgrund der fehlenden Informationen können die im vorliegenden Gutachten berechneten Konfidenzintervalle nicht korrekt sein, es wurde beispielsweise angenommen, dass alle Befragten auf die gleiche (simple) Art und Weise mittels *Probability-Sampling* ausgewählt wurden. Das ausgewählte Wilson-Intervall ist also der Komplexität des Surveys nicht angemessen, was der folgende Vergleich deutlich zeigt: Das von der BZgA im Glücksspiel-Survey 2019 angegebene Konfidenzintervall, das für das komplexe Design geeignet war, betrug für den Anteil an wahrscheinlich pathologisch Glücksspielenden: 0,11 % – 1,06 %. Das in diesem Gutachten ermittelte Wilson-Intervall liefert hingegen ein viermal schmaleres Intervall von 0.23 % bis 0.45 %. Das hier ermittelte Intervall ist also definitiv zu optimistisch, die tatsächlichen Konfidenzintervalle müssen größer sein.

Die Gründe dafür sind vielfältig: Wird mehr als ein Konfidenzintervall berechnet, liegt ein sogenanntes multiples Testproblem vor, was eine Korrektur (d.h. eine Verbreiterung) der einzelnen Konfidenzintervalle erfordert, um die Überdeckungswahrscheinlichkeit zu gewährleisten (sogenannte „Type-I-Error-Inflation“). Selbst wenn man die Inklusionswahrscheinlichkeiten für ein Probability-Sampling kennen

---

<sup>4</sup>Berücksichtigt man Fußnoten, sind es vier Erwähnungen.

würde, müsste eine Nonresponse-Korrektur durchgeführt und in der Inferenz berücksichtigt werden („Variance Estimation under Imputation“), um die sogenannte „Variance Inflation due to missingness“ auszugleichen. Allerdings darf bei der Interpretation der Konfidenzintervalle nicht vergessen werden, dass diese nur den *Sampling-Error* und nicht den *Nonsampling-Error* quantifizieren. Daher wird ein großer Teil des *Total-Survey-Errors* gar nicht in den Konfidenzintervallen berücksichtigt (Lohr 2021).

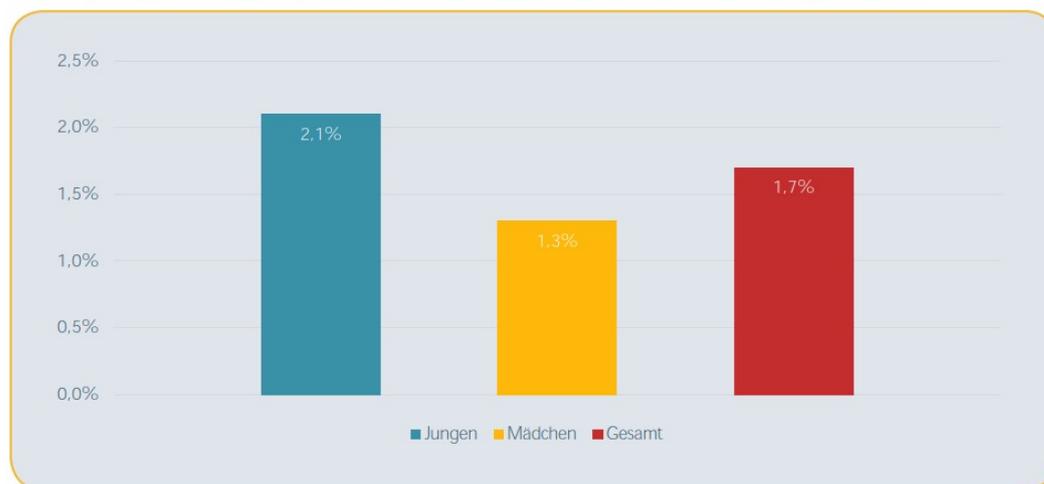
Zudem inkludiert der Stichprobenfehler nicht die durch den *Nonsampling-Error* induzierten Verzerrungen, welche die ohnehin vorhandene Unsicherheit weiter stark vergrößern würde. Unter Berücksichtigung der vorhandenen (unbekannten) Verzerrung sollten Aussagen basierend auf diesen Daten daher mit größter Zurückhaltung formuliert und ausschließlich unter Berücksichtigung und Nennung der durch die Unsicherheit und Verzerrung induzierten Limitationen getroffen werden. Dies gilt insbesondere aufgrund des genutzten *Nonprobability Samples*, dessen Stichprobenverzerrung sich nur auf Basis von starken Annahmen quantifizieren bzw. bestenfalls grob abschätzen lässt. Von substanziellen Interpretationen im Sinne von „*Kategorie X hat den höchsten Anteilswert und sollte daher ...*“ ist in diesen Fällen aus statistischer Sicht daher oftmals abzuraten. Ohne Sichtung der Daten der Glücksspiel-Survey 2021, was von den Studienautoren abgelehnt wurde, ist keine tiefergehende Einschätzung sowohl des Ausmaßes und als auch der letztlichen Auswirkung des gesamten Befragungsfehlers auf die Ergebnisse möglich. Dennoch illustriert das nachfolgende Beispiel nachdrücklich, warum die Kommunikation der Unsicherheit so essenziell ist.

### Geschlechterunterschied bei Jugendlichen

Man betrachte zur Illustration ein beliebig gewähltes Beispiel aus dem Survey: Abbildung 8 (hier Abbildung 4.2) zeigt den Anteil an Minderjährigen (16- bis 17-Jährige) mit Glücksspielproblemen nach DSM-IV-MR-J (Instrument für Jugendliche). Im Gegensatz zu allen anderen Grafiken wird in dieser Abbildung nicht erläutert, ob es sich um gewichtete oder ungewichtete Daten handelt. Die Studienautoren selbst beschreiben ihre Ergebnisse folgendermaßen: „*Nach dem Screening der glücksspielbezogenen Probleme bei Minderjährigen anhand der Kriterien des „DSM-IV-Multiple Response-Adapted for Juveniles (DSM-IV-MR-J)“ zeigt sich bei 1,7 % (95 %-KI: 0,7 % – 4,1 %) der Jugendlichen im Alter von 16-17 Jahren ein problematisches Spielverhalten (siehe Abbildung 8). Davon sind 2,1 % der männlichen und 1,3 % der weiblichen Jugendlichen betroffen.*“ (Buth, Meyer und Kalke 2022, S. 34 f)

Positiv hervorzuheben ist hier die Nennung des Konfidenzintervalls für den Gesamtanteil von Jugendlichen mit Glücksspielproblemen, insbesondere aufgrund der geringen Fallzahl ( $n = 280$ ). Überraschend ist jedoch, dass keine Konfidenzintervalle für den Anteil an Jungen und Mädchen genannt

**Abbildung 8:**  
**Minderjährige: Glücksspielprobleme nach DSM-IV-MR-J (Instrument für Jugendliche)**



Basis: 16- bis 17-Jährige; ungewichtete Fallzahl: N=280

Abbildung 4.2: Originalgrafik in Buth, Meyer und Kalke (2022)

werden. Diese wurden daher im Folgenden mit dem oben dargestellten Wilson-Intervall bestimmt<sup>5</sup>. Wie in Abbildung 4.3 an den Fehlerbalken zu erkennen ist, überlappen die Konfidenzintervalle, was eine erste Indikation dafür liefert, dass der Unterschied zwischen den Gruppen mutmaßlich nicht statistisch signifikant ist. Der exakte (zweiseitige) Test nach Fisher<sup>6</sup> (Fisher 1925) zeigt, dass der Unterschied zwischen Jungen und Mädchen tatsächlich nicht signifikant ist ( $p > 0.05$ ). Dies entspricht der Erwartung, wenn man bedenkt, dass in absoluten Zahlen sehr wenige Beobachtungen vorliegen. Beispielsweise gilt für die Mädchen: 1,3 % von 280 Beobachtungen sind 3,64 Personen, die Nachkommastellen ergeben sich aus der Gewichtung. Berücksichtigt man also den Stichprobenfehler, kann man keine statistisch fundierte Aussage darüber treffen, ob bei Jungen oder bei Mädchen mehr Glücksspielprobleme in der Bevölkerung vorhanden sind – dafür gibt es einfach zu wenige Daten. Diese essenzielle Information fehlt allerdings im Glücksspiel-Survey 2021, wodurch dem Leser ein Unterschied zwischen den Geschlechtern vermittelt wird, obwohl dieser statistisch nicht belegbar ist. Sowohl die Inklusion von Konfidenzintervallen in Text und Grafik als auch der klare Hinweis, dass die Konfidenzintervalle überlappen und/oder der Unterschied nicht statistisch signifikant ist, wäre nicht nur gemäß guter wissenschaftlicher Praxis, sondern insbesondere auch mit Blick auf die Nutzerorientierung zwingend notwendig gewesen.

<sup>5</sup>Diese sind, wie bereits erläutert, zu optimistisch und als Untergrenze für die Schätzunsicherheit unter optimalen Bedingungen anzusehen.

<sup>6</sup>Die Durchführung dieses Tests setzt ebenfalls optimale Bedingungen, d.h. insbesondere das Nichtvorliegen von Nonsampling-Errors, voraus.

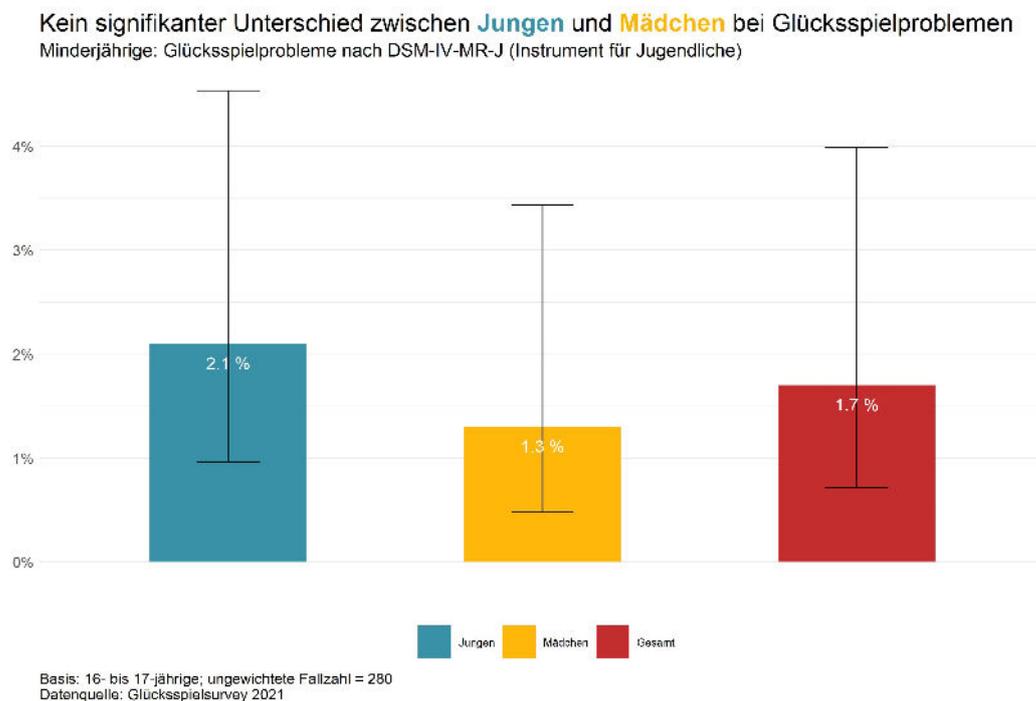


Abbildung 4.3: Grafik mit Konfidenzintervallen

Des Weiteren sollte der Stichprobenfehler auch bei der substantiellen Interpretation der Werte (abseits des Gruppenvergleichs) berücksichtigt werden. Dies wird erneut am Beispiel der 1,3 % der Mädchen, die gemäß der Befragung Glücksspielprobleme haben, illustriert. Diese Zahl ist, unter Berücksichtigung des Konfidenzintervalls, folgendermaßen zu lesen: Basierend auf der vorliegenden Stichprobe – unter der Annahme, dass die Stichprobe repräsentativ und nicht verzerrt ist<sup>7</sup> – schätzt man den Anteilswert in der weiblichen 16- bis 17-jährigen Bevölkerung auf 1,3 % und vermutet den wahren Anteilswert mit 95 % Konfidenz zwischen 0,5 % und 3,4 %. Auf den ersten Blick ist das kein großer Unterschied, allerdings werden die Zahlen aus dem Glücksspiel-Survey 2021 teilweise öffentlich durch politische Akteure auf die Bevölkerung hochgerechnet, zum Beispiel vom Innensenator für Inneres der Freien Hansestadt Bremen (Senator für Inneres der Freien Hansestadt Bremen 2023). Folgt man dieser Logik, hätten 1,3 % der 16- bis 17-jährigen Mädchen in Deutschland eine Glücksspielstörung, in absoluten Zahlen 61.768 Mädchen<sup>8</sup>. Diese Vorgehensweise ignoriert allerdings die Unsicherheit der Stichprobe, eine seriöse Angabe würde stattdessen das Konfidenzintervall kommunizieren. Es muss von einer erheblichen Spannweite der Unsicherheit in den Absolutzahlen ausgegangen werden, die eindrücklich die Relevanz von Konfidenzintervallen verdeutlicht, selbst wenn diese anhand der prozentualen Angaben klein bzw.

<sup>7</sup>Die getroffene Annahme ist für den Glücksspiel-Survey 2021 nicht haltbar, siehe Abschnitt 4.2.

<sup>8</sup>Basierend auf Bevölkerungsdaten vom Statistischen Bundesamt für 2021 (Statistisches Bundesamt (Destatis) 2023)

vernachlässigbar erscheinen.

Zusammengefasst wird der Stichprobenfehler im Glücksspiel-Survey 2021 nur selektiv kommuniziert, wodurch durchgängig eine falsche Genauigkeit vermittelt wird. Zudem müssten einige Aussagen der Studienautoren differenziert werden, wenn die Unsicherheit berücksichtigt wird – was eigentlich standardmäßig in der Wissenschaft geschehen sollte (Deutsche Forschungsgemeinschaft 2019). Da die Autoren allerdings keine Daten zur Verfügung gestellt haben, können die Aussagen aus methodischer Sicht nicht korrekt überprüft werden, was (erneut) nicht wissenschaftlichen Qualitätsstandards entspricht (Deutsche Forschungsgemeinschaft 2019). Das vorherige Beispiel lässt sich für nahezu beliebige Zahlen, Grafiken und Tabellen im Survey wiederholen, fast immer wäre mindestens eine Differenzierung der Aussagen notwendig. Allerdings quantifiziert der Stichprobenfehler nicht den *Nonsampling-Error*, wodurch die angegebene Unsicherheit immer zu gering ist.

## 4.2 Nonsampling-Error

### 4.2.1 Coverage-Error

Der *Coverage-Error* (oder auch Abdeckungsfehler) entsteht, wenn die zu untersuchende Population, also die Gesamtheit der Individuen, aus der eine Stichprobe gezogen werden soll, mit dem gewählten Sampling-Frame, d.h. der Auswahlgesamtheit, aus der gezogen werden kann, nicht vollständig erfasst wird. Mit anderen Worten: Ein Abdeckungsfehler liegt vor, wenn einige Individuen aus der Zielbevölkerung keine positive Wahrscheinlichkeit besitzen, in die Stichprobe aufgenommen zu werden und/oder wenn einige Individuen, die nicht der Zielbevölkerung angehören, eine positive Wahrscheinlichkeit besitzen in die Stichprobe aufgenommen zu werden. Der Abdeckungsfehler kann sich folglich in zwei unterschiedlichen Arten manifestieren: in der Unterabdeckung (Undercoverage) und der Überabdeckung (Overcoverage). Bei der (üblicheren) Unterabdeckung werden einige Mitglieder der Zielbevölkerung gar nicht oder zu selten in die Stichprobe einbezogen, beispielsweise können an einer Online-Befragung Menschen ohne Internetzugang nicht teilnehmen. Bei der Überabdeckung hingegen können einige Individuen mehrfach erfasst oder fälschlicherweise in die Population aufgenommen werden, etwa weil sie mehrere Telefonanschlüsse besitzen oder weil die Auswahlliste auch Telefonanschlüsse von Unternehmen beinhaltet. Beide Formen des Abdeckungsfehlers können die Präzision von Umfrageergebnissen erheblich beeinträchtigen. Sie können zu einer Verzerrung der Ergebnisse führen, da bestimmte Gruppen in der Bevölkerung über- oder unterrepräsentiert sein können (Lohr 2021). Daher ist es sehr wichtig, dass Autoren über die *Coverage* ihrer Befragungen transparent berichten.

#### Coverage-Error im Glücksspiel-Survey 2021

Im Glücksspiel-Survey wurde ein gemischter Ansatz zur Stichprobenauswahl verwendet: Knapp 40 % wurden online befragt, ca. 30 % durch eine einfache Zufallsauswahl von Mobilfunknummern und die übrigen knappen 30 % durch eine geschichtete Zufallsauswahl von Festnetznummern. Die einfache Zufallsauswahl kam bei der Mobilfunkstichprobe zum Einsatz, wobei der Sampling-Frame „*alle möglicherweise vergebenen Mobilfunknummern in allen von der Bundesnetzagentur freigegebenen Vorwahlbereichen*“ (Buth, Meyer und Kalke 2022, S. 15) darstellte. Bei der Festnetzstichprobe entsprach der Sampling-Frame „*alle[n] möglicherweise vergebenen deutschen Festnetznummern*“ (Buth, Meyer und Kalke 2022, S. 15), die nach Bundesland geschichtet und dann durch uneingeschränkte proportionale Zufallsauswahl ausgewählt wurden. Dies bedeutet, dass bei der Festnetzstichprobe der Anteil der

Befragten der jeweiligen Bundesländer ihrem Anteil an der deutschen Gesamtbevölkerung entsprechen sollte. Die Nutzung einer Festnetz- und Mobilfunkstichprobe wird „Dual-Frame-Verfahren“ genannt.

Grundsätzlich ist die beim Dual-Frame-Verfahren genutzte 50-50-Aufteilung der telefonischen Stichprobe positiv zu bewerten, sie entspricht damit den Empfehlungen der Arbeitsgemeinschaft ADM-Telefonstichproben (Arbeitsgemeinschaft ADM-Telefonstichproben 2016). Ebenso ist die Nutzung des geschichteten Auswahlverfahrens bei der Festnetzstichprobe positiv zu bewerten, da somit eine Repräsentation der Bevölkerung nach Bundesländern (zumindest teilweise) gewährleistet ist. Zudem wurde transparent über die gewählten Populationen berichtet, was dem Leser die Evaluation der Qualität des Surveys in dieser Hinsicht ermöglicht. Unklar bleibt allerdings, ob gemäß den ADM-Empfehlungen auch eine Gewichtung nach Babler-Häder umgesetzt wurde, da sich dies aus den Ausführungen der Autoren zur Gewichtung (Buth, Meyer und Kalke 2022, S. 16) nicht klar ergibt. Die Vorgehensweise der Autoren des Glücksspiel-Surveys 2021 bei der telefonischen Stichprobenziehung ist gleichwohl insgesamt positiv zu bewerten.

Anders sieht es bezüglich der Online-Stichprobe aus, die von einem starken (*Self-Selection-Error*) geprägt ist. Dies wird im Abschnitt 4.2.4 ausführlich diskutiert. Derartige Selektionsverzerrungen begründen die prinzipiell mangelnde Eignung von nichtprobabilistischen Online-Stichproben für bevölkerungsrepräsentative Schätzungen.

### 4.2.2 Unit-Nonresponse

Eine Verzerrung durch Unit-Nonresponse liegt vor, wenn sich die Befragten und diejenigen, die eine Teilnahme an der Befragung abgelehnt haben, substantiell unterscheiden. Diese Verzerrung kann einen starken Effekt auf Umfrageergebnisse haben, wie in der Literatur schon mehrmals festgestellt wurde (z.B. Lohr (2021) oder Lorant u. a. (2007)). Dazu soll eingangs ein Beispiel betrachtet werden: In einem Großraumbüro werden Zettel mit einer Befragung ausgelegt, um die Arbeitsbelastung zu messen. Da 90 % der Mitarbeiter aufgrund der hohen Arbeitsbelastung keine Zeit für die Befragung haben, beantworten nur die übrigen 10 % mit geringer Arbeitsbelastung die Befragung. Die Befragung liefert also ein verzerrtes bzw. falsches Bild der Realität, weil sich die Teilnehmer (niedrige Belastung) von den Nicht-Teilnehmern (hohe Belastung) substantiell unterscheiden. Diese Verzerrung wird „Nonresponse-Error“ genannt.

Dies wird auch beispielhaft in Abbildung 4.4 illustriert: Da die Teilnehmer einer fiktiven Befragung zur Glücksspielneigung (50 % spielen wenig) sich systematisch von den Nicht-Teilnehmern unterscheiden (hier spielen 100 % wenig), sind die Umfrageergebnisse entsprechend verzerrt (natürlich wäre auch eine umgekehrte Logik denkbar). Die realisierte Stichprobe der Befragten wäre hier also nicht repräsentativ für die Gesamtbevölkerung.

### Glücksspielprävalenz der Bevölkerung: 75% spielen wenig, 25% spielen viel

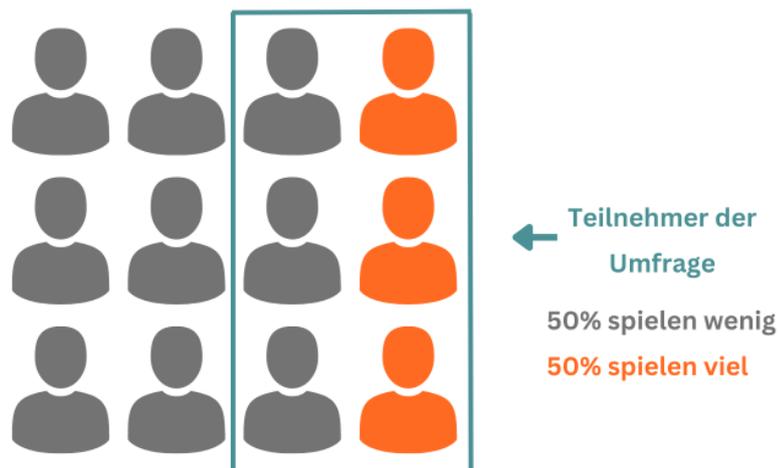


Abbildung 4.4: Beispiel Unit-Nonresponse

Trotz dieses in der Wissenschaft seit Jahrzehnten nachgewiesenen Problems publizieren viele Autoren dennoch ihre Befragungsergebnisse ohne Beachtung eines *Nonresponse-Error* und übertragen sie ohne Einschränkungen auf die Bevölkerung (Lohr 2021). Selbst wenn Autoren argumentieren, dass sich Teilnehmer und Nicht-Teilnehmer einer Befragung nicht substantiell unterscheiden, ist dies oft eine Illusion (Särndal und Lundström 2005). Insofern sind jegliche Aussagen in Befragungen mit substantieller Unit-Nonresponse-Quote, die sich auf Anteilswerte in der Bevölkerung beziehen, methodisch gesehen hoch problematisch. Inferenzen wären nur dann korrekt, wenn vom Anteil der Bevölkerung geschrieben würde, der an der jeweiligen Befragung auch teilnimmt. Dafür sind allerdings auch Informationen über die Teilnehmer und Nicht-Teilnehmer der Befragung notwendig. Die Größe der Stichprobe kann den Unit-Nonresponse-Bias übrigens nicht mitigieren: Wenn nur Männer und keine Frauen an einer Befragung teilnehmen, ist es egal, ob 100 oder 100.000 teilnehmen – der Unit-Nonresponse-Bias bleibt gleich (Lohr 2021).

Da *Unit-Nonresponse* bei jeder (freiwilligen) Befragung existiert, ist dessen transparente Kommunikation und (soweit möglich) Korrektur essenziell. Dazu müssen Informationen über die *Nonrespondents* – sofern möglich – präsentiert und analysiert werden, inwieweit der *Nonresponse* eine Verzerrung hervorruft. Nur so ist es dem Leser möglich, sich eine eigene Meinung zur Qualität der vorliegenden Daten zu bilden – so wie es beispielsweise auch die Deutsche Forschungsgemeinschaft (2019) empfiehlt.

### Unit-Nonresponse im Glücksspiel-Survey 2021

Im Glücksspiel-Survey 2021 ist potenziell ein starker *Unit-Nonresponse-Error* vorhanden, der allerdings entgegen wissenschaftlichen Grundprinzipien (Deutsche Forschungsgemeinschaft 2019) von den Studienautoren weder explizit genannt oder diskutiert noch zur Einordnung der Ergebnisse genutzt wird. Das ist bei einer Befragung mit dieser Zielsetzung (vgl. Kapitel 2) höchst problematisch. Denn es gibt gute Gründe anzunehmen, dass sich die tatsächlich Antwortenden – wie oben erläutert – von denjenigen unterscheiden, die nicht an der Befragung teilgenommen haben. Dies ist umso wahrscheinlicher, da die Autoren alle unvollständigen Interviews ausgeschlossen haben, wodurch sich der (vermutlich systematische) *Item-Nonresponse* in *Unit-Nonresponse* überträgt. Empfehlenswert gewesen wäre vielmehr, unvollständige Interviews für eine (zumindest) teilweise Nonresponse-Korrektur zu nutzen, was allerdings ein methodisch anspruchsvolles Vorgehen darstellt. Die Beschränkung auf vollständige Interviews resultiert in einer *Nonresponse-Quote* von 90 % für die Onlinebefragung und 73 % für die telefonische Befragung. Konkret bedeutet dies, dass von den Angerufenen 73 % das Interview verweigert oder vorzeitig abgebrochen haben und von den online zur Befragung eingeladenen

Personen 90 % die Befragung nicht oder nicht vollständig beantwortet haben.

Bemerkenswerterweise begründeten die Studienautoren den Wechsel zu einer *Mixed-Mode*-Erhebung unter anderem mit einer hohen Unit-Nonresponse-Quote bei telefonischen Befragungen, führen diesen Gedanken allerdings nicht mehr bei der Diskussion und Einordnung der eigenen Studienergebnisse aus (siehe Abschnitt 4.2.4). Es ist daher unklar, inwieweit Unit-Nonresponse die Umfrageergebnisse verzerrt, denn die Ausschöpfungsquote ist nicht linear mit der Unit-Nonresponse-Verzerrung verbunden (Lohr 2021): Eine Befragung kann eine hohe Ausschöpfung von 90 % aufweisen, zugleich jedoch aufgrund systematisch unterschiedlichen übrigen 10 % der Befragten einen extrem starken *Unit-Nonresponse-Error* besitzen. Umgekehrt könnte der Glücksspiel-Survey mit seinen niedrigen Ausschöpfungsquoten von 27 % und 10 % theoretisch – wenn auch sehr (!) unwahrscheinlich – einen geringen Unit-Nonresponse haben, sofern die Nicht-Teilnehmer der Befragung sich nicht systematisch von den Teilnehmern unterscheiden. Betrachtet man diesbezüglich die wissenschaftliche Evidenz, ist ein solch optimistisches Szenario allerdings nicht plausibel. Selbst eine Gewichtung kann eine solche Verzerrung nicht korrigieren (siehe z.B. Enderle, Münnich und Bruch (2013)).

Ein weiterer starker Hinweis auf vorliegenden Unit-Nonresponse ist der (vordergründig) nicht vorhandene (siehe 4.2.3) *Item-Nonresponse-Error*. Da von den Studienautoren nur vollständige Interviews berücksichtigt wurden, dürften unvollständige Interviews, die bei den genutzten sensiblen Daten häufiger vorkommen, entfernt worden sein. Insbesondere das genutzte DSM-IV-Instrument ist in Einzelfällen für recht hohe Item-Nonresponse-Quoten von 5 % oder 9 % bekannt, bei denen sich die Non-Responder signifikant in für Glücksspielprobleme relevanten Faktoren von den Respondern unterscheiden (Scholes u. a. 2008). Durch Adjustierung dieses Item-Nonresponse wurde im Report von Scholes u. a. (2008) die Glücksspielproblemprevalenz leicht nach unten korrigiert, auch innerhalb der untersuchten Subgruppen gab es Verschiebungen. Aufgrund desselben Frageinstruments ist ein ähnlicher Effekt auch für den Glücksspiel-Survey 2021 plausibel, aber da die Studienautoren *Item-Nonresponse* nicht untersucht haben (oder dies zumindest im Survey-Bericht nicht dokumentiert haben), lässt sich diese Vermutung nicht überprüfen.

Zusammengefasst kann eine derart hohe *Nonresponse-Quote* wie im Glücksspiel-Survey auch durch Gewichtung nicht mehr korrigiert werden, wodurch eine starke Verzerrung der Daten anzunehmen ist (Enderle, Münnich und Bruch 2013). Da die Studienautoren allerdings weder *Nonresponse* untersuchen noch relevante Daten diesbezüglich präsentieren, kann eine Einschätzung diesbezüglich nur näherungsweise geschehen. Diese Intransparenz ist zu kritisieren, insbesondere angesichts der Ziele des Surveys. Wenn Befragungsdaten der Information oder Entscheidungsunterstützung von Poli-

tik dienen sollen, müssen solche eklatanten Schwächen mitkommuniziert werden, um Fehlschlüsse zu verhindern (Münnich 2023). Dieses Vorgehen entspricht nicht den wissenschaftlichen Qualitätsstandards (Deutsche Forschungsgemeinschaft 2019) – der Leser hat keine Möglichkeit, die Qualität der im Glücksspiel-Survey 2021 präsentierten Daten zu evaluieren. Dieses Verhalten steht im Kontrast zum letzten BZgA-Survey (2019), welcher das Problem des *Nonresponse-Errors* zumindest ausdrücklich nennt. Es ist nicht nachvollziehbar, warum die Autoren des Glücksspiel-Surveys 2021 dies nicht ebenso getan haben.

### 4.2.3 Item-Nonresponse

*Item-Nonresponse-Error* liegt vor, wenn einzelne Fragen von bestimmten Individuen(-gruppen) nicht beantwortet werden. Ein typisches Beispiel dafür sind sensible Fragen, beispielsweise zum Einkommen oder (vermutlich) auch zum Geldeinsatz beim Glücksspiel (Lohr 2021). Wenn also Menschen mit Glücksspielstörung sich für ihre hohen Geldeinsätze schämen und diese Frage nicht beantworten, sind Spieler mit Glücksspielstörung folglich in der Stichprobe bezüglich dieser Frage unterrepräsentiert und die gemessenen Einsätze daher zu niedrig. Diese Verzerrung ist natürlich in jegliche Richtungen denkbar – vielleicht haben Menschen, die kein Glücksspiel spielen, keine Lust, Fragen zum Glücksspiel zu beantworten. Item-Nonresponse führt also zu verzerrten Ergebnissen, die nicht die Realität abbilden, insbesondere bei sensiblen Fragen.

Es gibt in der Literatur Evidenz, dass Item-Nonresponse insbesondere bei Fragen zu problematischem Glücksspielverhalten ein Problem darstellt (z.B. Scholes u. a. (2008)). Daher wird bei qualitativ hochwertigen Befragungen – auch außerhalb der Glücksspielforschung – diese Problematik bereits beim Design der Befragung berücksichtigt (Brooker, Clara und Cox 2009), beispielsweise durch Pretests, also das Testen des Fragebogens vor der eigentlichen Befragung<sup>9</sup>.

Themenspezifische Evidenz zum Glücksspielverhalten findet sich bei Scholes u. a. (2008), den Nonresponse in der British Gambling Prevalence Survey 2007 bezüglich des DSM-IV-Instruments untersuchen, das auch im Glücksspiel-Survey 2021 genutzt wird<sup>10</sup>. Dabei stellen die Autoren fest, dass 9 % (= 541) der Befragungsteilnehmer die DSM-IV-Fragen nicht beantwortet hatten. Dieser Item-Nonresponse war bei allen Instrumenten der Befragung, die Glücksspielstörungen messen, vorhanden und ist somit nicht mit einer generellen Umfragemüdigkeit erklärbar (Scholes u. a. 2008). Dabei unterschieden sich die Nicht-Antwortenden systematisch (substanziell und signifikant) von den Teilnehmern:

- Nicht-Antwortende waren jünger (16 % waren 16 bis 24 Jahre alt) als Antwortende (9 %).
- Nicht-Antwortende hatten verstärkt einen Migrationshintergrund.
- Nicht-Antwortende waren verstärkt im niedrigsten Einkommens-Quantil.

---

<sup>9</sup>Es ist anzumerken, dass ein für bestimmte Zwecke getesteter Fragebogen nicht zwangsweise für andere Zwecke geeignet ist. So kann die Eignung eines Fragebogens, der für die Diagnostik einer Störung getestet wurde, nicht ohne Weiteres auch für ein Screening angenommen werden. Denn es ist im Allgemeinen davon auszugehen, dass ein Individuum, das zum Zweck der medizinischen Diagnostik und eventuell anschließenden Therapie eines vermuteten Problems einen Fragebogen ausfüllt, ein höheres Interesse am Ergebnis besitzt als ein zufällig ausgewählter Befragter.

<sup>10</sup>Im Fragebogen des Glücksspiel-Surveys 2021 wird das DSM-IV Instrument genutzt, in der Analyse selbst wird allerdings DSM-5 genutzt, bei dem eine Kategorie weggelassen wird und folglich die Schwelle zur Störung sinkt.

- Nicht-Antwortende nahmen seltener an Glücksspiel teil: Nur 20 % der Nicht-Antwortenden nahmen in der vorherigen Woche an Glücksspiel teil, unter den Antwortenden waren es hingegen 65 %. Auch bei der Spielform gibt es starke Unterschiede, beispielsweise bei der nationalen Lotterie (Nicht-Antwortende 60 % vs. 87 % der Antwortenden). Eine genaue Aufschlüsselung findet sich bei Scholes u. a. (2008).

Diese Unterschiede sind besonders relevant, da Nicht-Antwortende teilweise eher den Faktoren (jung, Migrationshintergrund) entsprechen, die gemeinhin als Risiko für Glücksspielprobleme gelten (Buth und Stöver 2008; Banz 2019). Andererseits nehmen die Nicht-Antwortenden bedeutend seltener an Glücksspielen teil, was generell auf ein geringeres Risiko für eine Glücksspielstörung hindeutet (Scholes u. a. 2008).

Diese systematischen Unterschiede versuchen die Autoren mit der persönlichen Relevanz des Umfrage-themas zu erklären, also dass beispielsweise selten Spielende schlichtweg weniger Interesse an Fragen zur Glücksspielproblemen haben und diese daher nicht beantworten bzw. überspringen<sup>11</sup>. Dieser Effekt war in der untersuchten Studie nicht zu unterschätzen: Spieler die weniger als einmal im Monat spielen, hatten fünfmal niedrige Chancen (*Odds*) die DSM-IV Fragen zu beantworten als Spieler, die zwei- bis dreimal die Woche spielen. Im Ergebnis hat in der betreffenden Befragung das DSM-IV (ohne Gewichtung) ein verzerrtes Bild der Realität geliefert, da die Antworten auf die betreffende Frage nicht repräsentativ für alle Befragten sind, sondern nur für alle Antwortenden. Dies wird in Abbildung 4.5 gezeigt, die Item-Nonresponse-Bias anhand von zwei (fiktiven) Fragen illustriert. Deutlich erkennbar sind dort die beispielhaften Umfrageergebnisse für die Frage nach einer Glücksspielstörung erheblich verzerrt. Umso wichtiger ist es also, Informationen über eine potenziell vorhandene Verzerrung transparent zu kommunizieren.

In der British Gambling Prevalence Survey 2007 adjustierten die Autoren ihre Gewichte entsprechend dem gefundenen Item-Nonresponse-Bias, was zu einer niedrigeren Gesamtprävalenz an Glücksspielstörungen führte. Ebenso veränderten sich die Anteilswerte für die verschiedenen Subgruppen. Dies sollte nicht überraschen, da unter anderem die seltenen Spieler, die seltener auf die Frage zur Glücksspielproblemen geantwortet hatten, nun höher gewichtet wurden (Scholes u. a. 2008).

Diese Ergebnisse aus Großbritannien sind natürlich nicht automatisch auf alle anderen Befragungen, zum Beispiel den Glücksspiel-Survey 2021, übertragbar. Dennoch erscheint es unrealistisch, dass Item-Nonresponse beim DSM-IV-Instrument nur in Großbritannien bzw. nur in dieser Stichprobe ein erheb-

---

<sup>11</sup>Folglich ist diese Verzerrung unabhängig von der Erhebungsart (Scholes u. a. 2008).

liches Problem war. Die Autoren der Studie berichten zudem von einer weiteren Studie<sup>12</sup>, die ebenso eine Item-Nonresponse-Quote von 5 % beim DSM-IV Instrument hatte. Es ist also anzunehmen, dass das DSM-IV-Instrument zumindest potenziell anfällig für eine Verzerrung durch Item-Nonresponse ist.



Abbildung 4.5: Beispiel Item-Nonresponse

Zusammengefasst ist *Item-Nonresponse-Error* ein bekanntes Problem bei Befragungen, das insbesondere bei sensiblen Fragen auftritt. Aufgrund der potenziell starken Verzerrungen muss *Item-Nonresponse* bei Befragungen untersucht und bestenfalls ausgeglichen werden, sofern systematische Unterschiede zwischen Antwortenden und Nicht-Antwortenden vorliegen. Damit Leser die Existenz und das Ausmaß dieser Verzerrung einschätzen und bei ihrer Meinungsbildung bzw. Entscheidungsfindung berücksichtigen können, muss *Item-Nonresponse-Error* transparent kommuniziert und analysiert werden.

#### Item-Nonresponse im Glücksspiel-Survey 2021

Allerdings scheint Item-Nonresponse auf den ersten Blick im Glücksspiel-Survey nicht vorzuliegen: Die Grundgesamtheit  $n$  variiert nicht zwischen den einzelnen Fragen. Bei der Telefonbefragung werden

<sup>12</sup>Diese wird leider nicht genannt.

7.501 „vollständige“ (Buth, Meyer und Kalke 2022, S. 18) Interviews und bei der Online-Befragung „10 % [...] vollständig ausgefüllt[e]“ (Buth, Meyer und Kalke 2022, S. 417) Interviews genannt. Offensichtlich wurden von den Studienautoren keine Interviews berücksichtigt, bei denen mindestens eine Frage nicht beantwortet wurde. Dabei enthält der Glücksspiel-Survey durchaus einige Fragen(batterien), die dem wissenschaftlichen Konsens (Tourangeau, Rips und Rasinski 2000) nach als sensibel und demnach als besonders anfällig für Item-Nonresponse gelten, unter anderem Fragen zum Glücksspielverhalten, zum Alkoholkonsum oder zur psychischen Belastung. Beispielsweise werden folgende sensible Fragen gestellt<sup>13</sup>:

- Haben Sie andere Ihre Glücksspielschulden bezahlen lassen (d.h. sich aus der Klemme helfen lassen), wenn Sie wegen Ihrer finanziellen Lage verzweifelt waren?
- Haben Sie mehrmals ohne Erfolg versucht, mit dem Glücksspielen aufzuhören?
- Haben Sie Familienmitglieder, Freunde, Mitarbeiter oder Lehrer oft angelogen, wenn es um das Ausmaß Ihres Glücksspielens oder um die Höhe Ihrer Spielschulden ging?
- Haben Sie wegen Ihres Glücksspielens Arbeits- oder Schultage, soziale Aktivitäten oder Familienaktivitäten versäumt?
- Wie oft haben Sie im letzten Jahr an einem Tag sechs oder mehr alkoholische Getränke getrunken?
- Wenn Sie an einem Tag Alkohol trinken, wie viele alkoholhaltige Getränke trinken Sie dann typischerweise?
- Wie oft waren Sie in den vergangenen vier Wochen so niedergeschlagen, dass Sie nichts aufheitern konnte?
- Wie oft fühlten Sie sich in den vergangenen vier Wochen entmutigt und traurig?

Alle diese Fragen betreffen sensible Bereiche und haben daher eine höhere Wahrscheinlichkeit, keine (oder eine falsche) Antwort hervorzurufen, beides wurde bereits vielfach in der Literatur festgestellt (Tourangeau und Yan 2007). Das Entfernen aller unvollständigen Interviews eliminiert zwar auf den ersten Blick den Item-Nonresponse-Bias, vergrößert dafür aber den Unit-Nonresponse-Bias umso stärker. Denn nun gehören auch die Personen, die die Befragung abgebrochen haben oder bis auf eine Frage

<sup>13</sup>Zumindest sind dies Fragen aus den im Survey genutzten bzw. erwähnten Frageinstrumenten. Da der Fragebogen nicht veröffentlicht wurde und auch nicht geteilt wurde, können prinzipiell im Survey auch andere Fragen genutzt worden sein.

– zum Beispiel zum sensiblen Thema riskanter Alkoholkonsum – alle anderen Fragen beantwortet haben, zu den Nicht-Antwortenden und werden nicht mehr in den Befragungsergebnissen berücksichtigt.

Diese Vorgehensweise im Glücksspiel-Survey 2021 dürfte zu Verzerrungen geführt haben, über die der Survey allerdings keine Aussagen trifft. Dabei wäre es möglich, zu untersuchen, wer die Interviews abbricht oder sensible Fragen nicht beantwortet. Damit ließe sich nicht nur der Item-Nonresponse-Bias quantifizieren und einschätzen, sondern es ließen sich auch verschiedene statistische Methoden (zum Beispiel Imputation, Gewichtung) zur Mitigation dessen einsetzen (für einen Überblick, siehe Lohr (2021)). Diese Möglichkeiten haben die Autoren des Glücksspiel-Surveys 2021 nicht genutzt und die unvollständigen Interviews aus ihrer Stichprobe mutmaßlich entfernt, was im Vergleich zu den anderen Optionen zur größtmöglichen Verzerrung führt (Meinfelder 2014). Dies wird dem Leser allerdings nicht kommuniziert – liest man den Glücksspiel-Survey, scheint kein *Item Nonreponse* vorzuliegen, wodurch dem Leser auch keine daraus resultierenden Verzerrungen bewusst sind. Dieses intransparente Vorgehen, insbesondere aufgrund der rustikalen Entfernung aller nicht vollständigen Interviews (was definitiv zu Verzerrungen führt), ist ein erheblicher Kritikpunkt und entspricht nicht wissenschaftlichen Qualitätsstandards (Deutsche Forschungsgemeinschaft 2019). Gerade in einer Studie, die erwartungsgemäß derartige politische und mediale Aufmerksamkeit erfährt, sollte jede Chance genutzt werden, Fehlerquellen genauestens zu analysieren und resultierende Fehler mit allen Mitteln abzumildern.

#### 4.2.4 (Self-)Selection-Error

Knapp 40 % der Befragten im Glücksspiel-Survey stammen aus zwei Online-Panels. Solche nicht-probabilistischen Stichprobenerhebungen, auch *Nonprobability Sampling* genannt, beziehen sich auf Stichprobenerhebungsmethoden, bei denen die Elemente der Stichprobe nicht zufällig teilnehmen. Das heißt, die Partizipationswahrscheinlichkeit der Mitglieder der Zielpopulation ist unbekannt und/oder nicht für jedes Individuum positiv, was zu einer (unbekannten) Selektionsverzerrung, also einem *Selection-Error*, führt. Konkret werden die Teilnehmer beim *Nonprobability-Sampling* nicht zufällig, sondern gewöhnlich aufgrund ihrer leichten Verfügbarkeit ausgewählt, zum Beispiel in einem Online-Panel, bei dem die Teilnehmer freiwillig teilnehmen. Bei Online-Panels werden internet- und technikaffine Menschen überdurchschnittlich häufig befragt (Couper 2000) und Menschen ohne Internetzugang häufig überhaupt nicht. Bosnjak u. a. (2013) fanden Evidenz für solche systematischen Unterschiede in Online-Panels in Deutschland: Deutsche Internet-Benutzer sind jünger, gebildeter, extrovertierter, offener für neue Erfahrungen, weniger gewissenhaft und eher männlich. Diese systematischen Unterschiede führen zu Selektionsverzerrungen. Ebenso sind beim *Nonprobability-Sampling* die geschätzten Populationsparameter auch bei mehrfacher Ziehung verzerrt, da – im Gegensatz zum *Probability-Sampling* – durch die unterschiedlichen Partizipationswahrscheinlichkeiten keine durchschnittliche Angleichung an den Populationsparameter zu erwarten ist (Lohr 2021). Folglich kann auch kein Stichprobenfehler äquivalent zum Probability Sample errechnet werden, wodurch ein modellbasierter Ansatz mit starken Annahmen über die Eigenschaften der Befragten gewählt werden muss (Lohr 2021). Der Selektionsfehler ist insbesondere dann problematisch, wenn der Faktor, der zu einer höheren Teilnahmebereitschaft am Panel führt, mit dem zu untersuchenden Faktor korreliert (Lohr 2021; Bosnjak u. a. 2013). Wenn zum Beispiel vor allem junge Menschen am Panel teilnehmen und in einer Befragung Einkommen untersucht werden soll, werden diese Daten natürlich ein verzerrtes und nicht der Realität entsprechendes Bild der Einkommensverteilung zeigen.

Ein eindrucksvolles Beispiel für solche verzerrten Ergebnisse (und deren unkritische Rezeption) war die vergangene Studie von Plan International zu den Einstellungen junger Männer, nach der ein hoher Anteil Gewalt gegenüber Frauen befürwortet (Plan International 2023). Die Stichprobe basierte zu 100 % auf einem Online-Panel und lieferte entsprechend verzerrte Informationen (Tagesschau 2023) beziehungsweise (viel) höhere Angaben als ähnliche Studien<sup>14</sup>.

Dass *Nonprobability Samples* durchschnittlich eine geringere Präzision aufweisen als *Probability Samp-*

---

<sup>14</sup>Zur Kritik an der Methodik der Plan-International-Studie im Speziellen und an Online-Surveys im Allgemeinen siehe auch Schnell (2023)

les, ist Konsens in der Wissenschaft (Dutwin und Buskirk (2017), MacInnis u. a. (2018), Lavrakas u. a. (2022), Yeager u. a. (2011), Baker, Blumberg u. a. (2010)), nur in der Wahlforschung sind die Ergebnisse teilweise sogar besser als bei klassischen *Probability Samples* (Baker, Blumberg u. a. 2010). Bessere Ergebnisse liefern Online-Panels, bei denen die Teilnehmer mittels *Probability-Sampling* aus der Gesamtpopulation ausgewählt werden (zum Beispiel mit einer Registerstichprobe) und Teilnehmer ohne Internetzugang entsprechende Technik zur Verfügung gestellt bekommen (zum Beispiel: Blom, Gathmann und Krieger (2015)) – dennoch ist auch hier die Verzerrung und Repräsentativität stark von der Nonresponse-Quote abhängig (Enderle, Münnich und Bruch 2013).

Dennoch sind Online-Panels durch ihre einfache Durchführbarkeit und geringeren Kosten durchaus attraktiv für viele Forscher und Umfrageinstitute. Ob *Nonprobability Samples* eine geeignete Methode sind, entscheidet sich daher anhand der Ziele einer Befragung. So stellte die Taskforce zu *Nonprobability Samples* der *American Association for Public Opinion Research* mit Verweis auf den Report einer früheren Taskforce zu Online-Panels (Baker, Blumberg u. a. 2010) fest: „*such surveys have value for some types of research, but researchers „should avoid non-probability opt-in panels when a key research objective is to accurately estimate population values ... [and] claims of ‘representativeness’ should be avoided when using these sample sources.*“ (Baker, Brick u. a. 2013, S. 93). Online-Panels sind also nicht zur präzisen Schätzung von Anteilswerten in der Bevölkerung geeignet und können keine „Repräsentativität“ beanspruchen. Nicht-probabilistische Methoden werden daher häufig in explorativen oder qualitativen Forschungssituationen eingesetzt, wenn das Ziel nicht darin besteht, quantitativ genaue Schätzungen für die gesamte Population zu erstellen, sondern es vornehmlich darum geht, schnelle Einsichten zu erlangen oder Experimente durchzuführen. Sollten Online-Panels dennoch genutzt werden, ungeachtet ihrer Eignung für die Studienziele, ist Transparenz essenziell bzw. noch essenzieller als beim *Probability-Sampling* (Baker, Brick u. a. 2013). Informationen über die genutzten Methoden (Rekrutierung, Modellannahmen, etc.) müssen dem Leser mitgeteilt werden, damit dieser die Qualität der Daten beurteilen kann (Baker, Brick u. a. 2013).

Zusammengefasst sollten *Nonprobability Samples* aufgrund von Selektionsverzerrungen nicht genutzt werden, wenn präzise oder repräsentative Schätzungen für die Bevölkerung – abgesehen von Wahlprognosen – das Ziel einer Befragung sind (Baker, Brick u. a. 2013). Unabhängig von ihrer Eignung für das Ziel einer Befragung ist die transparente Kommunikation aller notwendigen Informationen zur Bewertung der Stichprobe unerlässlich (Baker, Brick u. a. 2013).

### (Self-)Selection-Error im Glücksspiel-Survey 2021

Wie bereits mehrfach erwähnt, müssen angesichts der Zielsetzung des Glücksspiel-Surveys 2021 hohe Qualitätsmaßstäbe angelegt werden (Münnich 2023), insbesondere bezüglich der Präzision der Ergebnisse (Baker, Brick u. a. (2013), Lohr (2021)). Den obigen Ausführungen folgend, ist *Nonprobability-Sampling* und daher auch das *Mixed-Mode-Design* keine geeignete Methode für den Glücksspiel-Survey, da aufgrund der starken potenziellen Verzerrungen präzise Schätzungen nicht zu erwarten sind und nicht auf die Gesamtbevölkerung inferenziert werden sollte (bzw. dies methodisch schlicht unmöglich ist).

Im Survey werden für die Online-Befragung die Online-Panels der Anbieter Gapfish und Dynata genutzt. Gapfish betreibt verschiedene Panels, die hier beispielhaft genannt werden sollen (GapFish 2021):

**Unsere Panel|welt**

**Starke Partner mit echten Mehrwerten auch für Sie!**

Neben unserem Panel „EntscheiderClub“, das wir seit 2013 in Deutschland, Österreich und der Schweiz betreiben, haben wir zahlreiche andere Panelmarken aufgebaut, u.a. auch im Kundenauftrag mit eigenem Branding, mit deren Betrieb und Vermarktung die GapFish GmbH betraut ist.

Die Panels unterscheiden sich hinsichtlich der Rekrutierungsquellen und auch hinsichtlich der Teilnehmerincentivierung.

**GAPFISH**  
WISSEN, WAS WIRKLICH ZÄHLT

Folgende Plattformen stehen aktuell zur Verfügung:

- EntscheiderClub**  
Unsere proprietären Panels mit monetärem Incentive  
<https://www.entscheiderclub.de/> <https://www.entscheiderclub.at/> <https://www.entscheiderclub.ch/>
- TrendsetterClub**  
Unsere proprietären Panels mit Punkte-Incentivierung  
<https://www.trendsetterclub.com/>
- boost Panel**  
Unsere proprietären Panel, bei dem das Incentive gespendet wird  
<https://www.boost-panel.de/>
- LOVE2SAY**  
Die Panels der Mediengruppe RTL (Rekrutierung zumeist durch TV-Kampagnen)  
<https://www.love2say.de/> <https://www.ilovemymedia.de/> <https://www.ilovemymedia.at/> <https://www.mynewspanel.de/>
- Kundenstimme**  
Das Panel von REWE (Rekrutierung durch REWE-Gruppe)  
<https://www.kundenstimme.com/>
- DIE HÖRER**  
Das RMS-Panel (Rekrutierung zumeist über Radiokampagnen)  
<https://www.diehoerer.de/>
- GALLUPFORUM**  
Das Panel des Österreichischen Gallup-Instituts  
<https://www.gallupforum.at/>
- POINTOU**  
Das Mediakraft Networks-Panel (Rekrutierung durch Social Media Kampagnen)  
<https://www.point-of-u.de/>

Abbildung 4.6: Rekrutierungsarten der Gapfish-Panels (Quelle: GapFish (2021))

Bei den drei proprietären (= Gapfish-eigenen) Panels, von denen (vermutlich) für den Glücksspiel-Survey 2021 die Befragten rekrutiert wurden, kann sich jeder anmelden. Beim sogenannten „TrendsetterClub“ erhalten Panelisten als Anreiz für das Werben weiterer Personen bis zu 2 Euro (Stand 11.07.2023). Auch beim zweiten genutzten Panel von Dynata ist die Teilnahme am Panel für jeden möglich. Bei den genutzten Panel-Anbietern haben die Panelisten also selbst entschieden, ob sie am

Panel teilnehmen. Ein problematischer Aspekt bei derlei Panels ist die Intransparenz über die genutzten Rekrutierungswege, die maßgeblich die Zusammensetzung und Verzerrung des Samples bestimmen (Baker, Brick u. a. 2013): Wird beispielsweise hauptsächlich über das Internet rekrutiert, werden sich mehrheitlich internetaffine Menschen im Panel befinden. Wird wie beim „TrendsetterClub“ (unter anderem) *Snowball-Sampling* benutzt, wird die Homogenität des Panels weiter verstärkt. Konkrete Informationen über die Rekrutierungsstrategien in beiden Panels sind in der Glücksspiel-Survey 2021 nicht zu finden, im Methodenbericht werden nur folgende allgemeine und wenig aussagekräftige Punkte genannt:

Zur Qualitätssicherung sind von den Panelanbietern eine Reihe von Maßnahmen umgesetzt worden. Insbesondere zu nennen sind in diesem Zusammenhang:

- breite Rekrutierungskanäle (online und offline)
- heterogene Panelstruktur
- Double-opt-in-Registrierung
- Identitätssicherung durch Bankverbindung
- zuverlässige Cash-Incentive-Programme
- ständige Überprüfung der Antwortkonsistenz
- ausschließliche Nutzung der Daten für Markt- und Sozialforschungszwecke
- regelmäßige Bereinigung der Panels
- kein Over- oder Underentertainment der Teilnehmer\*innen

Abbildung 4.7: Qualitätssicherungsmaßnahmen der genutzten Panels laut Glücksspiel-Survey 2021

Betrachtet man beispielsweise die Aussage „*breite Rekrutierungskanäle (online und offline)*“, so wird damit zwar die Information geliefert, dass Panelisten online und offline rekrutiert werden. Wie genau dies geschieht (Snowball-Sampling, Werbung auf (speziellen) Social-Media-Seiten, etc.), bleibt allerdings offen. Ähnliches gilt für die angebliche „*Heterogene Panelstruktur*“, für die kein Beleg geliefert wird. Die übrigen genannten Maßnahmen liefern auch keine diesbezüglichen Informationen. Dies erschwert es dem Leser erheblich, die Aussagekraft der Daten des Glücksspiel-Surveys 2021 zu beurteilen, da die Zusammensetzung der Panels – eine nicht zu vernachlässigende Quelle für eine mögliche Verzerrung – nicht evaluiert werden kann. Im Ergebnis lässt sich die Aussagekraft des Anteils der Menschen mit Glücksspielproblemen kaum beurteilen: Es wäre unter anderem plausibel, dass internetaffine Personen, die eher am Panel teilnehmen, auch eher Online-Glücksspiel spielen – ohne Daten über das Panel und insbesondere über die befragten Personen sind derlei Prüfungen allerdings nicht möglich. Es ist diesbezüglich unerheblich, ob die Panelanbieter solche Informationen den Autoren des Surveys nicht zur Verfügung gestellt haben (oder wollten) oder ob die Autoren diese Informationen aus unbekanntem Gründen nicht in den Survey aufgenommen haben – eine Befragung mit den eingangs

genannten Zielen erfordert die Bereitstellung von Informationen, die zur Bewertung der Datenqualität notwendig sind. Dadurch ist es weder im vorliegenden Gutachten noch für jeden anderen Leser möglich, die Güte der Zusammensetzung und Größe der Verzerrung durch die beiden Panels – wie in der Literatur gefordert (zum Beispiel Münnich (2023) und Baker, Brick u. a. (2013)) – zu bewerten. Aus ebendiesem Grund werden die Ergebnisse des Surveys in der vom Bundesministerium für Wirtschaft und Klimaschutz in Auftrag gegebenen Studie „Wissenschaftliche Studie zur Vorbereitung der Evaluierung der Sechsten Verordnung zur Änderung der Spielverordnung“ nur eingeschränkt verwendet (Bühringer u. a. 2023).

Es ist anzumerken, dass spezifische Forschung existiert, die den *Selection-Error* bei Befragungen im Glücksspielbereich untersucht: So stellte Sturgis und Kuha (2022) fest, dass Online-Befragungen aufgrund von *Selection-Error* höhere Anteile von Menschen mit Glücksspielproblemen messen.

Zusammenfassend nutzt der Glücksspiel-Survey zu knapp 40 % *Nonprobability-Sampling* durch zwei Online-Panels, obwohl diese Methode nicht dem notwendigen hohen Qualitätsstandard, der sich aus der Zielsetzung des Surveys ergibt, entspricht. Das Ausmaß dieses Problems kann allerdings nicht seriös quantifiziert werden, da der Survey die dafür notwendigen Informationen – entgegen der bei diesen Zielen erforderlichen Transparenz – nicht bereitstellt. Auf Seite 15 zeigt der Glücksspiel-Survey zwar die Verteilung von Geschlecht, Alter, Bildung und Migrationshintergrund in beiden Befragungsarten, zur Einschätzung der Verzerrung wäre allerdings die Verteilung der Glücksspielvariablen essenziell. So kann aktuell nur spekuliert werden, wie sich die Personen mit Glücksspielstörung über beide Befragungsarten verteilen, wodurch nicht evaluiert werden kann, inwieweit sich die Verzerrung der Online-Stichprobe auf die Gesamtstichprobe auswirkt. Eine Stichprobe ohne *Nonprobability-Sampling* wäre in diesem Fall aller Wahrscheinlichkeit nach weniger verzerrt gewesen (Lohr 2021; Baker, Blumberg u. a. 2010; Baker, Brick u. a. 2013; Sturgis und Kuha 2022).

### 4.2.5 Measurement-Error

*Measurement-Error* (Messfehler) ist dann vorhanden, wenn die in der Befragung gemessene Angabe für eine Größe (beispielsweise Glücksspielprobleme) nicht mit dem „wahren“ Wert dieser Größe übereinstimmt. Dieser Messfehler kann sowohl durch den Interviewer (in persönlichen Befragungen), als auch durch den Fragebogen sowie den Befragten verursacht werden und variiert je nach Befragungsart.

Interviewereffekte beschreiben eine durch die Interviewenden induzierte Verzerrung in den Antworten der Befragungsteilnehmer. Diese tritt beispielsweise in Telefonsurveys auf, wenn Befragte den Interviewer besonders sympathisch oder vertrauenswürdig finden und daher auskunftsfreudiger über negative Ereignisse sind. Gleichsam kann auch das Gegenteil eintreten. Befragte geben dann in persönlichen Befragungen gegenüber unsympathisch empfundenen Personen eher wenig Auskunft. Durch Erklärungen von Fragen kann ein Interviewer auch zum besseren Verständnis und dadurch zu akkurateren Antworten führen.

Ein besonders in Befragungen zum Thema Glücksspiel allerdings nicht außer Acht zu lassender Faktor bei *Measurement-Error* ist soziale Erwünschtheit in den Antworten von Teilnehmern – unabhängig von der Interviewform. In unzähligen Studien wurde nachgewiesen, dass Befragte dazu tendieren, unwahre Angaben über sensible Themen wie Alkoholkonsum, Glücksspiel oder Gehalt zu machen (für Beispiele, siehe (Lohr 2021; Tourangeau, Rips und Rasinski 2000; Tourangeau und Yan 2007)), um einer sozialen Norm zu entsprechen und eventuelles Fehlverhalten nicht zuzugeben. Am stärksten tritt dies in persönlichen Befragungen vor Ort auf – sofern keine geeigneten Gegenmaßnahmen getroffen werden – oder, wenn Angehörige bei der Beantwortung der (Telefon-) Befragung im Raum sind. Ein ähnlicher potenziell problematischer Faktor sind sensible Fragen (*sensitive questions*), die zum Entstehen des Messfehlers beitragen können. Diese betreffen als privat oder als persönlich empfundene Themengebiete wie Steuern oder gesundheitliche Themen, die Befragte aus Sorge vor unzureichendem Daten- und Persönlichkeitsschutz nicht oder bewusst falsch beantworten. Für eine Übersicht über sensible Fragen innerhalb der wenigen Lesern bekannten Frageinstrumente des Glücksspiel-Surveys, siehe 4.2.3.

Auch der verwendete Fragebogen kann dazu beitragen, dass Befragte Antworten geben, die von den wahren Antworten abweichen. So kann durch eine suggestive Formulierung von Fragen durch Vorstellen von Aussagen, wie „Wissenschaftler sagen, dass übermäßiger Konsum von Schokolade krank macht. Wie würden Sie ihren Schokoladenkonsum einschätzen?“ gezielt oder auch unbewusst das Antwortverhalten der Befragten beeinflusst werden. Oft reicht es aus, anstelle von Fragen zu Präferenzen

lediglich nach Zustimmung zu einer positiven/negativen Aussage zu fragen, um das Antwortverhalten in Richtung der Aussage zu beeinflussen und einen Messfehler hervorzurufen („Schokolade ist ungesund – stimmen Sie dieser Aussage zu?“; „Wie würden Sie die gesundheitlichen Auswirkungen von Schokolade einschätzen?“). Auch werden vage Formulierungen wie „manchmal“ oder „oft“ von verschiedenen Individuen möglicherweise unterschiedlich interpretiert (vgl. die Beispielfrage „Haben Sie Familienmitglieder, Freunde, Mitarbeiter oder Lehrer *oft* angelogen, wenn es um das Ausmaß Ihres Glücksspiels oder um die Höhe Ihrer Spielschulden ging?“.)

### Measurement-Error in der Glücksspiel-Survey 2021

Von einem (nicht zu vernachlässigenden) *Measurement-Error* allein aufgrund der Befragungsform ist in der Glücksspiel-Survey 2021 aufgrund des professionellen Dienstleisters nicht auszugehen: Einerseits schulen Befragungsinstitute ihre Interviewer für telefonische Befragungen professionell, andererseits ist bei Internetbefragungen keine große Beeinflussung anzunehmen. Anders sieht es im Survey bezüglich des Fragebogens aus: Da die Autoren des Surveys entsprechende Materialien nicht zur Verfügung gestellt haben, kann ein Messfehler aufgrund der Fragen nicht eingeschätzt werden. Erneut erschwert diese Intransparenz sowohl dem Leser als auch dem Gutachten die Einschätzung der vorhandenen Verzerrung im Glücksspiel-Survey 2021. Dies entspricht nicht guter wissenschaftlicher Arbeit (Deutsche Forschungsgemeinschaft 2019) und ist somit zu kritisieren.

### 4.2.6 Processing-Error

Der *Processing-Error* entsteht, wenn die Antworten von Befragten nicht korrekt in den Datensatz übertragen werden (Lohr 2021). Dieser Fehler wird durch die mittlerweile standardmäßige automatisierte Datenverarbeitung in telefonischen und internetbasierten Befragungen seltener, ist allerdings immer noch möglich. Denkbar sind verschiedene Fehlerquellen, zum Beispiel wenn sich ein Interviewer vertippt. Solche Fehler können nach der Datensammlung nicht mehr gefunden und korrigiert werden. Fehler bei der Datenverarbeitung und -analyse können allerdings auch ex post gefunden und korrigiert werden. In der modernen Umfrageforschung werden die Rohdaten für gewöhnlich in geeigneten Programmen (zum Beispiel SPSS, STATA, R) aufbereitet und weiterverarbeitet. Hierbei kann (un)bewusst jeder erdenkliche Fehler passieren, zum Beispiel kann ein Schwellenwert falsch gesetzt, können Variablen vertauscht oder auch ein simpler Rechenfehler passiert sein. Beispiele dafür gibt es in der Wissenschaft genug, so reformatierte beispielsweise das Tabellenkalkulationsprogramm Excel die Namen von Genen in 20 % der von den Autoren untersuchten Studien mit Exceldateien, was zu einer Verfälschung der Ergebnisse führte (Ziemann, Eren und El-Osta 2016).

#### Processing-Error im Glücksspiel-Survey 2021

Derlei Fehler könnten theoretisch auch im Glücksspiel-Survey 2021 vorgekommen sein. Da die Studienautoren entgegen den Empfehlungen guter wissenschaftlicher Praxis (Deutsche Forschungsgemeinschaft 2019) (dazu ausführlich Kapitel 8 Transparenz) für das vorliegende Gutachten eine Replikation und Überprüfung der Ergebnisse des Glücksspiel-Surveys 2021 nicht ermöglicht haben, kann die Fehlerfreiheit der Studie in diesem Gutachten nicht überprüft werden. Dies ist insbesondere im Hinblick auf die Ziele des Surveys, die nach Münnich (2023) und Baker, Brick u. a. (2013) Transparenz erfordern, kritikwürdig. Die Überprüfung und Replikation von Studienergebnissen ist keineswegs Ausdruck von Misstrauen, sondern ist im Gegenteil Teil des üblichen wissenschaftlichen Prozesses gemeinsamer Wahrheitsfindung und Qualitätssicherung. In qualitativ hochwertigen *Nature Journals* werden daher entsprechende Anforderungen an Veröffentlichungen gestellt: „*A condition of publication in a Nature Portfolio journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.*“ (Nature 2023).

Zusammengefasst kann die Existenz des *Processing-Error* im Glücksspiel-Survey 2021 aufgrund der nicht zur Verfügung gestellten Materialien nicht überprüft werden, was insbesondere aufgrund der Ziele des Surveys (Politikinformation und -beratung) kritikwürdig ist.

## 5 | Adjustierung von Verzerrungen

### Executive Summary

*Die existierenden Verzerrungen in den erhobenen Daten versuchen die Autoren des Glücksspiel-Surveys 2021 mit einer sehr simplen Poststratification-Gewichtung auszugleichen, die dies allerdings – wenn überhaupt – nur im telefonisch durchgeführten (Probability-Sampling)-Teil der Befragung und dort auch nur in begrenztem Ausmaß korrigieren kann. Angebracht wären stattdessen komplexere Gewichtungsverfahren (Ermittlung von Response Propensities in Kombination mit weiteren Modellannahmen), die jedoch bei Nonprobability Samples, wenn überhaupt, auch nur einen geringen Teil der Verzerrungen korrigieren können. Grundsätzlich hätten in der Glücksspiel-Survey 2021 von Beginn an mehr relevante Variablen (zum Beispiel Arbeitslosigkeit) abgefragt werden müssen, die die Effektivität der dringend notwendigen Gewichtung gesteigert hätten. Die Mode-Gewichtung im Survey überzeugt aufgrund der veralteten Literatur und der genutzten Aggregatdaten ebenso wenig. Zudem bleibt offen, warum die Survey-Autoren die hochgerechneten Ausgaben für Glücksspiel nur mit dem Umsatz des DLTB und nicht zusätzlich mit dem Umsatz anderer Spielformen vergleichen.*

Bei Befragungen, die auf Stichproben basieren, gibt es etliche Quellen möglicher Verzerrungen, die sich auf die Repräsentativität und somit die Präzision der Daten auswirken (Lohr 2021). Normalerweise sind, beispielsweise durch Selektionsverzerrung oder Nonresponse-Verzerrung, bestimmte Gruppen in der Stichprobe im Vergleich zur Population unter- oder überrepräsentiert. Somit ist die Stichprobe nicht mehr repräsentativ für die Gesamtpopulation und liefert entsprechend verzerrte und unpräzise Ergebnisse. Grundsätzlich gilt, dass eine schlecht designte Befragung auch ex post nicht mehr zu präzisen Ergebnissen führen kann (Lohr 2021). Gleichwohl gibt es verschiedene Möglichkeiten, die Verzerrungen in Befragungsdaten zumindest in Teilen zu korrigieren. Im Nachfolgenden wird vornehmlich die

Gewichtung, genauer gesagt *Poststratification*, beleuchtet, da dieses Vorgehen im Glücksspiel-Survey 2021 Anwendung findet.

*Poststratification* ist eine Technik, bei der die Stichprobe nach bestimmten Hilfsvariablen in *Poststrata* (z.B.: Männer – Frauen) aufgeteilt und dann so gewichtet wird, dass die Verteilung der Hilfsvariablen in der gewichteten Stichprobe der Verteilung in der Bevölkerung entspricht. Diese Hilfsvariablen können beispielsweise Geschlecht, Alter oder Bildungsniveau sein – für jede dieser Variablen müssen dann allerdings auch die entsprechende Verteilung in der Population bekannt sein. Für jedes *Poststratum* wird dann ein Gewichtungsfaktor berechnet als das Verhältnis der Anteile in der Population zur Stichprobe. Die Antworten der Befragten in jeder Kategorie werden dann mit diesem Faktor gewichtet. Diese Vorgehensweise kann unter bestimmten Umständen die Verzerrung verringern, allerdings nur für die zur Gewichtung verwendeten Hilfsvariablen (Lohr 2021). Die Verzerrung, die aus anderen Quellen stammt, wird nicht korrigiert (Lohr 2021). Dies soll an einem Beispiel illustriert werden.

Angenommen, es soll das Einkommen in der Bevölkerung untersucht werden, und die Hilfsvariable Geschlecht verteilt sich in der Stichprobe ( $S$ ) und der Population ( $P$ ) jeweils wie folgt:

- In der Stichprobe ( $\hat{p}$ ): 60 % Männer und 40 % Frauen.
- In der Bevölkerung ( $p$ ): 49 % Männer und 51 % Frauen.

Dann würden die Gewichtungsfaktoren ( $W$ ) für jedes *Poststratum* wie folgt berechnet:

Für Männer:

$$W_{\text{Männer}} = \frac{p_{\text{Männer}}}{\hat{p}_{\text{Männer}}} = \frac{0.49}{0.60} \approx 0.817 \quad (5.1)$$

Für Frauen:

$$W_{\text{Frauen}} = \frac{p_{\text{Frauen}}}{\hat{p}_{\text{Frauen}}} = \frac{0.51}{0.40} \approx 1.275 \quad (5.2)$$

Dabei steht  $p_{\text{Männer}}$  für den Anteil der Männer in der Bevölkerung und  $\hat{p}_{\text{Männer}}$  für den Anteil der Männer in der Stichprobe, analog für die Frauen. Die Gewichtungsfaktoren  $W_{\text{Männer}}$  und  $W_{\text{Frauen}}$  sind dann die Faktoren, mit denen die Antworten der Befragten in der entsprechenden Kategorie multipliziert werden. Die Antworten von Männern werden also weniger gewichtet (mit 0.817 multipliziert), da Männer in der Befragung überrepräsentiert werden. Die Antworten von Frauen werden hingegen

höher gewichtet. Damit soll die Verzerrung in der Stichprobe korrigiert werden.

*Poststratification* ist die einfachste Methode einer derartigen Korrektur. Sie beruht jedoch auf zwei starken (und stets fragwürdigen) Annahmen: Erstens, dass alle Mitglieder der jeweiligen *Poststrata* die gleiche Wahrscheinlichkeit haben, eine (Um)Frage zu beantworten, also zum Beispiel alle Frauen die gleiche Inklusionswahrscheinlichkeit besitzen. Zweitens, dass der Wert der Zielvariable, hier Einkommen, für alle Mitglieder eines *Poststratum*s gleich ist, also dass beispielsweise alle Frauen das gleiche Einkommen haben. Ein solches Modell wird alle Unterschiede innerhalb der Geschlechter (zum Beispiel aufgrund des Alters) ignorieren und entsprechende Verzerrungen nicht korrigieren (Lohr 2021). Die gewichtete Stichprobe wird daher immer noch verzerrt sein, beispielsweise aufgrund einer verzerrten Altersstruktur. Nur wenn eine Gewichtung alle Faktoren, die sowohl mit der Teilnahmewahrscheinlichkeit als auch mit der Zielvariable korreliert sind, berücksichtigt, kann die komplette Verzerrung korrigiert werden (McPhee u. a. 2022).

Grundsätzlich hängt der Erfolg einer Gewichtung also von den Gewichtungsvariablen (Hilfsvariablen) und deren Korrelation mit der Zielvariable sowie der Teilnahmewahrscheinlichkeit ab (Lohr 2021; Valliant 2020). Dies ist (vereinfacht) in Tabelle 5 zu sehen:

	Auxiliary variables predictive of outcome		
		No	Yes
Auxiliary variables predictive of selection	No	No effect	Reduce variance No effect on bias
	Yes	Increase variance No effect on bias	Reduce variance Reduce bias

Abbildung 5.1: Effekt einer Gewichtung mit Hilfsvariablen von MCPhee u. a. (2022)

Eine Hilfsvariable kann dann eine Verzerrung korrigieren, wenn sie sowohl mit der Teilnahmewahrscheinlichkeit der Teilnehmer als auch mit der Zielvariable korreliert ist. Ist sie hingegen mit keiner der beiden Variablen korreliert, hat eine Gewichtung keinen positiven Effekt (für eine ausführliche Diskussion, siehe MCPhee u. a. (2022)). Eine simple Gewichtung reicht also selten aus, existierende Verzerrungen zu korrigieren. Doch selbst eine ausdifferenzierte Gewichtung kann existierende Verzerrungen nur mäßig korrigieren (Lohr 2021; Mercer, Lau und Kennedy 2018; Valliant 2020; MCPhee

u. a. 2022). Mercer, Lau und Kennedy (2018) stellen zum Beispiel fest, dass selbst die besten Adjustierungsstrategien bei Online-Panels nur 30 % der Verzerrungen korrigieren können. Ergänzend fanden Fahimi u. a. (2015) heraus, dass selbst nach der Gewichtung sich Befragte in *Nonprobability Samples* noch in mehreren Aspekten (politisches Verhalten, Einkaufsverhalten, etc.) von Befragten in *Probability Samples* unterscheiden.

Unabhängig von der gewählten Gewichtung ist es essenziell, transparent die Gewichtungsmethode und damit getroffenen Annahmen zu kommunizieren. Autoren müssen also begründen, auf welchen Annahmen diese Gewichtung beruht und diese mit Evidenz belegen (Lohr 2021). Wird zum Beispiel nach Geschlecht und Alter gewichtet, um Einkommen zu untersuchen, müssten Studienautoren belegen, dass Menschen mit gleichem Geschlecht und Alter auch ein ähnliches Einkommen haben. Zudem werden bei komplexen Surveydesigns häufig verschiedene Gewichtungen ausprobiert und auch mehr als eine Gewichtung kommuniziert, insbesondere wenn die Gewichtung die Ergebnisse stark beeinflusst.

Zusammenfassend lässt sich sagen, dass *Poststratification* eine geeignete Methode zur Korrektur von Verzerrungen sein kann, sofern nach den richtigen Hilfsvariablen gewichtet wird (McPhee u. a. 2022). Dies gilt jedoch nur für *Probability Samples* mit keinem oder geringem *Nonresponse* (Lenau u. a. 2021). Liegt kein (annähernder) Full-Response vor, kann bei *Probability Samples* auf komplexere Gewichtsverfahren (Ermittlung von Response-Propensities in Kombination mit weiteren Modellannahmen) zurückgegriffen werden; bei *Nonprobability Samples* führen selbst Response-Propensity-Ansätze nur in Ausnahmefällen und mit starken Modellannahmen zum Ziel. Das heißt, selbst eine nach bestem Wissen gewählte und methodisch anspruchsvolle, modellbasierte Gewichtung kann Verzerrungen nur teilweise korrigieren (Lohr 2021; MCPhee u. a. 2022) und wird in hohem Maße von subjektiven (d.h. nicht zwingend richtigen) Annahmen beeinflusst. Unabhängig von der gewählten Gewichtung ist eine transparente Kommunikation der zugrundeliegenden Annahmen essenziell, damit der Leser die Güte der Gewichtung einschätzen kann. Hat die Gewichtung einen großen Einfluss auf die Ergebnisse, sollten auch alternative Gewichtungen sowie die ungewichteten Ergebnisse präsentiert werden.

## Adjustierung im Glücksspiel-Survey 2021

Im Glücksspiel-Survey 2021 wird eine Gewichtung genutzt, um existierende Verzerrungen der Daten (unter anderem wurden junge Menschen intentional überrekrutiert) zu adjustieren. Relevant sind insbesondere die genutzte *Poststratification*, also die Gewichtung nach Haushaltsgröße, Alter, Geschlecht, Schulabschluss (Abitur) und Bundesland, sowie die Mode-Gewichtung, also die unterschiedliche Ge-

wichtung der telefonisch und online Befragten.

Für die *Poststratification* wurde die Stichprobe anhand von bekannten Verteilungen der oben genannten Hilfsvariablen für Deutschland<sup>1</sup> gewichtet. Wie ausführlich im vorigen Abschnitt 5 Gewichtung ausgeführt, ist der Erfolg einer Gewichtung von der Qualität der Hilfsvariablen, also der Korrelation mit der Teilnahmewahrscheinlichkeit und den Zielvariablen, abhängig (McPhee u. a. 2022). Dafür sind also Informationen über den *Nonresponse* im Glücksspiel-Survey 2021 notwendig – dieser wird allerdings im Survey nicht explizit untersucht (bzw. dies ist nicht im Bericht dokumentiert). Allerdings werden auf S. 18 „ausgewählte Charakteristika der ungewichteten und gewichteten Stichprobe“ in einer Tabelle präsentiert (Buth, Meyer und Kalke 2022), die zeigt, dass sich in der Stichprobe des Surveys – verglichen mit der deutschen Gesamtbevölkerung in 2019 – überproportional viele Abiturienten und 16- bis 25-Jährige<sup>2</sup> befinden. Dementsprechend sind der Bildungsgrad Abitur und das Alter mutmaßlich mit *Nonresponse* korreliert, das Geschlecht hingegen nicht (oder nur marginal). Inwieweit die übrigen zur Gewichtung verwendeten Variablen (Haushaltsgröße und Bundesland) mit dem *Nonresponse* korreliert sind, kann aufgrund nicht kommunizierter bzw. nicht zur Verfügung gestellter Informationen nicht evaluiert werden. Dies wird ebenso bezogen auf die Korrelation der Hilfsvariablen mit den Zielvariablen – das vorliegende Gutachten konzentriert sich auf die Prävalenz von Glücksspielneigung und Glücksspielstörungen – betrachtet. Die Ergebnisse werden in Tabelle 5 gezeigt:

	Haushaltsgröße	Geschlecht	Alter	Schulabschluss (Abitur)	Bundesland
Korrelation mit <i>Nonresponse</i>	unbekannt	marginal	ja	ja	unbekannt
Korrelation mit Glücksspielprävalenz	unbekannt	ja	ja	ja	unbekannt
Korrelation mit Glücksspielstörung	unbekannt	ja	ja	unbekannt	unbekannt

Tabelle 5.1: Korrelation der der zur Gewichtung genutzten Hilfsvariablen mit den Zielvariablen

Zusammengefasst kann die Eignung der zur Gewichtung herangezogenen Hilfsvariablen aufgrund fehlender bzw. nicht zur Verfügung gestellter Informationen nur teilweise eingeschätzt werden. Für die Hilfsvariablen, für die Informationen vorliegen, liegt in sieben von acht Fällen eine Korrelation vor.

<sup>1</sup>Dafür wurde der Mikrozensus 2019 herangezogen.

<sup>2</sup>Diese Gruppe wurde von den Studienautoren bewusst überrekrutiert.

Insofern ist die Gewichtung nach Geschlecht, Alter und Schulabschluss mutmaßlich sinnvoll, für die Bewertung der Gewichtung nach Haushaltsgröße und Bundesland fehlen Informationen.

Allerdings bleibt fraglich, ob diese einfache Gewichtung ausreicht, um die bestehenden Verzerrungen auch nur ansatzweise zu korrigieren. Betrachtet man die einschlägige Literatur (zum Beispiel Calado und Griffiths (2016), Hodgins, Stea und Grant (2011)) bezüglich der Prävalenz von Glücksspielproblemen, treten diese insbesondere bei Personen auf, die:

- männlich,
- alleinstehend,
- geschieden,
- arbeitslos,
- jüngeren Alters sind,
- einer Minderheit angehören,
- eine niedrigere Bildung,
- ein niedriges Einkommen,
- einen Migrationshintergrund haben,
- Suchtmittel einnehmen.

Insofern wäre es aus methodischer Sicht sinnvoll, zusätzlich zu Geschlecht, Alter und Bildung auch mindestens nach Beziehungsstatus (alleinstehend, geschieden), Migrationshintergrund, Einkommen und Beschäftigungsstatus zu gewichten. Zu all diesen Hilfsvariablen liegen amtlich erhobene Verteilungen des statistischen Bundesamts vor und sie wurden auch (teilweise) von anderen Studien im Glücksspielbereich genutzt (zum Beispiel Meyer, Bischof u. a. (2015)). Allerdings bleibt unklar, ob diese Variablen (abgesehen von Migrationshintergrund) überhaupt im Glücksspiel-Survey 2021 gemessen wurden – da der Fragebogen nicht zur Verfügung gestellt wurde, kann dies nicht beantwortet werden. Weitere sinnvolle Hilfsvariablen für eine Gewichtung wären riskanter Alkoholkonsum oder Beeinträchtigung der psychischen Gesundheit, die beide im Glücksspiel-Survey 2021 gemessen wurden. Es gibt etliche weitere Studien, die diese Konzepte während der Pandemie messen (z.B. Orth und Merkel (2022) und Mauz u. a. (2023)), allerdings hätten die Autoren der Glücksspiel-Survey 2021 auch dieselben Messinstrumente wie diese aktuellen Studien nutzen müssen, um eine Vergleichbarkeit

herzustellen. Da sie dies nicht getan haben, wird im Survey der Anteil an Menschen mit Alkoholkonsum mit einer Studie aus dem Jahr 2014 – obwohl die BZgA ca. alle zwei Jahre einen Alkoholsurvey durchführt – und der Anteil an Menschen mit Beeinträchtigungen der psychischen Gesundheit mit einer Studie aus dem Jahr 2012 verglichen. Mit derart alten, insbesondere prä-pandemischen Daten ist eine sinnvolle Gewichtung schwer möglich.

Im Falle des Glücksspiel-Surveys hätte es sich auch angeboten, die mittels *Nonprobability-Sampling* durchgeführten Interviews anhand mehrerer Hilfsvariablen nach ebenjenen Variablenverteilungen in der *Probability-Sampling* Telefon-Stichprobe zu gewichten. Allerdings ist auch diese etablierte Gewichtungsmethode (Lohr 2021) von den abgefragten Hilfsvariablen abhängig und hätte einer vorausschauenden Planung bedurft. Es gibt mehrere weitere Variablen (zum Beispiel Arbeitslosigkeit), die die Gewichtung vermutlich verbessern könnten. Allerdings haben die Autoren des Glücksspiel-Surveys diese Variablen entweder nicht genutzt oder überhaupt nicht gemessen, obwohl deren Einbeziehung die Datenqualität mutmaßlich substanziell hätte verbessern können (Lohr 2021).

Weiterhin gewichten die Autoren der Glücksspiel-Survey die Online-Stichprobe zu 33 % (39 % der Befragten) und die Telefonstichprobe zu 67 % (61 % der Befragten). Zur Begründung dieses Verhältnisses *„flossen zum einen aktuelle wissenschaftliche Erkenntnisse (Sturgis, 2020) wie auch die Erfahrungen aus anderen Befragungen mit ähnlichem Erhebungsdesign ein. Des Weiteren wurden für unterschiedliche Verhältnisse von Telefon- und Onlinedaten die daraus jeweils erzielten Ergebnisse hinsichtlich ihrer Plausibilität – insbesondere mit Bezug auf die Prävalenz der Glücksspielteilnahme und -probleme sowie des aufsummierten Geldeinsatzes für bestimmte Spielformen bewertet.“* (Buth, Meyer und Kalke 2022, S. 17).

Diese Argumentation ist aus mehreren Gründen nicht plausibel: Erstens sind die vermeintlichen *„aktuelle wissenschaftliche Erkenntnisse“* (Buth, Meyer und Kalke 2022, S. 17) von Sturgis (2020) veraltet und Sturgis‘ neuer Artikel (Sturgis und Kuha 2022) – den die Autoren des Surveys an anderer Stelle sogar zitieren (S. 13) – gibt keine derartigen Empfehlungen<sup>3</sup>. Zweitens nennen die Autoren bei *„Erfahrungen aus anderen Befragungen mit ähnlichem Erhebungsdesign“* nur die Wählerbefragung von infratest-dimap. Diese Befragung ist allerdings nur bedingt vergleichbar, da die infratest-dimap die 33 %/67 %-Aufteilung direkt für die durchgeführten Interviews und nicht die Gewichtung nutzt – insofern stellt sich auch die Frage, wieso der Glücksspiel-Survey 2021 diese Aufteilung nicht direkt bei der Befragung berücksichtigt hat. Zudem sind – wie ausführlich in Abschnitt 4.2.4 diskutiert – Wählerprognosen die einzigen Befragungen, bei denen *Nonprobability Samples* bessere Ergebnisse erzielen

<sup>3</sup>Dies wurde ausführlich in Abschnitt 4.2.4 diskutiert.

können als *Probability Samples*, daher lassen sich Erfahrungen von Wahlbefragungen nicht auf den Glücksspiel-Survey 2021 übertragen.

Drittens haben die Autoren die Spieleinsätze der Befragten für Produkte des Lotto- und Toto-Blocks (DLTB) hochgerechnet und mit dem Bruttoumsatz des DLTB verglichen. Nähere Informationen finden sich dazu nicht im Glücksspiel-Survey selbst, sondern nur in einem auf dem Survey basierenden Forschungspapier: „*For 2021, the German Lotto and Toto Block reports a turnover of 7.9 billion Euro. This amounts to an average of 658.3 million Euro per month. The stakes for products of the German Lotto and Toto Block which were extrapolated according to the responses given by the participants of the present study add up to 631.3 million Euro per month [95 %-CI: 567.1 million Euro – 695.6 million Euro]. It should be noted in this context that the stakes of persons over 70 years of age who were not included in the study at hand would need to be added for full comparability. However, compared with younger age cohorts, the stakes of persons above the age of 70 usually range far below average.*“ (Buth, Meyer und Kalke 2022, S. 6). Diese Vorgehensweise beruht allerdings auf zwei methodischen Fehlschlüssen: Einerseits summieren die Autoren alle Produkte des DLTB, um sie mit dem Gesamtumsatz des Blocks zu vergleichen. Dabei ist unklar, inwieweit der Umsatz bezüglich der einzelnen Spielformen übereinstimmt – denkbar wäre auch, dass Keno 60 % des Geldeinsatzes der Befragten ausmacht, während Keno nur 10 % des Bruttoumsatzes ausmacht, dies aber aufgrund der Summierung nicht auffällt. Eine methodisch korrekte Vorgehensweise würde also den Geldeinsatz mit dem Umsatz pro Spielform vergleichen, ein aggregierter Vergleich hat geringe Aussagekraft. Zudem ist (siehe die obigen Ausführungen) zu bezweifeln, dass das von den Studienautoren ermittelte Konfidenzintervall der Komplexität des Surveydesigns angemessen ist. Es bleibt offen, wieso die Autoren des Glücksspiel-Surveys die eingesetzten Geldeinsätze der Befragten nur selektiv mit dem Bruttoumsatz des DLTB vergleichen, nicht aber mit dem Umsatz anderer Spielbranchen, beispielsweise mit Einsätzen bei Sportwetten. So kann nicht ausgeschlossen werden, dass mit den anderen Spielformen keine Übereinstimmung vorliegt.

Im Ergebnis haben die Autoren des Glücksspiel-Surveys 2021 mit der *Poststratification* das einfachste Gewichtungsverfahren gewählt, welches bestehende Verzerrungen in den Daten, wenn überhaupt, im telefonisch durchgeführten (*Probability-Sampling*) Teil der Befragung (und dort auch nur in begrenztem Ausmaß) korrigieren kann. Bereits bei der Erstellung des Fragebogens hätten die Autoren die geplante Gewichtung berücksichtigen müssen, um geeignete Hilfsvariablen für eine effektive Gewichtung nutzen zu können (Lohr 2021). Dies ist entweder nicht geschehen oder wurde bei der Gewichtung nicht umgesetzt. Aufgrund der nicht zur Verfügung gestellten Materialien kann dies im vorliegenden Gutachten nicht beurteilt werden. Die *Mode*-Gewichtung der Autoren ist ebenfalls nicht überzeugend

begründet, da veraltete oder konzeptionell nicht übertragbare Studien zur Begründung herangezogen werden. Der (selektive) Vergleich mit dem Bruttoumsatz des DLTB überzeugt zudem aufgrund der genutzten Aggregatdaten nicht. Auch wurden keine weiteren Vergleiche – trotz vorhandener Daten – mit anderen Spielformen durchgeführt, wodurch die Plausibilität der Gewichtung diesbezüglich nicht überprüfbar ist. Insgesamt sollte der Glücksspiel-Survey aufgrund seiner Ziele transparenter mit der Gewichtung umgehen, und bei Befragungen mit solch gewichtigen Zielen sollte insbesondere eine Sensitivitätsanalyse durch einen Vergleich verschiedener Gewichtungsvariablen und -verfahren durchgeführt werden, die – insbesondere bei großen Veränderungen – auch transparent dokumentiert werden sollten.

## 6 | Weitere methodische Kriterien

### 6.1 Querschnittsbefragung

#### Executive Summary

*Obwohl für die Erreichung eines Teils der Ziele des Glücksspielsurveys 2021 eine Längsschnittbefragung alternativlos gewesen wäre, haben die Autoren des Surveys Querschnittsbefragungen durchgeführt. Dadurch können auf Grundlage des Surveys und der gewählten Methodik keine kausalen Aussagen in Bezug auf die erhobenen Daten getroffen werden, was die Autoren des Surveys illegitimerweise dennoch machen, beispielsweise, „dass es auch unter den Teilnehmer\*innen von Glücksspielen mit geringem Gefährdungspotential einen nennenswerten Anteil von Problemspielenden gibt – auch wenn überwiegend andere Glücksspielformen diese glücksspielbezogenen Probleme verursacht haben dürften.“ Für diese und ähnliche Kausalaussagen kann eine solche Querschnittsbefragung jedoch keine Evidenz liefern, für die Ziele des Surveys ist die gewählte Methode damit ungeeignet. Ein Hinweis auf diese Limitation von Querschnittsbefragungen, die keine über Korrelationen hinausgehenden Schlussfolgerungen zulassen, fand sich noch in den vergangenen BZgA-Surveys, allerdings nicht mehr im Glücksspielsurvey 2021.*

Je nach Ziel und Ressourcen einer Befragung sollte entweder eine Querschnittsbefragung, also eine Befragung zu einem bestimmten Zeitpunkt, oder eine Längsschnitt- oder Panelbefragung<sup>1</sup>, bei der Individuen mehrmals an verschiedenen Zeitpunkten befragt werden, durchgeführt werden. Nach einer kurzen Darstellung beider Befragungsformen wird im Folgenden erläutert, welche der beiden Methoden besser für die Ziele des Surveys geeignet wäre, und die Entscheidung der Autoren für eine

<sup>1</sup>Im engeren Sinne ist eine Längsschnittbefragung die Befragung eines Individuums zu mehreren Zeitpunkten; oft wird der Begriff jedoch synonym zu einer Panelbefragung verwendet.

Querschnittsbefragung wird in diesem Lichte kritisch gewürdigt.

Querschnittsbefragungen ermöglichen eine punktuelle Datenerhebung (Glücksspiel-Survey 2021: August bis Oktober 2021) durch die Befragung einer Stichprobe zu einem bestimmten Zeitpunkt. Die Methode ist kostengünstig und zeiteffizient, wodurch sie sich gut zur punktuellen Erfassung der aktuellen Verteilung von Merkmalen und Verhaltensweisen in der untersuchten Bevölkerung eignet. Während sie jedoch nützliche Momentaufnahmen und die Untersuchung der Korrelation verschiedener Merkmale ermöglicht, ist sie ungeeignet, Kausalbeziehungen oder zeitliche Veränderungen zu untersuchen (Pforr und Schröder 2015). In einer Längsschnittbefragung hingegen werden dieselben Personen mehrmals im zeitlichen Abstand befragt. Dies ermöglicht die Untersuchung von Veränderungen, Kausalbeziehungen und Trends im Zeitverlauf. Dafür geht diese Methode allerdings mit einem erheblichen Zeit- und Ressourcenaufwand einher und ist entsprechend teurer. Zudem kommen bei der Längsschnittbefragung eines Panels neue potenzielle Verzerrungen durch Panel Attrition (Befragte nehmen nicht mehr teil) oder Konditionierung (Befragte lernen aus vorherigen Befragungen) hinzu.

Betrachtet man die Zielsetzung des Glücksspiel-Surveys 2021, so muss man zwischen den einzelnen Zielen differenzieren. Eine Querschnittsbefragung wäre geeignet,

- die Prävalenz verschiedener Merkmale bezüglich Glücksspiels zu erfassen,
- die Bekanntheit und Akzeptanz von Jugend- und Spielerschutzmaßnahmen zu erheben sowie
- korrelative Zusammenhänge zwischen verschiedenen Faktoren zu untersuchen.

Allerdings ist eine Querschnittsbefragung nicht geeignet,

- Maßnahmen des Spieler- und Jugendschutzes,
- die Auswirkungen des Glücksspielstaatsvertrags oder
- den Einfluss von Werbung zu evaluieren.

Denn bei diesen Zielen sollen kausale, also Ursache-Wirkungs-Zusammenhänge untersucht werden (Pforr und Schröder 2015). Folglich wäre eine Längsschnittbefragung die bessere Methode gewesen, da sie zur Erreichung aller Ziele zugleich beitragen kann. Generell geben die Nationale Akademie der Wissenschaften Leopoldina, acatech – Deutsche Akademie der Technikwissenschaften und Union der deutschen Akademien der Wissenschaften (2015) an: *„Bevölkerungsweite Längsschnittstudien bilden das Rückgrat der empirischen Forschung in den Sozial-, Wirtschafts- und Verhaltenswissenschaften*

sowie der *Epidemiologie und der Gesundheitsforschung*“. Diese Einschätzung ist Konsens in der einschlägigen Fachliteratur, so stellte die BZgA bereits im Survey von 2019 klar: „*Da bislang ausreichend gesicherte Längsschnittdaten fehlen, ermöglichen solche Modelle allerdings keine kausalen Aussagen*“ (Banz 2019, S. 26). Ebenso subsumierten Hayer u. a. (2022) in einer Bewertung vergangener Glücksspielregulierungsevaluationen, dass „*in der Vergangenheit wie Gegenwart mehrere Chancen verpasst [wurden], eine wissenschaftlich fundierte Evaluation der gesetzlichen Rahmenbedingungen unter Berücksichtigung forschungstechnischer Mindeststandards (u.a. mit Longitudinaldaten) einzuleiten.*“ (S. 240). Weiterhin schreiben die Autoren bezüglich optimaler Forschungsdesigns, dass „*turnusmäßige Erhebungen zum Glücksspielverhalten in der Bevölkerung einschließlich der Bestimmung des Ausmaßes glücksspielbezogener Probleme sowie riskanter Spielmuster zu priorisieren [sind]. Dieses Forschungsdesign eröffnet die Möglichkeit der Abbildung individueller Entwicklungsverläufe unter veränderten regulatorischen Rahmenbedingungen. Wünschenswert wäre eine Kombination aus wiederholten repräsentativen Erhebungen sowie echten Längsschnittstudien (also eine Wiederholungsmessung bei den gleichen Personen) mit Messzeitpunkten alle zwei oder drei Jahre.*“ (S. 242 f.). In anderen Ländern werden diesen Qualitätsmaßstäben entsprechend auch Längsschnittbefragungen durchgeführt, beispielsweise *Swelogs* (The Swedish Longitudinal Gambling Study) (Abbott, Romild und Volberg 2018) in Schweden.

Der Glücksspiel-Survey 2021 ist entgegen den Anforderungen seiner Zielsetzung eine Querschnittsbefragung. Dennoch werden im Survey Zusammenhänge untersucht und Aussagen getroffen, die sich aus den mit einem solchen Design generierten Daten nicht ableiten lassen<sup>2</sup>. Zum Beispiel schreiben die Autoren, „*dass es auch unter den Teilnehmer\*innen von Glücksspielen mit geringem Gefährdungspotential einen nennenswerten Anteil von Problemspielenden gibt – auch wenn überwiegend andere Glücksspielformen diese glücksspielbezogenen Probleme verursacht haben dürften.*“ (Buth, Meyer und Kalke 2022, S. 7). Diese im Fazit befindliche Aussage impliziert, dass glücksspielbezogene Probleme bei Nutzern von Glücksspiel mit geringerem Gefährdungspotential (bspw. Lotto) überwiegend von der Teilnahme an Glücksspiel mit höherem Gefährdungspotential (bspw. Poker) stammen. Allerdings liefert der Survey keinerlei Daten, die diese Aussagen stützen – wie oben erläutert, ist er aufgrund des Querschnittsdesigns auch nicht dazu geeignet. Diese Limitation ist bedingt durch die einmalige Befragung, wodurch keine zeitliche Abfolge zwischen Faktoren – eine Grundbedingung für Kausalität (Opp 2010) – festgestellt werden kann. Aus diesem Grund handelt es sich um eine spekulative Behauptung, die nicht ausreichend als solche kenntlich gemacht wird. Vielmehr ist es gute wissenschaftliche Praxis, solche Vermutungen als Hypothesen mit weiterem Forschungsbedarf zu formulieren.

---

<sup>2</sup>Sollten zukünftig Folgebefragungen an denselben Individuen erfolgen, so ist die Situation neu zu bewerten, wobei die oben getätigten Kritikpunkte zur Repräsentativität bestehen bleiben.

Zur Illustration soll Person A betrachtet werden, die sowohl Glücksspielform 1 und 2 spielt und eine leichte Glücksspielstörung hat. Person A wird an Tag  $t$  in einer Querschnittsbefragung befragt. Aufgrund dieses Studiendesigns kann die Herkunft der Glücksspielprobleme nicht erklärt werden, da keine zeitliche Reihenfolge etabliert werden kann. Sowohl könnte Glücksspielform 2 die Ursache der Spielstörung sein als auch eine vorhandene Spielstörung erst zu Glücksspielform 2 geführt haben (oder eine dritte, unbeobachtete Variable könnte die Glücksspielstörung verursacht haben). Tatsächlich kann nur festgestellt werden, dass A eine leichte Glücksspielstörung aufweist. Eine Längsschnittbefragung wäre geeigneter gewesen, solche Aussagen zu treffen, denn mit diesem Design hätte man (im Idealfall) beobachten können, wie Person A zum Zeitpunkt  $t$  nur Glücksspielform 1 spielte und keine Probleme hatte, zum Zeitpunkt  $t + 1$  allerdings nun auch Glücksspielform 2 spielt und zudem eine leichte Glücksspielstörung entwickelt hat. Dieses Szenario ist in Abbildung 6.1 dargestellt.

## Längsschnittbefragung an Tag $t$ & $t + 1$

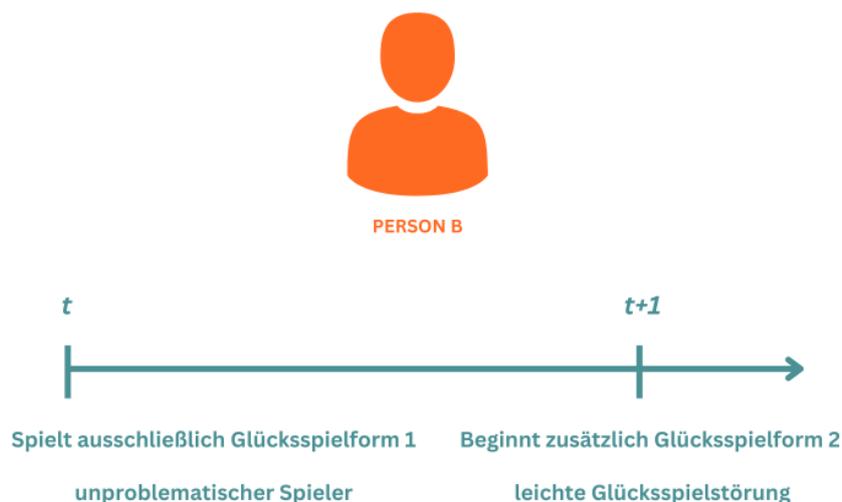


Abbildung 6.1: Vorteile einer Längsschnittbefragung

Selbiges gilt natürlich auch für den Zusammenhang von pathologischem Spielverhalten mit psychischen Problemen, Alkoholkonsum, Maßnahmen des Spieler- und Jugendschutzes und Werbung. Zwischen allen Faktoren können lediglich korrelative und keine kausalen Zusammenhänge untersucht werden: Ob Alkohol Spielstörungen fördert oder vice versa, kann mit den Daten des Surveys nicht beantwortet werden. Ebenso ist das Querschnittsdesign nicht geeignet, den Glücksspielstaatsvertrag zu evaluieren (ein explizites Ziel des Surveys), da keine individuelle Veränderung im Zeitverlauf – wie es bei

einer Längsschnittstudie geschieht – betrachtet werden können. Es sei nochmals deutlich darauf verwiesen, dass hierfür die fortlaufende Befragung derselben Individuen notwendig ist, d.h. eine echte Panelbefragung und keine Folge von Querschnittsbefragungen. Andernfalls besteht das Risiko eines ökologischen Fehlschlusses. Mit den geplanten weiteren Befragungen in 2023 und 2025 können deshalb nur grundsätzliche Veränderungen in der Bevölkerung gemessen werden, deren Ursprung aber mit Querschnittsbefragungen nicht erklärt werden kann.

Der Vollständigkeit halber sei angemerkt, dass gewisse methodische Ansätze (*Rubin Causal Model, RCM*) existieren, die unter bestimmten Annahmen auch in Querschnittsuntersuchungen Kausalität durch kontrafaktische Analysen nachzuweisen versuchen. Allerdings erfordert dieser Ansatz die konditionale Unabhängigkeit der Zuweisung zum Experimentalstatus von der interessierenden Zielgröße, vgl. (Gangl und DiPrete 2004, S. 7). Im vorliegenden Fall würde das bedeuten, dass die Entscheidung für eine bestimmte Glücksspielform (= hypothetischer Einflussfaktor auf das Risiko einer Glücksspielstörung), kontrolliert auf die Kovariaten wie z.B. Alter und Geschlecht, unabhängig ist vom Risiko eines Spielers für die Entwicklung glücksspielbezogener Probleme. Anders gesagt, müsste man die Daten des Glücksspiel-Surveys als „Quasi-Experiment“ auffassen, in dem jeder Spieler, nach Berücksichtigung relevanter soziodemografischer Merkmale, „frei“ einer Glücksspielform „zugewiesen“ wurde und aufgrund dieser Zuweisung ein geringeres oder höheres Risiko für die Entwicklung einer Glücksspielstörung aufweist. Diese Annahme ist nach aktuellem Forschungsstand kaum plausibel zu begründen.

Zusammenfassend lässt sich feststellen, dass die Ziele des Glücksspiel-Surveys 2021 eine Längsschnittbefragung erfordern, die Autoren des Surveys sich allerdings für eine (Serie von) Querschnittsbefragung(en) entschieden haben. Damit lassen sich einige Ziele des Surveys nicht erfüllen, da die gewählte Befragungsform zur Beantwortung der gestellten Fragen nicht geeignet ist. Ein Hinweis auf diese Tatsache ist allerdings im Glücksspiel-Survey nicht zu finden – anders als noch im BZgA-Survey von 2019. Dieser fehlende Hinweis ist keine Trivialität, denn bei einer Querschnittsbefragung können keine über Korrelationsaussagen hinausgehenden Schlussfolgerungen gezogen werden.

## 6.2 Kausalität

### Executive Summary

*Kausalität beschreibt die Beziehung von Ursache und Wirkung. Um kausale Schlüsse ziehen zu können, ist neben der Kontrolle für konfundierende Faktoren in der Beziehung zwischen zwei Variablen vor allem die zeitliche Abfolge relevant – dafür sind zeitlich versetzte Messpunkte notwendig. Durch das Studiendesign der Querschnittsbefragung bedingt, ist der Glücksspielsurvey nicht dafür geeignet, Daten zu generieren, auf deren Basis sich kausale Aussagen treffen lassen. Dennoch geschieht dies für den statistischen Zusammenhang zwischen Glücksspielsucht und möglichen Risikofaktoren wie Alkohol und psychischen Problemen. Dort wird ein kausaler Zusammenhang suggeriert, der grundsätzlich aus den Daten und durchgeführten Analysen wissenschaftlich nicht abgeleitet werden kann. Neben Zusammenhängen, die durch gemeinsame, unbeobachtete Faktoren, etwa eine erhöhte Risikoneigung oder andere Persönlichkeitsdimensionen, verursacht sein können, kann auch eine umgekehrte Kausalität im Survey nicht ausgeschlossen werden, was eine Identifikation von „riskanten Glücksspielformen“ und ihrem Gefährdungspotenzial unmöglich macht. Zudem beschränkt sich der Glücksspielsurvey auf ein einziges Modell zur Entstehung von Glücksspielproblemen und beleuchtet nicht andere in der Literatur verbreitete Erklärungsansätze.*

Kausalität beschreibt das grundlegende Konzept, dass eine Ursache einen bestimmten Effekt oder eine bestimmte Wirkung hervorruft. Es geht also darum, die Beziehung zwischen Ursache und Wirkung zu verstehen und zu erklären (Opp 2010). Um kausale Schlüsse ziehen zu können, müssen bestimmte Voraussetzungen erfüllt sein.

Nachfolgend finden sich einige wichtige Konzepte, die bei der Analyse kausaler Zusammenhänge zu beachten sind. Dazu gehört das Konzept der *Korrelation*: Eine Korrelation zwischen zwei Variablen bedeutet, dass sie miteinander in Beziehung stehen<sup>3</sup>. Eine positive Korrelation beinhaltet, dass eine Zunahme einer Variable mit einer Zunahme der anderen Variable einhergeht, während eine negative Korrelation eine Abnahme der einen Variable bei Zunahme der anderen Variable bedeutet. Eine Korrelation allein lässt jedoch noch keine Rückschlüsse auf kausale Zusammenhänge zu. So sind beispielsweise die Variablen Absatz von Speiseeis und Sonnenbrand stark korreliert, aber nicht kausal

<sup>3</sup>Häufig werden implizit zusätzliche Annahmen über die Form des Zusammenhangs getroffen, beispielsweise Linearität.

verbunden. Stattdessen werden beide Variablen von einer dritten Variable — starke Sonneneinstrahlung — beeinflusst. Sonnenbrand ist in diesem Beispiel beliebig gegen andere durch Sonne bedingte Faktoren (Absatz von Sonnencreme, Besucher in Freibädern, etc.) austauschbar. *Konfundierende Faktoren* (confounding factors) sind andere Variablen, die sowohl mit der unabhängigen als auch mit der abhängigen Variable in Verbindung stehen können und den vermeintlichen kausalen Zusammenhang erklären könnten. Es ist wichtig, potenzielle konfundierende Variablen zu berücksichtigen und zu kontrollieren, um sicherzustellen, dass der beobachtete Effekt tatsächlich auf die unabhängige Variable zurückzuführen ist (Kühnel und Dingelstedt 2022). Dies geschieht beispielsweise mithilfe einer Regressionsanalyse.

Um eine kausale Beziehung zu postulieren, muss die Ursache der Wirkung zudem zeitlich vorausgehen. Das bedeutet, dass die unabhängige Variable vor der abhängigen Variable stattfinden oder auftreten muss. So kann eine Korrelation und sogar eine kausale Beziehung zwischen zwei Faktoren existieren, aber es ist nicht möglich, zu identifizieren, welche Variable die Änderungen in der anderen bedingt. Dies lässt sich exemplarisch anhand der Beziehung von pathologischem Glücksspiel und der Wahrnehmung von Werbung für Glücksspiel verdeutlichen. Je mehr Werbung für Glücksspiel eine Person wahrnimmt, desto stärker ist ihre Glücksspielstörung laut Glücksspielsurvey. Plausibel wäre aber auch eine umgekehrte Beziehung: Je stärker die Glücksspielstörung einer Person, desto mehr Werbung für Glücksspiel nimmt sie wahr, da sie eine erhöhte Sensibilität für das Thema besitzt. Ohne weitere Untersuchung — insbesondere ohne mehrere, zeitlich versetzte Messpunkte (Opp 2010) — ist somit unklar, welche Variable einen kausalen Effekt auf die andere hat. Die zeitliche Abfolge ist daher ein entscheidendes Kriterium, um eine kausale Beziehung zu begründen.

### Kausalität im Glücksspiel-Survey 2021

Wie in Abschnitt 6.1 ausführlich dargelegt, ist der Glücksspiel-Survey qua Studiendesign nicht dafür geeignet, kausale Aussagen zu treffen. Bedingt durch die Ziele des Surveys (siehe Abschnitt 2), insbesondere

1. das Untersuchen der Zusammenhänge zwischen Glücksspielproblemen, psychischer Belastung und Alkoholkonsum (Buth, Meyer und Kalke 2022, S. 8) sowie
2. das Untersuchen der Beziehung von glücksspielbezogener Werbung (bzw. deren Wahrnehmung) und Glücksspielverhalten (Buth, Meyer und Kalke 2022, S. 8),

werden allerdings solche Kausalaussagen getroffen oder es wird lediglich ungenügend über die korrelative Natur der Daten aufgeklärt. Die getroffenen Aussagen genügen gemeinhin nicht den Anforderungen, die in der Wissenschaft an solche gestellt werden (siehe Abschnitt 6.2). Dies werden im Folgenden für jede Kausalaussage ausgeführt und es wird dargelegt, warum das gewählte Studiendesign ungeeignet ist, die jeweilige Aussage zu treffen.

Bei der folgenden Ursache-Wirkungs-Aussage handelt es sich beispielsweise um eine unbewiesene These: *„Die Ergebnisse zur multiplen Spielteilnahme haben deutlich gemacht, dass die Mehrheit der Spieler\*innen riskanter Glücksspiele auch an weiteren Spielformen — insbesondere den Lotterien — teilnimmt. Dies wiederum bedeutet, dass es auch unter den Teilnehmer\*innen von Glücksspielen mit geringem Gefährdungspotential [sic!] einen nennenswerten Anteil von Problemspieler\*innen gibt — auch wenn meistens andere Glücksspielformen die Probleme verursacht haben dürften —, die in den zugehörigen terrestrischen und onlinegestützten Spielstätten mit präventiven Maßnahmen erreicht werden könnten“* (Buth, Meyer und Kalke 2022, S. 49). Es geht aus den Daten allerdings nicht hervor, aus welchen Gründen Lotto-Spieler ein problematisches Spielverhalten entwickelt haben. Der Survey gibt an, diese Hypothese der riskanten Glücksspiele (zu denen Lotto nicht gehört) zu untersuchen, tut dies aber tatsächlich aufgrund der bereits diskutierten Natur der erhobenen Daten gar nicht.

Die Bezeichnung „riskante Glücksspiele“ oder „Glücksspiele mit geringem Gefährdungspotenzial“ geht auf eine Theorie zurück, die auf der Annahme einer Monokausalität zwischen einzelnen Glücksspielarten und Glücksspielstörungen ausgeht. Es existieren in der Wissenschaft allerdings ebenso andere Theorien zu den Einflussfaktoren auf die Entstehung von Glücksspielstörungen. So schlussfolgern Bühringer u. a. (2023, S. 9) – nicht zuletzt aufgrund der spärlichen wissenschaftlichen Datenlage – in einer vom BMWK in Auftrag gegebenen Studie, dass es *„keine empirische Evidenz für ein unterschiedlich kausales Risikopotential verschiedener Glücksspiele“* gibt. Sie behandeln das sog. *Vulnerabilitäts-Risiko-Modell*, nach dem Personen eine individuelle Vulnerabilität aufweisen, die sie im Zusammenspiel mit äußeren Risikofaktoren (Umwelt- und/oder Glücksspielfaktoren) anfälliger für eine Glücksspielstörung machen kann. Es existieren weitere Modelle, die beispielsweise psychosoziale Faktoren oder Umweltfaktoren beinhalten.

An dieser Stelle erfolgt lediglich der Verweis auf die Existenz weiterer Theorien (Manrique u. a. 2023) – das Gutachten erhebt keinen Anspruch auf Vollständigkeit in dieser Hinsicht. In einer umfassenden Untersuchung sollten jedoch verschiedenste Erklärungsansätze in Betracht gezogen und anhand geeigneter Daten untersucht werden. Diesem Anspruch genügt der Glücksspiel-Survey 2021 nicht. So schreibt auch der Autor des Glücksspiel-Surveys der BZgA von 2019 in einer weiterführenden Pu-

blikation: „Bei der mechanischen Zuordnung von zumindest problematisch eingestuften Spielern zu den Spielformen, die dieser Spieler spielt, ist kein direkter Kausalzusammenhang gegeben“ (Banz und Becker 2019, S. 221). Vielmehr müssten „aus wissenschaftlicher Sicht weitere Schlussfolgerungen mit Vorsicht und den dargelegten Limitationen vorgenommen werden.“ (Banz und Becker 2019, S. 222). Diese gebotene Vorsicht kann dem Glücksspiel-Survey 2021 nicht uneingeschränkt attestiert werden. Grundsätzlich ist der Glücksspiel-Survey qua Design ungeeignet, Vulnerabilität, Angebot oder eine Kombination von beidem als Ursache eines problematischen Verhaltens zu identifizieren.

Welche Art von Glücksspiel in den Daten den größten Einfluss auf Glücksspielstörungen hat – also am stärksten mit Glücksspielstörungen korreliert<sup>4</sup> – ließe sich mithilfe einer multiplen Regressionsanalyse überprüfen. Doch im Gutachten unterbleibt grundsätzlich die Nutzung statistischer Modelle zur Untersuchung von Korrelation, sodass diese an keiner Stelle abgeleitet werden kann. Aus diesem Grund können Unterschiede zwischen Gruppen im Survey (z.B. Frauen und Männer) – sofern sie überhaupt statistisch signifikant sind (siehe Abschnitt 4.1) – ebenfalls nicht genauer untersucht werden. Es besteht die Möglichkeit, dass diese durch dritte Variablen (z.B. Arbeitslosigkeit, die unter Männern größer als unter Frauen ist) verursacht werden.

Fehlende Regressions- oder Korrelationsanalysen hindern die Identifikation von korrelativen und kausalen Beziehungen in den Daten ebenso wie das Studiendesign und die Nutzung von Querschnittsdaten 6.1. Obwohl die Anwendung solcher gängigen statistischen Methoden, gemessen am Aufwand, einen hohen Erkenntnisgewinn verspricht, unterbleibt die Anwendung dieser gänzlich. Bemerkenswert ist allerdings, dass die Autoren im bereits angesprochenen Forschungspapier (Meyer, Kalke und Buth 2023) mit den Surveydaten eine multinomiale logistische Regression rechnen und sich folglich des Nutzens komplexerer Modelle bewusst zu sein scheinen. Die vorgenommene Regressionsanalyse untersucht allerdings lediglich die Beziehung von Geschlecht und Glücksspielstörung und beinhaltet ausschließlich die Kontrollvariable Alter. Ein solches Modell ermöglicht keinen Erkenntnisgewinn hinsichtlich des Risikos einzelner Glücksspielarten. Erforderlich wäre u.a. eine Untersuchung, welche Faktoren und Arten von Glücksspiel den stärksten (statistischen) Einfluss auf das Vorhandensein bzw. die Schwere einer Glücksspielstörung haben. Einige dieser Modelle wurden beispielsweise in den früheren BZgA-Surveys gerechnet.

Ferner ist das Problem der fehlenden Längsschnittdaten bereits seit längerem bekannt und wurde im letzten Glücksspiel-Survey der BZgA thematisiert: „Aufgrund der uneinheitlichen Ergebnisse sowie durch das der Befragung zugrunde liegende Querschnittsdesign lassen sich auch im Jahr 2019 durch

---

<sup>4</sup>Anmerkung: Es kann nur um Korrelation gehen, nicht um Kausalität, dafür herrschen wie erläutert höhere Ansprüche an die Daten.

*Eurojackpot bedingte Kanalisierungseffekte abschließend weder nachweisen noch widerlegen [sic]. Um hierzu weiteren Aufschluss zu erlangen, wäre zu überlegen, inwieweit ein longitudinal angelegtes Studiendesign mit einer ausreichenden Anzahl und zu mindestens zwei Zeitpunkten identischen Befragten realisiert werden kann. Diese wären dann detailliert zu ihrem Spielverhalten zu befragen“* (Banz 2019, S. 127). Dieser sachlich richtigen Argumentation sind die Autoren des Glücksspiel-Surveys aus unbekanntem Gründen nicht gefolgt.

Wegen dieser Gegebenheiten kann nicht verlässlich ausgeschlossen werden, dass im Survey untersuchte Kategorien, wie Geschlecht, Alkoholkonsum und psychische Erkrankungen, überhaupt ursächlich für die von den Autoren gezeigten und nach Grad der Glücksspielstörung differenzierten Unterschiede sind. Für den statistisch nicht versierten Leser erwecken einige Grafiken im Survey dennoch diesen Eindruck (siehe Abschnitt 4.1).

Der Survey hat nach Aussage der Studienautoren *„einzelne Risikofaktoren wie beispielsweise psychische Belastung und Alkoholkonsum erfasst, um mögliche Zusammenhänge mit bestehenden Glücksspielproblemen zu untersuchen“* (Buth, Meyer und Kalke 2022, S. 8). Bezüglich des Zusammenhangs zwischen Glücksspielstörungen und psychischen Erkrankungen stellen die Autoren fest, dass *„mit zunehmendem Schweregrad der glücksspielbezogenen Störung ein kontinuierlich steigender Anteil mit psychischen Beeinträchtigungen (von 36,3 % bis 68,3 %) verbunden ist.“* (Buth, Meyer und Kalke 2022, S. 39). Als Beleg für diese Aussage dient Abbildung 14 auf Seite 40. Dabei wird dem statistisch nicht versierten Leser suggeriert, es sei eine Korrelation zwischen dem Schweregrad der Glücksspielprobleme und der Beeinträchtigung der psychischen Gesundheit vorhanden – ganz im Sinne von *„je stärker die Glücksspielstörung, desto höher der Anteil an psychischen Beeinträchtigungen“*.

Allerdings führen die Autoren lediglich eine deskriptive Analyse durch. Einen Schluss auf Korrelation oder Kausalität lässt diese aus statistischer Sicht nicht zu. Zum besseren Verständnis und zur Vorbeugung von Fehlinterpretationen beispielsweise durch Medien und (politische) Entscheider (siehe dazu Kapitel 10) sollte daher ein Kommentar der Autoren ähnlich zum BZgA Survey von 2019 erfolgen: *„Bei der Analyse der einzelnen Glücksspiele ist zu beachten, dass es sich um eine rein deskriptive Beschreibung handelt und die Fallzahlen durch die Differenzierung z. T. relativ klein sind.“* (Banz 2019, S. 13) Sehr ähnliche Formulierungen finden sich für die Beziehung von riskantem Alkoholkonsum und dem Schweregrad der Glücksspielstörung (S. 40). Auch hinsichtlich der Wahrnehmung von Glücksspielstörungen im sozialen Umfeld der Befragten und deren Schweregrad wird im Glücksspiel-Survey eine Korrelation suggeriert. Dies erfolgt durch die Aussage, ein Anteilswert von Befragten, die wegen Personen mit Glücksspielstörung in ihrem Umfeld schon einmal einen Beratungstermin wahr-

genommen hätten, würde mit „*zunehmender Glücksspielproblematik steigen*“ (Buth, Meyer und Kalke 2022, S. 42). Es unterbleibt auch an dieser Stelle eine über die deskriptive Darstellung hinausgehende Analyse.

Die Prüfung auf fälschlicherweise getätigte kausale Aussagen hat nennenswerte Defizite bei einzelnen Formulierungen hervorgebracht und aufgezeigt, dass die notwendigen Grundlagen für das Ableiten von Kausalaussagen im Glücksspiel-Survey 2021 nicht gegeben sind. Demnach sind keine gesicherten Rückschlüsse auf Risikofaktoren von Glücksspielstörungen sowie auf das Gefährdungspotenzial einzelner Glücksspielarten möglich. Ein Hinweis wie der folgende, den zwei der Autoren des Surveys in einer anderen Publikation zu Glücksspielverhalten anbringen, wäre auch im Glücksspiel-Survey 2021 notwendig gewesen: „*Die Ergebnisdarstellung erfolgt ausschließlich deskriptiv. Das heißt, dass auf die Prüfung von Unterschieden zwischen den Verteilungen [...] wie auch der Veränderungen im Zeitverlauf [...] im Rahmen dieses vornehmlich explorativ angelegten Berichtes – zugunsten einer besseren Lesbarkeit – verzichtet wurde*“ (Buth, Schütze und Kalke 2021, S. 22).

## 7 | Inhaltliche Evaluation

### 7.1 DSM-5 Kriterien

#### Executive Summary

*Im Survey wird das Instrument DSM-5 zur Evaluation von glücksspielbezogenen Problemen der Befragten verwendet. Die bisher im deutschen Sprachraum unter Verwendung der DSM-5-Kriterien durchgeführten Untersuchungen gingen von problematischem Glücksspielverhalten (präklinische und klinische Störung) aus, wenn mindestens vier der neun Kriterien erfüllt waren. Der Glücksspiel-Survey reduziert diesen Schwellenwert auf ein zu erfüllendes Kriterium. Die Gruppe der neu als „problematisch“ eingestuften Spieler, die ein bis drei Kriterien erfüllen, bezeichnen die Autoren als „riskante Spieler“. Diese Kategorie ist derzeit Gegenstand wissenschaftlicher Diskussion, wird aber ohne hinreichende Rechtfertigung der Gruppe der „problematischen Spieler“ zugeordnet. Die primäre Begründung, dass die geringen individuellen glücksspielinduzierten Kosten von „riskanten Spielern“ zusammengenommen hohe gesellschaftliche Kosten verursachen würden, ist gemäß Literatur mit erheblichen methodischen Problemen verbunden. Für eine Sensitivitätsanalyse bräuchte es mindestens eine Darstellung der genauen Verteilung der Befragten innerhalb der Kategorien, die im Gegensatz zu vergleichbaren Studien hier nicht vorgenommen wird. Vor diesem Hintergrund muss die Absenkung des Schwellenwertes, ab dem einer befragten Person „problematisches Spielverhalten“ attestiert wird, zwar nicht abgelehnt werden. Sie ist aber zumindest diskussionswürdig und bedarf klarer Validierung, ist es doch erst die Ausdehnung des Begriffs „problematisches Spielverhalten“, welche die nach der Survey-Veröffentlichung vielfach öffentlich rezipierte Zahl von über fünf Millionen Personen in Deutschland mit einer Glücksspielproblematik erzeugt.*

In der wissenschaftlichen Literatur zu Glücksspielverhalten wird eine Vielzahl von Instrumenten zur Evaluation glücksspielbezogener Probleme der Bevölkerung vorgestellt und diskutiert. Der Glücksspiel-Survey der Bundeszentrale für gesundheitliche Aufklärung (BZgA) von 2019 wendet die South Oaks Gambling Screen (SOGS) Kriterien an (Banz 2019), während der vorliegende Survey ein anderes Instrument benutzt – die DSM-5-Kriterien. Der SOGS ist ebenso wie DSM ein international vielfach eingesetztes, standardisiertes Instrument. Ob eine Glücksspielstörung vorliegt, wird beim SOGS aus 20 Fragen ermittelt, die mit jeweils einem Punkt bewertet werden. Erreicht ein Teilnehmer fünf oder mehr Punkte, liegt höchstwahrscheinlich ein pathologisches Glücksspielverhalten vor. Bei drei oder vier Punkten werden Befragte als „problematisch Glücksspielende“ eingestuft und bei ein oder zwei Punkten wird ihnen im BZgA-Glücksspiel-Survey von 2019 ein „auffälliges bzw. riskantes Glücksspielverhalten“ attestiert, das sich jedoch klar vom „problematischen Spielverhalten“ abgrenzt. Diese letzte Kategorie war in der ursprünglichen Version des SOGS nicht vorgesehen (Lesieur und Blume 1987) und wird in der Literatur mitunter kritisch gesehen (Delfabbro, Australia. u. a. 2010).

Die DSM-5-Kriterien sind mit einer zusätzlichen Kategorie etwas differenzierter als der SOGS. Als wissenschaftlicher Konsens ist derzeit anzusehen, dass ab vier von neun erfüllten Kriterien eine leichte Glücksspielstörung vorliegt, ab sechs von neun erfüllten Kriterien eine mittlere Glücksspielstörung und ab acht von neun erfüllten Kriterien eine schwere Glücksspielstörung. Der Schwellenwert für das Vorliegen „problematischen Spielverhaltens“ (präklinische/klinische Störung) liegt nach allgemeiner Auslegung von DSM-5 also bei vier von neun erfüllten DSM-5-Kriterien. Zusätzlich kategorisiert der Glücksspiel-Survey auch jene Personen, die in der Befragung lediglich ein bis drei DSM-5-Kriterien erfüllen: Ihnen attestiert der Survey „*riskantes Glücksspielverhalten*“, ähnlich der niedrigsten Kategorie des BZgA-Surveys von 2019 („auffälliges“ bzw. „riskantes Spielverhalten“) (Banz 2019). Diese Kategorisierung von Fällen, „die unterhalb der Schwelle für eine (Abhängigkeits-)diagnose“<sup>1</sup> (Falk Kiefer 2011, S. 46) liegen, ist wissenschaftlich bislang nicht abschließend diskutiert. Indem die Survey-Autoren diese Kategorie der „riskanten Spieler“ der Gruppe der „problematischen Spieler“ zuordnen (Buth, Meyer und Kalke 2022, S. 6), reduzieren sie den Schwellenwert für die Annahme problematischen Spielverhaltens entgegen allen vergleichbaren vorherigen Untersuchungen von der Erfüllung von mindestens vier auf ein DSM-5-Kriterium. Hierdurch vervielfacht sich der Anteil der Personen in Deutschland zwischen 18 und 70 Jahren, die der Survey als problematische Spieler einstuft, von 2,3 % auf 8,0 %.

Die Kategorie des „riskanten Spielverhaltens“ wird im Glücksspiel-Survey unter anderem damit begründet, dass „auch Risiko- und Nicht-Problem-Spieler\*innen“ von den „*potentiell schädlichen Auswir-*

<sup>1</sup>Im Falle von Substanzgebrauchsstörungen teils als „diagnostische Waisen“ bzw. „diagnostic orphans“ bezeichnet.

*kungen des Glücksspiels*“ (Buth, Meyer und Kalke 2022, S. 19) betroffen seien. Weshalb bzw. inwiefern genau jedoch Spieler, die nach allgemeiner Auslegung von DSM-5 explizit kein problematisches (d.h. unproblematisches!) Spielverhalten aufweisen, von diesen schädlichen Auswirkungen betroffen sind, wird in der Studie unzureichend dargelegt. Auch werden mit der Angabe einer einzigen Quelle keine hinreichenden Belege für diese Aussage angeführt. Hier ist anzumerken, dass durch DSM-5 lediglich bzw. hauptsächlich Symptome für eine Glücksspielstörung bzw. ein Risiko für Glücksspielstörung erfasst werden können; diese sind nicht gleichzusetzen mit glücksspielbezogenen Schäden. Letztere sind zudem im Glücksspiel-Surveys nicht näher spezifiziert worden – angesichts des fehlenden Konsenses für eine Definition von glücksspielbezogenen Schäden wäre dies jedoch angebracht gewesen. Browne und Rockloff (2017) weisen auf die damit verbundene Problematik deutlich hin und thematisieren in diesem Kontext die methodische Schwierigkeit der klaren Unterscheidbarkeit von Risikofaktor und Outcome.

Dass die Autoren des Glücksspiel-Surveys die Kategorie der „riskanten Spieler“ anschließend damit begründen, eine „*Abgrenzung zu den unproblematischen Spieler\*innen*“ (Buth, Meyer und Kalke 2022, S. 19) (siehe oben!) schaffen zu wollen, ist daher erstaunlich. Zumindest wäre eine Sensitivitätsanalyse angebracht, um den Einfluss dieser Kategorie auf das Ergebnis transparent zu machen. Im einfachsten Fall wäre dies eine Darstellung der Verteilung der Befragten innerhalb der Kategorie der „riskanten Spieler“, d.h. eine Übersicht, welcher Anteil der Befragten welche Kriterien im Detail erfüllt. Dies ist bedeutsam, weil die verschiedenen Items der Skala unterschiedlich schwere individuelle Belastungen durch das Glücksspielverhalten messen könnten (analog zur folgenden Erläuterung der „low harm items“ vs. „serious harm items“).

Eine weitere Begründung für die Absenkung des Schwellenwertes nach DSM-5, ab dessen Erreichen „problematisches Spielverhalten“ vorliegt, und die Erweiterung der Gruppe der „Problem-“ um „riskante Spieler“ lautet, dass Risikospieler zwar individuell geringere glücksspielbezogene Schäden erlitten als solche mit Glücksspielstörung, aber in ihrer Gesamtheit aufgrund der höheren Prävalenz größere Auswirkungen auf die Schadensbelastung der Gesellschaft hätten. Dieses Argument wird in der Literatur kontrovers diskutiert. Die Einschätzung basiert auf einer Studie von Browne, Greer u. a. (2017), die eine Kategorie „low-risk gamblers“ benutzt, welche die Kategorie der „riskanten Spieler“ inspiriert hat. Dafür ist eine Einstufung von eins bis zwei (von neun) auf dem sogenannten PGSI (Problem Gambling Severity Index) (Ferris und Wynne 2001) nötig. Die Studie gibt an, den gesamtgesellschaftlichen Einfluss von Glücksspiel anhand von Prävalenz und Stärke der Glücksspielprobleme einzuschätzen. Dabei basiert die Abschätzung auf einem Katalog von 73 Fragen<sup>2</sup>, ursprünglich entwickelt von Lang-

<sup>2</sup>Je nach Version enthält der Katalog 72 oder 73 Items.

ham u. a. (2016). Diese beinhalten glücksspielinduzierte Einschränkungen wie geringere Ausgaben für andere Freizeitaktivitäten, generell weniger verfügbares Einkommen oder geringere Spareinlagen sowie weniger Zeit für Freizeitaktivitäten, die nichts mit Glücksspiel zu tun haben.

Bei diesen „low-risk harms“ handelt es sich Delfabbro und King (2017) nach eher um Opportunitätskosten, die den einzelnen Spielern entstehen, als um wirkliche Belastungen für die Gesellschaft. Die Autoren merken an, dass es dabei um Einschränkungen geht, die auch andere gängige Freizeitaktivitäten zur Folge haben. Es sei daher fraglich, wie sich die „low-risk harms“ von den Auswirkungen von regelmäßigen Besuchen in Fußballstadien oder Shoppingtrips unterscheiden. In der Tat handelt es sich bei einigen zur Ermittlung der gesellschaftlichen Belastung von Glücksspiel abgefragten Eigenschaften um Dinge, die auf jeden leidenschaftlichen (Fußball-)Fan zutreffen können. Wer regelmäßig Eintrittskarten für das Fußballstadion erwirbt oder gar Auswärtsspiele besucht, wendet ebenfalls viel Zeit auf und gibt substanzielle Summen an Geld aus, das dadurch nicht mehr für andere Freizeitaktivitäten zur Verfügung steht. Kritisiert wird in der Literatur auch, dass „low-risk gamblers“ entgegen solchen Spielern mit Glücksspielstörung grundsätzlich und ganz konkret niedrige Zustimmungsanteile bei „serious harm items“ aufweisen und die (Gesellschafts-)Belastung nur durch Fragen zu beispielsweise Opportunitätskosten zustande kommt (Delfabbro und King 2017). Das Resultat einer ähnlichen Studie von Li u. a. (2016) bestätigt dies. Die einzigen Schäden, die mehr als 5 % aller „low-risk gamblers“ angaben, waren reduziertes verfügbares Einkommen und reduzierte Ausgaben für nicht mit Glücksspiel assoziierte Freizeitaktivitäten. Analog zu den obigen Ausführungen, ist es mindestens diskussionswürdig, ob diese Begründung die Absenkung des Schwellenwertes für „problematisches Spielverhalten“ nach DSM-5 rechtfertigt.

Im Glücksspiel-Survey 2021 wird sodann auf ein Ergebnis der Studie von Browne, Greer u. a. (2017) referenziert, wonach 85 % aller Kosten auf Ebene der Gesellschaft von Spielern mit niedrigem und moderatem Risiko einer Glücksspielstörung verursacht seien. Dieser Referenz sind zwei Kritikpunkte zu attestieren. Erstens kommt die Zahl nur zustande, indem o.g. Einschränkungen als Kosten für die Gesellschaft klassifiziert werden, ohne jedoch den gesellschaftlichen Nutzen von Glücksspiel diesen gegenüber zu stellen. Browne und Rockloff (2017) stimmen in einer Replik auf Delfabbro und King (2017) zu, dass eine solche Abwägung erfolgen sollte. Der Glücksspiel-Survey unterlässt hingegen eine klare Abgrenzung zu Opportunitätskosten und eine deutliche Definition von glücksspielbezogenen Schäden – verwiesen wird ausschließlich auf die Studie von Browne, Greer u. a. (2017), wobei unerwähnt bleibt, dass diese Problematik Gegenstand einer größeren Fachdiskussion ist.

Zweitens ist eine derartige Argumentation bereits dem Grunde nach umstritten. Nur durch die Zu-

ordnung des „riskanten Spielverhaltens“ zum Sachverhalt des „problematischen Spielverhaltens“, d.h. durch die Absenkung des Schwellenwertes für „problematisches Spielverhalten“ nach DSM-5, generiert die Studie die häufig (wenngleich nicht immer) unkritisch kommunizierte Zahl von über fünf Millionen Menschen in Deutschland, die ein Glücksspielproblem hätten bzw. *„ein riskantes Spielverhalten an der an der Schwelle zur Abhängigkeit“* zeigen sollen (Deutsches Ärzteblatt 2023). Eine breite wissenschaftliche Basis für diese Einschätzung ist, wie erläutert, nicht gegeben. Zusätzlich fehlt es bislang für die gesamtgesellschaftliche Bewertung von Glücksspiel an geeignet operationalisierten Messinstrumenten (Adams und Fiedler 2018), d.h. die generelle Messbarkeit gesellschaftlicher Kosten beruht in weiten Teilen auf nicht oder kaum prüfbar Annahmen. *„Insbesondere Faktoren wie z.B. das psychische Leid von Angehörigen pathologisch Glücksspielender sind schwer zu quantifizieren“* (Banz 2019, S. 24).

Die bereits existierenden Messinstrumente zu „gambling-related harm“ sind deutlich abgetrennt von DSM-5 Kriterien, siehe beispielsweise der durch Latvala, Browne et al. validierte *Short Gambling Harm Screen (SHGS)* (Latvala u. a. 2021). Die Autoren weisen erneut darauf hin, dass kein internationaler Konsens zur Definition von glücksspielbezogenen Schäden besteht. Jedoch bewerten sie die Definition glücksspielbezogener Schäden nach (Langham u. a. 2016) als praxisgerecht: *„any initial or exacerbated adverse consequence due to an engagement with gambling that leads to a decrement to the health or wellbeing of an individual, family unit, community or population“*. An dieser Definition wird erneut deutlich, dass einige dieser Aspekte nicht durch die DSM-5-Kriterien gemessen werden können. Neueste Studien wie Syvertsen u. a. (2023) betonen deutlich, wie wichtig es ist, problematisches Glücksspielverhalten unabhängig von glücksspielbezogenen Schäden zu messen bzw. zu erfassen, nicht zuletzt, weil Messungen von problematischem Glücksspielverhalten nicht das Spektrum möglicher glücksspielbezogener Schäden reflektieren können. Negative Folgen des Glücksspielens können auch in Abwesenheit von problematischem Spiel auftreten – dies macht deutlich, dass die Konzepte nicht vermischt werden sollten.

Neben den erläuterten niedrighwelligen Einschränkungen der Freizeit von Spielern, welche laut den Autoren u.a. zu den außerordentlichen gesamtgesellschaftlichen Auswirkungen von Glücksspiel führen sollen, sind auch die Fragen zur Einstufung in die verschiedenen Kategorien des DSM-5 in Kombination mit der Neukategorisierung des „riskanten“ als „problematisches“ Spielverhalten selbst zumindest diskussionswürdig. Da der Fragebogen nicht zur Verfügung gestellt wurde, bedient sich das vorliegende Gutachten an dieser Stelle der Fragen aus dem Erhebungsinstrument zu DSM-IV von Slecza u. a. (2015), auf das die Autoren des Glücksspiel-Surveys referenzieren<sup>3</sup>.

<sup>3</sup>Die tatsächlich im Glücksspiel-Survey 2021 angewendeten Fragen könnten also unter Umständen abweichen.

Für die Einstufung als riskanter Spieler müssen lediglich ein bis drei der neun Item-Kategorien des DSM-5 zutreffen. Da jede Item-Kategorie als zutreffend gilt, wenn mindestens eins der beiden damit assoziierten Items mit „ja“ beantwortet wird, reicht die Beantwortung eines einzigen Items mit „ja“ bereits aus, damit das eigene Spielverhalten nicht mehr als „unproblematisch“ gilt. Folgende Items erscheinen isoliert betrachtet nicht zweifelsfrei geeignet, das Vorliegen einer Spielstörung zu beurteilen; vielmehr ließen sie sie auf nahezu jede beliebige Freizeitaktivität übertragen:

- Gab es Phasen, in denen Sie sehr viel Zeit damit verbrachten, über Ihr vergangenes oder zukünftiges Glücksspielen nachzudenken/zu grübeln?
- Hatten Sie den Eindruck, dass Ihnen Ihr Glücksspielen geholfen hat, unangenehme Gefühle wie Angst oder Depression zu mildern?
- Hatten Sie das Gefühl, dass Sie gespielt haben, um vor persönlichen Problemen zu fliehen?

Ersetzt man „Glücksspiel“ durch „Stadionbesuch“, würde vermutlich auch ein Fußballfan, den ein Stadionbesuch erheitert und mit Freude erfüllt, der ggf. mit seinem Lieblingsverein sogar tagelang zu Auswärtsspielen (durch ganz Europa) reist, diese Frage(n) mit „ja“ beantworten, ohne objektiv gefährdet zu sein, ein problematisches Verhältnis zum Fußballschauen zu entwickeln („Hatten Sie den Eindruck, dass Ihnen Ihr Stadionbesuch geholfen hat, unangenehme Gefühle wie Angst oder Depression zu mildern?“). Dies bedeutet nicht, dass die genutzten Fragen in dem Kontext gänzlich ungeeignet sind und zusammengenommen keine Indikation einer Glücksspielstörung liefern können, doch isoliert betrachtet eignen sie sich nicht. Nicht grundlos ist die unterste Kategorie im Rahmen des DSM-5 für problematisches Spielverhalten („leichte Glücksspielstörung“) ursprünglich erst ab vier zutreffenden Kategorien definiert. Bislang erfolgt keine Validierung dieser Absenkung des Schwellenwertes für problematisches Spielverhalten nach DSM-5 durch die Autoren des Glücksspiel-Surveys. Allerdings bedarf die Kategorisierung von „Risiko-“ als „Problemspielern“ mindestens einer detaillierten Aufschlüsselung, um Transparenz darüber zu schaffen, welche Items in dieser Kategorie ggf. überproportional häufig bejaht wurden.

Angemessen wäre es also, die Fallzahlen bzw. Anteile von Befragten mit einer bis drei mit „ja“ beantworteten Item-Kategorien anzugeben, wie es Bühringer u. a. (2023) tun, und diese beispielsweise detailliert mit durchschnittlichen Spieleinsätzen abzugleichen. Im besten Fall erfolgt eine Langzeitbetrachtung, um nachzuweisen, dass insbesondere die Klassifikation von Befragten in die niedrigste Kategorie der „riskanten Spieler“ entweder ein Frühindikator für eine beginnende Glücksspielstörung ist (denn nur wenn ein solches Risiko nachgewiesen werden könnte, ließe sich der Begriff „riskant“

überhaupt rechtfertigen) oder zumindest stabil bleibt, d.h. dass die Befragten nicht im Laufe der Zeit wieder ein „nicht riskantes Spielverhalten“ entwickeln. So könnte auch der Kritik entgegenge- wirkt werden, dass es sich bei der Bezeichnung der Kategorie als „riskantes“ Spielverhalten und deren Zuordnung zum „problematischen Spielverhalten“ vornehmlich um Framing eines (eigentlich) unpro- blematischen Spielverhaltens handelt. Eben aufgrund des Fehlens einer solchen Betrachtung, sowie des Faktes, dass der Glücksspiel-Survey der erste in der Auftragsreihe ist (2021, 2023, 2025), wäre eine nuanciertere Darstellung und Analyse der Autoren wünschenswert gewesen.

Zusammenfassend ist festzuhalten, dass die Kategorie der „riskanten Spieler“ grundsätzlich in der Literatur umstritten ist. Die Autoren des Surveys führen keine hinlänglich überzeugenden Belege an, um die Klassifizierung des Spielverhaltens von Befragten, auf die lediglich ein bis drei DSM-5- Kriterien zutreffen, ebenfalls als „problematisch“ zu rechtfertigen. Zusätzlich erscheinen die Kriterien für eine Einstufung isoliert betrachtet zu niedrigschwellig. Die wenige von den Autoren referenzierte Literatur weist bei einer genauen Überprüfung erhebliche Kritikpunkte auf. So ist die Angabe, dass 85 % aller Kosten auf Ebene der Gesellschaft von Spielern mit niedrigem und moderatem Risiko zu Glücksspielproblemen stammen würden, sehr wahrscheinlich deutlich zu hoch. Dies führt dazu, dass der im Glücksspiel-Survey genannte Anteil an riskanten Spielern von 5,7 % (bezogen auf die Bevölkerung in Deutschland von 18 bis 70 Jahren) auf keiner klaren wissenschaftlichen Grundlage beruht. Insbesondere die gesellschaftlichen Auswirkungen des Glücksspiels dieser 5,7 % – welche laut Autoren ursächlich für eine Einstufung dieser individuell kaum Schäden erleidenden Gruppe sind – sind lediglich annahmen- und nicht faktenbasiert.

## 7.2 Kategorisierung nach Spielform

### Executive Summary

*Im Glücksspiel-Survey 2021 wird keine einheitliche Kategorisierung der einzelnen Glücksspielarten vorgenommen, die sich im gesamten Dokument wiederfindet. Die eingangs definierten acht Kategorien mit insgesamt 24 einzelnen Spielformen werden in dieser Form nicht in der Analyse im Survey benutzt. Stattdessen werden fortlaufend sowohl einzelne Spielformen als auch Kategorien ergänzt und/oder weggelassen sowie Spielformen anderen Kategorien zugeordnet. Dies erschwert bzw. verunmöglicht die Vergleichbarkeit einzelner Analysen und Glücksspielarten im Survey. Zusätzlich ist die Zuordnung der Lotterie Keno (ein Produkt der Landeslotteriegesellschaften) fragwürdig und nicht durchweg einheitlich. Entgegen der Einteilung der BZgA in den vorherigen Surveys wird Keno nicht durchgehend in der Gruppe der Lotterien geführt und ebenfalls nicht unter riskanten Glücksspielformen aufgelistet, obwohl die BZgA in ihrem letzten Glücksspiel-Survey zu einem solchen Ergebnis kommt.*

Die Autoren nehmen zu Beginn des Surveys eine grundlegende Einteilung der „insgesamt 24 abgefragten Glücksspielformen“ (Buth, Meyer und Kalke 2022, S. 20) in acht Kategorien vor, um „für einen Teil der im Ergebnisteil berichteten Befunde“ (Buth, Meyer und Kalke 2022, S. 20) in Gruppen berichten zu können. Die Einteilung erfolgte laut Autoren „aufgrund des Zugangswegs und/oder des Suchtpotentials der Spielformen“ (Buth, Meyer und Kalke 2022, S. 20).

Bemerkenswert ist hierbei, dass die vorab definierten Kategorien in ihrer Gesamtheit nicht ein einziges Mal im Ergebnisteil des Surveys analysiert werden. Stattdessen werden andere Kategorisierungen vorgenommen oder bestimmte Kategorien in der Betrachtung gänzlich ausgelassen. Im Folgenden werden daher jegliche Nennungen von Glücksspielformen im Survey auf ihre Vergleichbarkeit zur ursprünglichen Kategorisierung überprüft.

Nachdem auf Seite 20 im einleitenden Teil insgesamt 24 Spielformen gelistet werden und der Eindruck entsteht, über diese Liste hinaus würden keine anderen Glücksspiele in der Befragung vorkommen, sind es auf den Seiten 25 (Tabelle 3: 12-Monatsprävalenz einzelner Glücksspielformen nach Geschlecht und Altersgruppen) und 28 (Tabelle 4: 12-Monatsprävalenz einzelner Glücksspielformen nach Migrationshintergrund und höchstem Schulabschluss) nun 26 Spielformen, die teilweise anders zusammengefasst werden. Es kommen drei Spielformen neu hinzu: „Toto Fußballwetten“, „Live-Wetten“ und „Pferde-

wetten". Die Spielformen „Kasinospiele im Internet“ und „Sportwettenangebote im Internet“ sowie Automaten Spiele in Raststätten (hier nur „Geldspielautomaten (Spielhalle/Gaststätte)“) werden dafür nun unter „andere Sportwetten“ oder „sonstige Glücksspiele“ zusammengefasst. Es erstaunt, dass somit bereits in der ersten spezifischen Nennung von Glücksspielformen im Survey von der eingangs vorgenommenen Kategorisierung abgewichen wird.

In Tabelle 5 (*12-Monatsprävalenz zusammengefasster Glücksspielformen nach Zugangsweg und Geschlecht*) auf Seite 29 werden nur vier zusammengefasste Spielformen gezeigt, jeweils aufgeteilt in terrestrisch und online sowie gesamt. Es gibt eine Kategorie der riskanten Spiele (Automaten Spiele, Kasinospiele, Sportwetten), aber Kasino und Sportwetten bilden zusätzlich jeweils eine eigene Kategorie. Virtuelle Automaten Spiele werden allerdings hier zu den online-Kasinospielen gezählt, was zuvor nicht der Fall war.

Abbildung 7 auf Seite 32 des Surveys (*Anzahl gespielter Glücksspielformen (letzte 12 Monate) nach Alter und Geschlecht*) fasst die 24 abgefragten Einzel-Glücksspiele in acht Kategorien zusammen, die ebenfalls nicht mehr der eingangs auf S. 20 eingeführten Kategorisierung entsprechen. Keno und Sofortlotterien werden hier aus der Kategorie Lotterien (damit ist wohl die Kategorie „Lotterien (insgesamt)“ von S. 20 gemeint, da Lotto- und andere Lotterien keine eigene Kategorie sind) herausgenommen und jeweils einzeln betrachtet. Eine neue Kategorie der sonstigen Glücksspiele wird eingeführt.

Diese fortlaufende Neukategorisierung der Spiele setzt sich in Tabelle 6 auf S. 33 (*Multiple Teilnahme am Glücksspiel von Personen, die in den zurückliegenden 12 Monaten Glücksspiele spielten*) mit dem Wegfallen der Kategorie „Lotterien“ (äquivalent zu Lotterien insgesamt von S. 20) fort. Dafür ist eine neunte Kategorie („sonstige Lotterien“) ergänzt worden, die wohl der Kategorie „andere Lotterien“ von S.20 entspricht. Zusätzlich werden die „Lotto-Lotterien“ nun wieder separat betrachtet, ebenso wie die Sofortlotterien. Soweit folgt man bis hierhin tatsächlich der Klassifizierung von S. 20, allerdings wird dann ohne nähere Begründung Keno isoliert aufgeführt und nicht mehr unter Lotto-Lotterien. Die Unterscheidung Lotto-Lotterien vs. Keno und Sofortlotterien führt dazu, dass die Anteilswerte für Lotto-Lotterien niedriger ausfallen, als dies bei einer zusammengefassten Betrachtung der Fall wäre. Denn Kunden klassischer Lotto-Produkte spielen zu 57,6 % nur dieses eine Produkt und konsumieren „riskantere“ Glücksspiele allenfalls in geringem Maße (Geldspielautomaten zu 4,8 %, riskante Online-Spiele zu 12,0 % etc.). Würden die Keno- und Sofortlotterie-Kunden hier einfließen, wären diese Werte höher. Umgekehrt lenkt die getrennte Betrachtung verschiedener Lotto-Produkte den Blick darauf, dass Keno- und Sofortlotteriekunden mehr Affinität für riskantere Spielformen aufweisen, diese Produkte somit „riskanter“ sein könnten als das klassische Lotto "6 aus 49" (Keno-Spieler spielen zu 10,4

% auch an Geldspielautomaten und zu 30,9 % riskante Onlinespiele; bei den Sofortlotteriekunden liegen diese Werte bei 11,3 % und 20,6 %).

In Abbildung 13 auf Seite 38 werden selektiv Anteile von Spielern mit Glücksspielstörungen für neun verschiedene Spielarten angegeben, aber nicht für alle (bspw. Poker, aber keine anderen Kasinospiele). Auf der linken Seite in der Grafik stehen ausschließlich Angebote der staatlichen Lotterien, rechts andere, die einen deutlich höheren Anteil an Spielern mit gestörtem Spielverhalten aufweisen. Es geht aus dieser Grafik nicht hervor, wie die Schweregrade der Glücksspielstörungen innerhalb der nicht aufgeführten Spielarten verteilt sind. Zwar wird von den Autoren angemerkt, dass bei der Interpretation zu berücksichtigen sei, dass ein Teil der Spieler an mehreren Glücksspielen teilnehme. Beispielhaft wird genannt, dass "17,1 % der Teilnehmer\*innen an Lotterien parallel auch riskante Glücksspiele" (Buth, Meyer und Kalke 2022, S. 38) spielen. Diese beiden Kategorien finden sich allerdings nicht in dieser Form in der Abbildung wieder. Es wird also nicht klar, welche Anteile in der Grafik auf Spieler multipler Spielarten zurückzuführen sind. Zudem wird dem Leser aufgrund der selektiven Darstellung die Bildung einer fundierten eigenen Bewertung verwehrt.

Eine weitere Auffälligkeit hinsichtlich der Kategorisierung liegt auf Seite 39 in Tabelle 8 vor. Hier wird in ähnlicher Darstellung wie in Tabelle 5 auf Seite 29 (zusätzlich mit der Ergänzung einer Gesamtbetrachtung „jegliche Glücksspielteilnahme“) des Surveys der „*Schweregrad der Glücksspielprobleme (DSM-5) einzelner Glücksspielgruppen nach Zugangsweg*“ gezeigt. Dabei werden die Glücksspielarten in vier Kategorien eingeteilt: riskante Glücksspiele, Kasinospiele (inkl. Online-Automatenspiele), Sportwetten und Lotterien. Unterschieden wird jeweils noch in online und terrestrischen Zugangsweg sowie insgesamt. Zusätzlich wird im Bereich Lotterien eine Zeile speziell für „ausschließlich Lotto-Lotterien“ (ohne Rubbellose) angegeben, die deutlich niedrigere Anteile von Spielern mit Glücksspielstörung aufweist als andere Kategorien. Solch eine Differenzierung nach „ausschließlich Lotto-Lotterien“ gibt es im Survey sonst kein einziges Mal. Unverständlich ist hierbei, dass diese Kategorie nur Spieler enthält, die ausschließlich (staatliche) Lotterien spielen, dabei also an keinen anderen Glücksspielen teilnehmen. Damit erfasst man lediglich 57,6 % der Lotto-Spieler (siehe Tabelle 6, S.33 im Survey). Würde man dieses Vorgehen für die gesamte Tabelle übernehmen, deckte man beispielsweise nur 7,4 % der terrestrischen Sportwettenden ab. Denn die anderen in der Tabelle dargestellten Kategorien enthalten mutmaßlich durchgängig Spieler, die mehrere Glücksspiele spielen, z.B. Lotto-Produkte und Sportwetten (zumindest ist dies nicht anders angegeben). Damit verhindern die Autoren jegliche Vergleichbarkeit innerhalb der Tabelle – ohne dies transparent zu kommunizieren – und fördern damit Fehlinterpretationen.

Dass Lotto somit eine Quote an unproblematischen Spielern von 83,1 % erreicht und damit die höchste aller Glücksspielarten in der Tabelle, ist wenig verwunderlich. Nähme man die restlichen 42,4 % der Lotto-Spieler hinzu, die auch andere Glücksspiele spielen, läge die Zahl mutmaßlich höher. So spielen 71,9 % aller Spieler von terrestrischen Sportwetten und 60,1 % aller Spieler von riskanten Onlinespielen auch Lotto. Der Anteil an problematischen Spielern liegt bei diesen Arten bei ungefähr 30 %. Eine Aufnahme dieser Befragten in der Kategorie Lotto würde also sehr wahrscheinlich zu erhöhten Spielstörungsprävalenzen bei Lotto führen. Unklar bleibt, was die Motivation der Autoren diesbezüglich ist, denn es existiert keine Erläuterung. Wichtig zu beachten ist, dass wie in Abschnitt 6.2 erläutert, aus dem relativ hohen Anteil von unproblematischen Lotto-Spielern aufgrund der Natur der Daten sowie der fehlenden statistischen Analyse keine Rückschlüsse über die „Gefährlichkeit“ der einzelnen Glücksspielformen abgeleitet werden können.

Diskussionswürdig erscheint zudem die Kategorisierung der Lotterie Keno des deutschen Lotto- und Totoblocks. Denn die Lotterie wird, entgegen der Einteilung der BZgA in den vorherigen Surveys, nicht durchgehend in der Gruppe der Lotterien geführt und ebenfalls nicht unter riskanten Glücksspielformen aufgelistet, obwohl die BZgA in ihrem letzten Glücksspiel-Survey auf ein solches Ergebnis kommt: *„Insgesamt zeigt sich über alle Datenerhebungen, dass besonders Automaten- und Casinospiele mit einem deutlich höheren Risiko für auffälliges bzw. risikoreiches und mindestens problematisches Glücksspiel verbunden sind, während Lotterieangebote bis auf die Lotterie Keno ein deutlich geringeres Risiko aufweisen.“* (Banz 2019, S. 130). Da die Einteilung in Glücksspielkategorien nach Angabe der Autoren des Glücksspiel-Surveys 2021 aufgrund des Zugangswegs und/oder des Potentials für Spielstörungen der Spielformen erfolgte, stellt sich umso mehr die Frage, warum Keno als weniger risikoreiche Lotterie kategorisiert ist.

Falls die Autoren die Risikohaftigkeit von Keno erst (selbst) messen wollten, wäre es vor dem Hintergrund der Betrachtung des Glücksspiel-Surveys 2021 als initialem Messpunkt unter Beachtung der gebotenen Transparenz legitim, Keno konsequent den (Lotto-) Lotterien zuzuordnen. *„Keno gilt aufgrund der täglichen Möglichkeit zur Spielteilnahme als Lotterie, an der gesperrte Spielerinnen und Spieler nicht teilnehmen dürfen“*, ist allerdings selbst auf der Webseite *lotto.de* unter Bezugnahme auf den Glücksspielstaatsvertrag zu lesen (Toto-Lotto Niedersachsen GmbH 2023). Eine solche Konsequenz ist jedoch nicht gegeben. Stattdessen kommt Keno häufig – aber nicht ausschließlich und auch nicht in der initialen Kategorisierung auf Seite 20 – eine eigene Kategorie unabhängig von den anderen Lotto-Produkten oder den Lotterien insgesamt zu. Dies erscheint willkürlich und wird von den Autoren nicht begründet. Die vorgenommene isolierte Betrachtung von Keno senkt häufig den berichteten Anteil an Spielern mit Glücksspielstörung (oder anderen Faktoren) bei Produkten der Landeslotte-

riegesellschaften. Grundsätzlich ist es infolge dieser Praxis im Survey schwierig zu erkennen, ob Keno bei gewissen Analysen nun zu den Lotterien gezählt wird oder nicht. So wird auf Seite 39 in Tabelle 8 nicht erläutert, ob Keno zu Lotto-Lotterien zählt oder wie zuvor geschehen von den Lotto-Produkten getrennt behandelt wird und daher in der Tabelle nicht auftaucht.

Insgesamt behindert die inkonsistente und bestenfalls willkürliche Umkategorisierung die Vergleichbarkeit zwischen den Glücksspielarten und -kategorien im Gutachten und stiftet Verwirrung beim Leser. Zusätzlich ist unklar, warum die Autoren keiner klaren Linie folgen und fortlaufend unterschiedliche Kategorien in Tabellen und Grafiken darstellen.

## 7.3 Nullpunktmessung

### Executive Summary

*Der Glücksspielsurvey 2021 bildet entgegen seiner erklärten Intention aufgrund der Vermischung von Daten vor und nach dem Inkrafttreten des Glücksspielstaatsvertrags keine valide Nullpunktmessung in Bezug auf den Glücksspielstaatsvertrag 2021. Somit können mit dem gewählten Studiendesign (inklusive der folgenden Erhebungen), anders als intendiert, keine Auswirkungen des aktuellen Glücksspielstaatsvertrags erfasst werden. Weiter wird die zum Zeitpunkt der Befragung anhaltende Covid-19-Pandemie und die damit verbundene mangelnde Vergleichbarkeit mit den Folgebefragungen nicht thematisiert – dies muss im nächsten Survey definitiv diskutiert werden.*

Der Glücksspiel-Survey 2021 ist qua Studiendesign ungeeignet, eine trennscharfe Nullpunktmessung für die Evaluierung des Glücksspielstaatsvertrags durchzuführen. Dabei war es ein explizites Ziel der Studienautoren, „einen soliden Ausgangspunkt für die Erfassung von Auswirkungen der neuen gesetzlichen Regelungen auf das Glücksspielverhalten der Bevölkerung“ (Buth, Meyer und Kalke 2022, S. 8) zu bieten. Mit der neuen gesetzlichen Regelung ist der Glücksspielstaatsvertrag 2021 gemeint, der am 1. Juli 2021 in Kraft trat, insbesondere verschiedene Online-Glücksspiele legalisierte und die behördlichen Eingriffsmöglichkeiten erweiterte.

Die fehlende Eignung des Studiendesigns als Nullpunktmessung ergibt sich aus dem Zeitpunkt der Befragung (3. August – 16. Oktober 2021), die nach dem Inkrafttreten des neuen Glücksspielstaatsvertrags (1. Juli 2021) durchgeführt wurde. Da die Frage bezüglich einer Glücksspielteilnahme auf die letzten 12 Monate abzielte, lagen im abgefragten Zeitraum somit – je nach genauem Zeitpunkt der Befragung – zwei bis dreieinhalb Monate nach der neuen Regelung und zehn bis achteinhalb Monate vor der neuen Regelung. Damit wird es unmöglich, im Glücksspiel-Survey 2021 sowie in den folgenden Glücksspiel-Surveys Veränderungen dem Glücksspielstaatsvertrag (kausal) zuzuweisen, denn es gibt keinen Datenpunkt vor Inkrafttreten des Vertrags, also genau keine Nullpunktmessung, die der Survey vorgibt zu sein. Zur Veranschaulichung betrachte man Figur 7.1, die den hypothetischen Anteil an Glücksspielteilnehmenden in den Jahren 2020 bis 2022 zeigt. In dem gezeigten Szenario ist der Anteil an Glücksspielenden bis zum neuen Glücksspielstaatsvertrag stabil bei 20 %, steigt durch den Glücksspielstaatsvertrag auf 40 % und stagniert dort. Ein adäquates Studiendesign, das die Nullpunktmessung vor dem Glücksspielstaatsvertrag durchführt, würde diese Werte entsprechend messen.

Survey kann Auswirkungen des Glücksspielstaatsvertrags qua Design nicht akkurat erfassen  
Hypothetisches Szenario - Daten sind nicht real

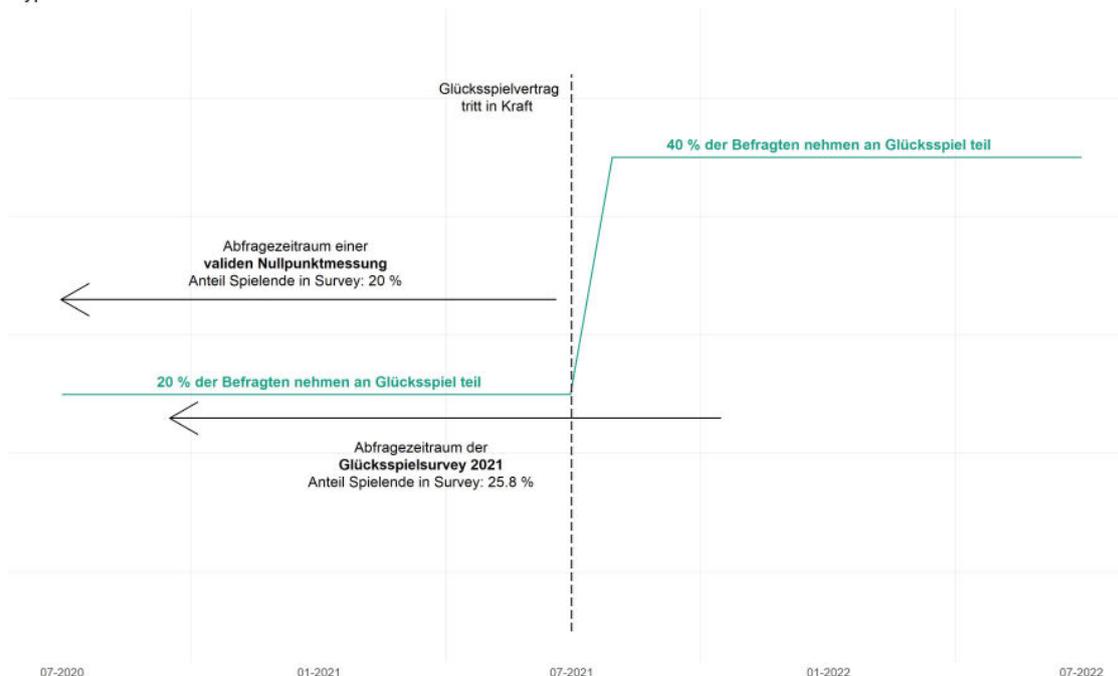


Abbildung 7.1: Auswirkung der unterschiedlichen Befragungszeiträume auf die Messung rund um den Glücksspielstaatsvertrag 2021

Der Glücksspiel-Survey 2021 hingegen ist ungeeignet, den realen Anteil akkurat zu messen, da ein Teil der Auswirkungen des Glücksspielstaatsvertrags in der „Nullpunktmessung“ enthalten ist und diese damit ad absurdum führt. Etwaige Unterschiede zu späteren Messungen können daher nicht akkurat dem Glücksspielstaatsvertrag zugerechnet werden.

Dazu kommt die Problematik, dass der Staatsvertrag aufgrund des „Übergangsbeschlusses“ der Bundesländer („Umlaufbeschluss der Chefinnen und Chefs der Staats- und Senatskanzleien der Länder vom 8. September 2020“) bei einigen Spielformen zumindest in Teilen vorweggenommen wurde. Dies betrifft den Zeitraum ab Oktober bzw. Dezember 2020. Auch aus diesem Grund kommt die Untersuchung zu spät, um eine Nullpunktmessung darstellen zu können.

Des Weiteren fand die Befragung während der Corona-Pandemie statt, deren Auswirkungen die Autoren des Glücksspielsurveys im Fazit folgendermaßen zusammenfassen: *„Die in diesem Bericht präsentierten Ergebnisse lassen daher nur eine begrenzte Vergleichbarkeit mit den Befunden der früheren Glücksspiel-Surveys zu. Die Jahre 2020 und 2021 waren von der Coronapandemie und deren Auswirkungen auf das politische, wirtschaftliche und gesellschaftliche Leben geprägt. Unter anderem mussten*

*viele terrestrische Glücksspielstätten aufgrund pandemiebedingter Lockdowns für längere Zeiträume schließen. Auch wenn die Auswirkungen der Pandemie auf das Spielverhalten und die glücksspielassoziierten Probleme nicht exakt quantifiziert werden können, sind diese bei der Interpretation der Ergebnisse des Glücksspiel-Surveys 2021 zu berücksichtigen.*“ (Buth, Meyer und Kalke 2022, S. 49).

Den Autoren ist demnach offensichtlich bewusst, dass die Pandemie das Spielverhalten und die Prävalenz an Glücksspielproblemen beeinflusst. Diese Kontextualisierung der Ergebnisse mit Verweis auf die mangelnde Vergleichbarkeit mit den Vorjahren, ist positiv zu bewerten. Umso mehr verwundert die fehlende entsprechende Einordnung bezüglich der „Nullpunktmessung“, die der Survey sein will: Denn wenn der Survey (abgesehen von der veränderten Methodik) nur eingeschränkt mit prä-pandemischen Umfragen vergleichbar ist, inwieweit ist er dann mit post-pandemischen Umfragen – also den für 2023 und 2025 geplanten Surveys – vergleichbar? Diese Frage beantwortet der Glücksspiel-Survey 2021 nicht, diese Diskussion sollte allerdings in den nächsten Surveys geführt werden.

Zusammengefasst ist der Glücksspiel-Survey 2021 aufgrund der Vermischung von Daten vor und nach dem Inkrafttreten des Glücksspielstaatsvertrags keine valide Nullpunktmessung und somit können mit dem gewählten Studiendesign (inklusive der folgenden Erhebungen) keine Auswirkungen des Glücksspielstaatsvertrags erfasst werden. Der Survey ist mithin ungeeignet, das selbstgesetzte Ziel einer „Erfassung von Auswirkungen der neuen gesetzlichen Regelungen auf das Glücksspielverhalten der Bevölkerung“ (Buth, Meyer und Kalke 2022, S. 8) zu erreichen. Weiterhin bleibt offen, inwieweit der Glücksspiel-Survey 2021 aufgrund seines Befragungszeitraums während der Corona-Pandemie mit den nachfolgenden Surveys vergleichbar ist.

## 7.4 Evaluierung der Legalisierung des Online-Glücksspiels

### Executive Summary

*Der Survey ist aufgrund seiner mangelnden Differenzierung zwischen legalem und illegalem Glücksspiel nicht geeignet, das selbstgesetzte Ziel einer „Erfassung von Auswirkungen der neuen gesetzlichen Regelungen auf das Glücksspielverhalten der Bevölkerung“ (Buth, Meyer und Kalke 2022, S. 8) zu erreichen. Zudem steigt die Gefahr für Fehlinterpretationen oder Scheineffekte durch die Nichterhebung des Faktors „Legalität“ der individuell genutzten Glücksspielart.*

Der Glücksspiel-Survey 2021 will die „Auswirkungen der neuen gesetzlichen Regelungen auf das Glücksspielverhalten der Bevölkerung“ (Buth, Meyer und Kalke 2022, S. 9) erfassen. Dies ist neben der in Abschnitt 7.3 und 6.1 beschriebenen Nullmessungs- sowie Querschnitts-Problematik auch aufgrund der mangelnden Differenzierung zwischen legalem und illegalem Online-Glücksspiel nicht möglich.

Das Problem lässt sich an einem einfachen Beispiel illustrieren: Person A spielt ungeachtet der Legalisierung des Online-Glücksspiels weiterhin illegales Glücksspiel im Internet, während Person B nur legales Online-Glücksspiel spielt. Nun werden beide befragt und es stellt sich heraus, dass Person A eine Glücksspielstörung und Person B – der legale Spieler – ein sogenannter unproblematischer Spieler ist. Nach dem Glücksspiel-Survey würde die Prävalenz einer Glücksspielstörung unter den Teilnehmern des untersuchten Glücksspiels somit bei 50 % liegen. Differenziert man hingegen nach der Legalität des Glücksspiels, wäre der Anteil an Spielern mit einer Glücksspielstörung bei illegalem Glücksspiel bei 100 % und bei legalem Glücksspiel bei 0 %. Wenngleich es sich um ein überspitztes Beispiel handelt, verdeutlicht es doch die Konsequenzen einer mangelnden Differenzierung zwischen legalem und illegalem Glücksspiel.

Folglich können die Auswirkungen der Legalisierung des Online-Glücksspiels nicht mit dem Glücksspiel-Survey erfasst werden, denn um den Effekt einer Legalisierung zu messen, muss konsequent erfasst werden, wer legal und wer illegal spielt<sup>4</sup>. Zudem bräuchte man trennscharfe Datenpunkte einer Panelbefragung vor und nach der Legalisierung, was, wie in den Abschnitten 7.3 Nullmessung und 6.1 Querschnitt beschrieben, nicht der Fall ist. Doch selbst wenn dies gegeben wäre, könnte der Effekt der Legalisierung nicht identifiziert werden: Betrachte man erneut zur Illustration Person C, einen Spieler mit leichter Spielstörung, der vor dem Glücksspielstaatsvertrag illegal Online-Glücksspiel spielte.

<sup>4</sup>Es sei denn, man wolle argumentieren, dass mit der Legalisierung die Nutzung illegaler Glücksspiele auf Null gesunken sei, was aber als unrealistisch einzuordnen wäre.



Abbildung 7.2: Keine Differenzierung nach Legalität des Glücksspiels

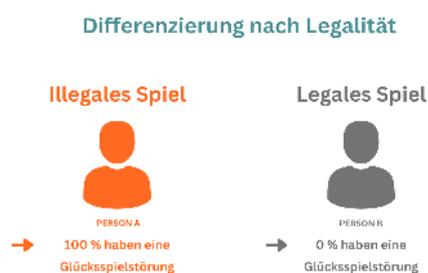


Abbildung 7.3: Differenzierung nach Legalität des Glücksspiels

Aufgrund der Legalisierung des Online-Glücksspiels wechselt Person C ins legale Glücksspiel und wurde aufgrund der Regularien zu einem unproblematischen Spieler. Ohne Differenzierung nach der Legalität würden die Daten nur zeigen, dass Person C ohne Veränderung des Glücksspielverhaltens – der Wechsel zu legalem Glücksspiel wird nicht erfasst – seine Glücksspielstörung verloren hat. Erklärt werden könnte die Veränderung somit nicht, auch nicht durch den kausal ursächlichen Glücksspielstaatsvertrag, da die dafür notwendige Differenzierung nach der Legalität des genutzten Glücksspielangebots fehlt.

Insgesamt lässt sich festhalten, dass der Glücksspiel-Survey 2021 aufgrund seiner mangelnden Differenzierung zwischen legalem und illegalem Glücksspiel nicht geeignet ist, die Auswirkungen des Glücksspielstaatsvertrags zu evaluieren. Zudem steigt die Gefahr für Fehlinterpretationen oder Scheineffekte durch die Nichterhebung des Faktors Legalität.

## 8 | **Transparenz**

### Executive Summary

*Die Autoren des Glücksspiel-Surveys haben eine Anfrage der Gutachter auf Bereitstellung der detaillierten Survey-Daten, des Feldberichts und des Fragebogens abgelehnt, obwohl sie in ihrem Forschungspapier zum Survey (Buth, Meyer und Kalke 2022) noch ihre grundsätzliche Bereitschaft zur Herausgabe dieser Daten bekundet hatten. Schriftlich begründet wurde dies mit der angeblich fehlenden Transparenz der Anfrage seitens der Gutachter. Aus diesem Grund lassen sich Berechnungen im Survey teilweise nicht überprüfen sowie die Datenqualität nicht abschließend einschätzen. Die Haltung der Autoren verstößt gegen den Kodex „Leitlinien zur Sicherung guter wissenschaftlicher Praxis“ der Deutschen Forschungsgemeinschaft (DFG), der explizit die Replizierbarkeit von Forschung durch Herausgabe der Forschungsdaten und relevanten Materialien betont sowie die Bedeutung einer kritischen wissenschaftlichen Diskussion über Forschungsergebnisse hervorhebt. Durch die Verweigerung der Herausgabe der Daten auch auf Nachfrage erschweren die Autoren des Surveys auch eine offene und kritische wissenschaftliche Debatte über ihre Studienergebnisse. Darüber hinaus missachten die Survey-Autoren auch die Bedeutung einer kritischen Diskussion der Limitation der eigenen wissenschaftlichen Arbeit, wie sie die DFG ebenfalls in ihrem Kodex hervorhebt: „Zu den Prinzipien gehört es [...], alle Ergebnisse konsequent selbst anzuzweifeln sowie einen kritischen Diskurs in der wissenschaftlichen Gemeinschaft zuzulassen und zu fördern“ (Deutsche Forschungsgemeinschaft 2019, S. 9). Der Glücksspiel-Survey lässt nicht erkennen, dass sich die Autoren einer transparenten und angemessenen kritischen Auseinandersetzung mit ihrer Arbeit stellen, denn die im Gutachten skizzierten, offenkundigen Limitationen der Studie werden von ihnen weder im Survey noch in der begleitenden Kommunikation angegeben und diskutiert.*

Eine schriftliche Anfrage der Gutachtenersteller auf Bereitstellung der dem Glücksspiel-Survey zugrunde liegenden Daten sowie des benutzten Fragebogens und eines zugehörigen Feldberichts haben die Autoren abgelehnt. Nach mehrfacher schriftlicher und ausführlicher Beantwortung von sukzessive neuen (Rück-)Fragen der Autoren zu Inhalt, Rahmen und Förderern der Arbeit, erfolgte eine Ablehnung der Anfrage auf Herausgabe der Forschungsdaten mit Verweis auf mangelnde Transparenz seitens der Gutachter.

Die Mitteilung, das Gutachten werde von „Akteuren aus dem Bereich des Glücksspielwesens – und zwar spielformübergreifend“ finanziert, wurde von den die Autoren des Glücksspiel-Surveys als nicht hinreichend eingestuft und es wurde auf der Nennung der genauen Namen der „Auftraggeber bzw. finanziellen Förderer“ bestanden; eine Bereitstellung der Daten wurde abgelehnt. Das Argument, dass zur Vorbeugung einer Einflussnahme durch Dritte auf das Gutachten über die Auftraggeber eine solche Angabe bis zur Veröffentlichung nicht angezeigt ist, wurde nicht akzeptiert. Unabhängig von den konkreten Förderern befremdet die Verweigerung der Überlassung von Daten und Hintergrundinformationen aufgrund der Finanzierung dieses Gutachtens durch (beliebige) Akteure der Glücksspielbranche. Einerseits wird der Glücksspiel-Survey 2021 ebenfalls durch einen Akteur der Glücksspielbranche finanziell gefördert (DLTB), zum anderen wird diesem Gutachten durch diese Ablehnung a priori und unbegründet eine seriöse und neutrale wissenschaftliche Herangehensweise (Vorwurf des „Auftragsgutachtens“) abgesprochen.

Der zweite von den Autoren herangeführte Grund zur Ablehnung der Bereitstellung der Daten ist die fehlende Transparenz hinsichtlich der genauen Fragestellung des Gutachtens. In der Anfrage wurden sowohl der Titel des Gutachtens als auch die Zielsetzung einer „Analyse des derzeitigen Stands des wissenschaftlichen Monitorings in Bezug auf das Glücksspielverhalten in Deutschland“ explizit benannt; gleichwohl stuften die Autoren das als nicht ausreichende und unklare Ausführung ein. Ohne die Möglichkeit einer näheren Beantwortung der Fragen einzuräumen, entschieden die Autoren nach zwei E-Mails, der Bitte um Bereitstellung der abgefragten Materialien „final“ nicht nachzukommen.

Aufgrund der abgelehnten Bereitstellung der angefragten Materialien lassen sich die Berechnungen im Glücksspiel-Survey 2021 im vorliegenden Gutachten nicht vollständig überprüfen. Insbesondere die vielen Anteilen aufgrund geringer Stichprobengrößen zugrundeliegende Unsicherheit lässt sich ohne die Originaldaten nur approximativ abschätzen. Auch Bühringer u. a. (2023) sehen bzgl. des Glücksspiel-Surveys 2021 „methodologische Einschränkungen“ (S. 203) aufgrund von Intransparenz. Hinsichtlich der Zusammensetzung der Gesamtstichprobe schreiben sie, diese sei „nicht nachvollziehbar dargestellt“ (S. 3). Bemerkenswerterweise signalisieren die Autoren in ihrem Forschungspapier zum Glücksspiel-

Survey eine grundsätzliche Bereitschaft zur Herausgabe der Daten an Dritte: „The authors agree to make their data available upon reasonable request“ (Meyer, Kalke und Buth 2023, S. 15).

Bei der von den Autoren erhobenen Datenart, den Befragungsdaten, ist es gängige wissenschaftliche Praxis, anonymisierte Daten, Codebuch und Fragebogen für Dritte bereitzustellen. Vielerorts geschieht dies bereits proaktiv auf frei zugänglichen Servern, ohne dass eine explizite Anfrage erforderlich wäre. Wie in der Literatur festgestellt wurde, mag die Formulierung „*data available upon reasonable request*“ allerdings in einigen Fällen eher eine Phrase darstellen, die lediglich den Eindruck der Datenverfügbarkeit erwecken soll, insbesondere weil dieser Hinweis von Förderern oder Journals vorgeschrieben ist. Bei einer Untersuchung der tatsächlichen Bereitstellung von Daten durch Autoren von in den Fachzeitschriften *Nature* und *Science* veröffentlichten Werken fanden die Autoren heraus, dass weniger als die Hälfte aller Daten, die angeblich auf (begründete) Anfrage erhältlich waren, auch zur Verfügung gestellt wurden (Tedersoo u. a. 2021).

Dabei ist die Bereitstellung von Forschungsdaten von enormer wissenschaftlicher Bedeutung. „*Dass Ergebnisse beziehungsweise Erkenntnisse durch andere Wissenschaftlerinnen und Wissenschaftler repliziert bzw. bestätigt werden können [...], ist – abhängig von dem betroffenen Fachgebiet – essenzieller Bestandteil der Qualitätssicherung*“, stellt auch die Deutsche Forschungsgemeinschaft (DFG) in ihrem „Kodex: Leitlinien zur Sicherung guter wissenschaftlicher Praxis“ von 2019 fest (Deutsche Forschungsgemeinschaft 2019, S. 15). Im vorliegenden Fall des Glücksspiel-Surveys ist die Bereitstellung der Daten eine unerlässliche Voraussetzung für die Replikation oder Bestätigung der Ergebnisse. Durch die Verweigerung der Herausgabe der Daten stehen die Autoren somit einer offenen und kritischen wissenschaftlichen Debatte im Weg und verstoßen gegen allgemeine wissenschaftliche Standards. Im Kodex der DFG heißt es weiter: „*Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nachnutzbarkeit hinterlegen Wissenschaftlerinnen und Wissenschaftler, wann immer möglich, die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien – den FAIR-Prinzipien („Findable, Accessible, Interoperable, Re-Usable“) folgend – zugänglich in anerkannten Archiven und Repositorien*“ (Deutsche Forschungsgemeinschaft 2019, S. 19). Diesem Transparenzgrundsatz folgen die Autoren explizit nicht und erläutern auch nicht ihre Beweggründe.

Darüber hinaus betont die DFG in ihrem Kodex die Bedeutung einer kritischen Diskussion der Limitation der wissenschaftlichen Arbeit: „*Zu den Prinzipien gehört es [...] alle Ergebnisse konsequent selbst anzuzweifeln sowie einen kritischen Diskurs in der wissenschaftlichen Gemeinschaft zuzulassen und zu fördern*“ (Deutsche Forschungsgemeinschaft 2019, S. 9). Im Glücksspiel-Survey 2021 ist nicht zu erkennen, dass sich die Autoren einer transparenten und angemessenen kritischen Auseinanderset-

zung mit ihrer Arbeit stellen, denn Limitationen werden nicht diskutiert. Ferner kann die kritische Auseinandersetzung in der Wissenschaft nicht erfolgen, wenn die Daten nicht frei zugänglich gemacht werden.

Diese Praxis guter wissenschaftlicher Arbeit gilt insbesondere für Studien wie den Glücksspiel-Survey 2021, die für die breite Rezeption in der Öffentlichkeit und durch politische Akteure gedacht sind. Da allerdings für solche, nicht in anerkannten Journals publizierten Studien standardmäßig kein Peer-Review erfolgt, ist die Nennung und Diskussion der Limitationen keine Pflicht, sondern obliegt der freien Entscheidung der Studienautoren. Es gibt jedoch keine überzeugenden Argumente, die gegen eine Nennung und Diskussion der studieninhärenten Limitationen sprechen (Babbie 2015). Vielmehr gibt es etliche Argumente, die für einen transparenten Umgang mit den Schwächen einer Studie sprechen. Im Folgenden werden vier kurz genannt:

Erstens dienen die Nennung und Diskussion der Limitationen dazu, die Vertrauenswürdigkeit der wissenschaftlichen Forschung zu stärken. Wenn Forscher ihre Arbeit mit einer offenen Anerkennung der Unvollkommenheiten präsentieren, zeugt das von Integrität und dem Willen, ein ehrliches und realistisches Bild ihrer Erkenntnisse zu zeichnen. Dieses Maß an Transparenz unterstreicht die Glaubwürdigkeit der Forschung und stärkt das Vertrauen der Öffentlichkeit in die Wissenschaft. Zweitens ermöglichen die Nennung und Erörterung von Einschränkungen eine fundierte öffentliche Diskussion über die Forschungsergebnisse. Die Kenntnis der Grenzen einer Studie hilft der Öffentlichkeit und politischen Akteuren, deren Ergebnisse und Implikationen besser zu verstehen und zu interpretieren. Drittens kann die Offenlegung von Limitationen dazu beitragen, dass Forschungsergebnisse nicht über- oder fehlinterpretiert oder gar missbraucht werden. Es ist elementar, die Öffentlichkeit klar und korrekt darüber zu informieren, was die Studie tatsächlich aussagt, und welche Schlussfolgerungen gezogen werden können – und welche nicht. Viertens trägt die Transparenz hinsichtlich der Limitationen zur Förderung des wissenschaftlichen Fortschritts bei. Sie liefert anderen Wissenschaftlern wichtige Hinweise darauf, welche Aspekte weiterer Untersuchungen bedürfen, und bildet so die Grundlage für nachfolgende Studien.

Die Autoren des Glücksspiel-Surveys 2021 haben sich gegen eine transparente Nennung und Diskussion der Limitationen im Survey entschieden – eine aus wissenschaftlicher Sicht fragwürdige Entscheidung. Die Beweggründe der Studienautoren sind nicht nachvollziehbar, insbesondere wenn man berücksichtigt, dass sie in einem Forschungsartikel im *International Journal of Gambling Studies* Daten aus dem Glücksspiel-Survey 2021 nutzen und verschiedene Limitationen nennen und diskutieren:

„A considerable limitation of online samples is that online respondents cannot be randomly recruited

from the population. The corresponding selection Bias can lead to Biases in prevalence estimates. Even if attempts are made to minimize this Bias (e.g. random selection of respondents in panel samples, subsequent weighting, etc.), this problem can hardly be resolved completely. Furthermore, the COVID-19 pandemic and the new State Gambling Treaty may also have affected the results. Pandemic-related lockdowns occurred in 2020 and 2021, accompanied by massive restrictions in everyday life and their social, emotional, and financial effects. Terrestrial gambling venues were forced to close and there were fewer options for betting on sports outcomes (Georgiadou et al., 2021). With the State Gambling Treaty in force since July 2021, there have also been numerous changes in the market, such as the legalization of online gambling as a major novelty which is associated with a higher availability of gambling services. Finally, although the DSM-5 criteria achieve satisfactory reliability, validity, and classification accuracy (Stinchfield et al., 2016), they may be less suitable for epidemiological studies as they were originally developed for the clinical context“ (Meyer, Kalke und Buth 2023).

All diese Limitationen wurden in dem Glücksspiel-Survey 2021 weder benannt noch diskutiert, obwohl die Studienautoren sich dieser offensichtlich bewusst waren. Dabei wäre die Diskussion der Limitationen in einer für die Öffentlichkeit und politische Akteure gedachten Studie sogar noch relevanter als in einem wissenschaftlichen Artikel, der von einem Fachpublikum gelesen und entsprechend eingeordnet werden kann – der durchschnittliche Bundesbürger, Journalist oder Politiker kann das nicht. Der Vorsitzende des ADM (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.) betonte in diesem Kontext zum wiederholten Male: „*Es ist heute wichtiger denn je, dass wir uns an Regeln für die Veröffentlichung von Umfrageergebnissen halten, um allen Stakeholdern die Möglichkeit zu geben, die Ergebnisse einer Befragung angemessen einzuschätzen und zu bewerten. Die Verantwortung für eine korrekte Kommunikation liegt bei denjenigen, die die Ergebnisse veröffentlichen. Gleichzeitig sollten die Medien die Ergebnisse kritisch hinterfragen und sich über ihre Entstehung informieren, insbesondere wenn eine große öffentliche Aufmerksamkeit vermutet wird oder es sich um kontroverse Themen handelt. Dies erfordert natürlich ein hohes Maß an Datenkompetenz. Wenn es Zweifel an der Aussagekraft einer Studie gibt, muss von einer weiteren Verbreitung abgesehen werden*“ (ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. 2023). Die Verantwortung für die Kommunikation von Studienergebnissen liegt also explizit bei den Studienautoren. Dennoch müssen Medien und Politik die Studien kritisch hinterfragen – das dafür erforderliche Maß an Datenkompetenz ist jedoch (noch) nicht vorhanden. Umso größer wiegt also die Verantwortung der Forscher. Die Auswirkungen des Ausbleibens einer korrekten und umfangreichen Kommunikation seitens der Wissenschaftler werden im nächsten Kapitel thematisiert.

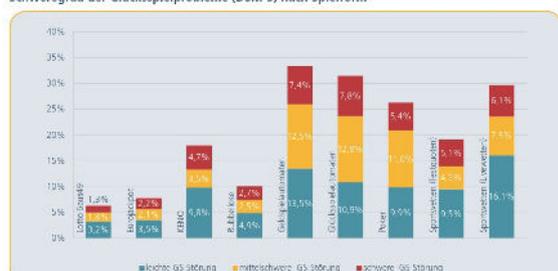
Die Glücksspiel-Surveys der BZgA vergangener Jahre zeigen sowohl, dass es durchaus möglich ist, ei-

ne transparente und kritische Auseinandersetzung mit Befragungsergebnissen zu betreiben, als auch, dass eine Veröffentlichung der erhobenen Daten zweifelsfrei möglich ist. So sind Datensätze, Codebücher und Fragebögen früherer Glücksspiel-Surveys im GESIS Data Archive ohne vorherige Anfrage frei zugänglich (allerdings nur bis einschließlich des Surveys des Jahres 2013 (Bundeszentrale für gesundheitliche Aufklärung 2017)). Vor diesem Hintergrund erscheint die Ablehnung durch die Autoren des Surveys von 2021 mit dem Hinweis auf Transparenz umso fragwürdiger, da die BZgA keinerlei Informationen von Nutzern der Daten fordert, bevor der Zugang ermöglicht wird.

## Relative und absolute Zahlen

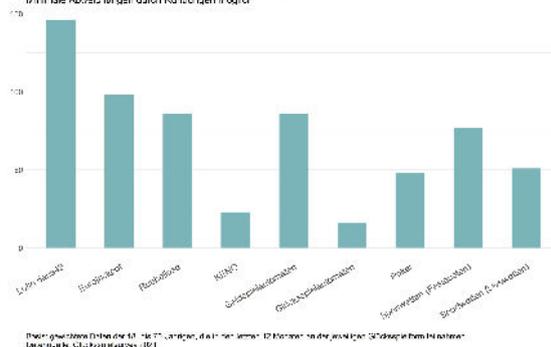
Zur Transparenz gehört es auch, nicht nur relative, sondern auch absolute Fallzahlen zu kommunizieren. Dies muss nicht notwendigerweise im Survey selbst geschehen, diese Informationen sollten allerdings mindestens im Appendix (oder in anderer Form) vorhanden sein. Insbesondere bei der Beurteilung von Gesamtauswirkungen sind absolute Zahlen unverzichtbar und können einen wichtigen Kontext liefern. Zum Beispiel kann ein kleiner Prozentsatz immer noch eine große absolute Anzahl repräsentieren (und vice versa), was für den Menschen allerdings schwer greifbar ist. So hatte zum Beispiel das Coronavirus laut dem Robert-Koch-Institut (RKI) eine Infektions-Sterbe-Rate<sup>1</sup> von 0,4 % bis 0,9 % (Stand 23.11.2021). Diese Zahl wirkt auf den ersten Blick klein, ganz im Gegensatz zur absoluten Zahl von knapp 100.000 Fällen. Daher ist es wichtig, beide Zahlenarten zu kommunizieren, damit Leser sich selbst ein vollständiges Bild ohne *Framing* machen können. Zu Illustrationszwecken zeigen wir den Unterschied von relativen und absoluten Zahlen in der folgenden Grafik:

Abbildung 13:  
Schweregrad der Glücksspielprobleme (DSM-5) nach Spielform



Basis: gewichtete Daten der 18- bis 70-Jährigen, die in den letzten 12 Monaten an der jeweiligen Glücksspielform teilnahmen

Aggregierte Glücksspielprobleme (DSM-5) nach Spielform in absoluten Zahlen  
Minimale Absolutzahlen sind durch Funktionen rundet



Die absolute Zahl der 18- bis 70-Jährigen, die in den letzten 12 Monaten an der jeweiligen Glücksspielform teilnahmen

Abbildung 8.1: Relative Zahlen.

Abbildung 8.2: Absolute Zahlen.

<sup>1</sup>Der Anteil der Infizierten, der verstorben ist.

## 9 | Evidenzbasierte Politik

### Executive Summary

*Die Forderung nach evidenzbasierter Politik wurde von den Vereinten Nationen bereits 1954 formuliert; als Grundlage müssen statistische Informationen höchster Qualität zur Verfügung stehen. In jüngster Vergangenheit wurde diese Forderung von der Leopoldina, der Nationalen Akademie der Wissenschaften, bekräftigt. Dafür bedarf es repräsentativer nationaler Expositionsdaten und evidenzbasierter Risikoschätzer als Basis eines wissenschaftlichen Nachweises, dass eine politische Maßnahme die Ursache eines Gesundheitsproblems wirksam beeinflussen kann. Das „Stimmungsbild Sucht“ des Beauftragten der Bundesregierung für Sucht- und Drogenfragen sowie die „Studie zur Evaluierung der Abwanderungsthese von terrestrischem Glücksspiel während des Covid-19 Lockdowns“ der Autoren des Glücksspiel-Surveys liefern keine derartigen Grundlagen und sind deshalb als Negativ-Beispiele anzusehen.*

*„The demand for statistics of the highest quality by Governments . . . has increased enormously. . . . To function efficiently, such Governments must have a basis of sound statistical information to assist them in formulationg their policies. . . . In all fields, both national and international, the work of the statistician is fundamental . . . for substituting realistic data for wishful thinking, biased claims and political oratory.“*

Bereits in den 1950er Jahren plädierten die Vereinten Nationen (United Nations 1954, S. 55) mit diesen Worten für evidenzbasierte Politik als Gegenentwurf zu reinem Wunschdenken und bloßer politischer Rhetorik und betonten dabei explizit die Relevanz qualitativ hochwertiger Datengrundlagen.

Evidenzbasierte Politik definiert sich durch die Nutzung von empirischen Daten und wissenschaftlichen Erkenntnissen zur Steuerung politischer Entscheidungsprozesse. Hierbei spielen Umfragen und

Studien eine Schlüsselrolle, da sie bei sorgfältiger Planung und Qualitätskontrolle aussagekräftige Daten liefern, die es ermöglichen, die Meinung der Bevölkerung oder bestehende Probleme zu erfassen. Umfragen dienen dabei oft als grundlegende empirische Instrumente. Sie stellen quantitative Daten bereit, die einen (repräsentativen) Einblick in die Haltung der Bevölkerung zu bestimmten politischen Themen bieten oder einen Überblick über die Prävalenz von verschiedenen Problemen in der Bevölkerung liefern können. So können Politiker und politische Analysten eine bessere Vorstellung davon bekommen, welche Themen die Bürger als dringlich erachten und welche politischen Maßnahmen sie bevorzugen. Studien, insbesondere wissenschaftliche Forschungsstudien, können weitere Beiträge zur Evidenz liefern. Sie können tiefergehende und detailliertere Informationen liefern als Umfragen, indem sie zum Beispiel langfristige Auswirkungen politischer Maßnahmen analysieren, kausale Zusammenhänge aufdecken und komplexe sozio-politische Dynamiken untersuchen. Sie können auch dazu dienen, Theorien zu entwickeln und zu testen, um das Verständnis der politischen Phänomene zu verbessern und fundierte Vorhersagen zu machen.

Die Nutzung von Umfragen und Studien als Grundlage evidenzbasierter Politik erfordert jedoch hohe fachliche, statistische und (forschungs-)ethische Kompetenz. Einerseits müssen diese Instrumente sorgfältig konzipiert und implementiert werden, um valide und zuverlässige Daten zu liefern. Dies erfordert ein tiefes Verständnis von Forschungsmethoden und Statistik. Andererseits müssen die Daten, die durch Umfragen und Studien gewonnen werden, kritisch interpretiert und in den fachspezifischen Kontext eingebettet werden. Das heißt, es ist notwendig, die Grenzen und möglichen Verzerrungen dieser Daten zu erkennen, sie in Beziehung zu anderen Arten von Evidenz und Wissen zu setzen und sie mit den sozio-politischen Realitäten und ethischen Überlegungen zu verbinden.

Umfragen und Studien können eine wichtige Grundlage für die Entwicklung, Umsetzung und Bewertung von politischen Maßnahmen bilden. Aber sie sind kein Selbstzweck. Sie sind Werkzeuge, die dazu dienen, fundierte, effektive und gerechte politische Entscheidungen zu treffen. Es ist daher unerlässlich, dass sie in einem breiteren Rahmen der Evidenzbasierung, Transparenz, Partizipation und Reflexion genutzt werden.

## 9.1 Grundprinzipien evidenzbasierter Politik

Die Nationale Akademie der Wissenschaften Leopoldina (2021) beschreibt evidenzbasierte Politik als „auf einer Überprüfung der Wirksamkeit eines bestimmten Programms oder einer Maßnahme“ aufbauend. Verfüge man noch nicht über entsprechende Daten über die Wirksamkeit einer Maßnahme, so

ließen sich diese mithilfe einer Evaluation gewinnen. Dabei, so die Leopoldina, sollten sowohl fortlaufend die festgelegten Ziele der gesetzlichen Maßnahmen und Programme evaluiert werden als auch die Implikationen und Auswirkungen zukünftiger Maßnahmen beurteilt werden. Falls festgestellt wird, dass die Ziele nicht in vorgesehenem Maße erreicht werden, ist eine Adaption und gegebenenfalls Beendigung der Maßnahmen geboten. Auf dieser Grundlage heißt es weiter: *„Um eine informierte Debatte zu gewährleisten, müssen Datengrundlage, Auswertungsmethode und Ergebnisse von Evaluationen zugänglich und nachvollziehbar sein“* (Nationale Akademie der Wissenschaften Leopoldina 2021). Ausführliche Erläuterungen zur Notwendigkeit der transparenten Forschung sowie Implikationen einer Nichtbeachtung dieses Grundsatzes findet sich in Kapitel 8.

Evidenzbasierte Politik erfordert darüber hinaus die systematische Analyse und Integration möglichst vieler empirischer Befunde aus qualitätsgesicherter, insbesondere peer-reviewter wissenschaftlicher Forschung, um eine bestimmte Fragestellung zu beantworten. Im Kontext der Evaluation des Glücksspielstaatsvertrags, wie beispielsweise bei der Beurteilung der Wirksamkeit von Maßnahmen zum Spielerschutz, bedeutet dies, dass es notwendig ist, erkenntnisoffen verschiedene Erklärungsansätze für Glücksspieltörungen zu berücksichtigen. So sollten u.a. Erkenntnisse zum Einfluss psychosozialer Faktoren (Manrique u. a. 2023; Bühringer u. a. 2023) in die Evaluation einbezogen werden, um die erwartete Wirksamkeit zukünftiger Maßnahmen zu maximieren.

Im Konzeptpapier der Nationale Akademie der Wissenschaften Leopoldina (2021) zur „Initiative zur evidenzbasierten Politikgestaltung“ wird allerdings bereits im ersten Absatz darauf hingewiesen, dass die von der Politik genutzten Informationsquellen *„in ihrer Fülle kaum überschaubar und in ihrer Qualität zum Teil sehr zweifelhaft sind“*. Weiter heißt es: *„In Deutschland wie auch in anderen Ländern können Formen der Evidenzbasierung noch besser als bisher in der Politikgestaltung verankert werden.“* Insbesondere der Nachweis von Ursache-Wirkungs-Beziehungen ist unerlässlich, wenn politische Maßnahmen durch Evidenz begründet werden sollen. Hierfür werden verlässliche Daten und angemessene Methoden benötigt. Unterschiedliche Methoden und Replikationsstudien maximieren die Wahrscheinlichkeit, robuste und damit relativ verlässliche Ergebnisse zu liefern. Dazu sollten neben Daten aus administrativen Quellen repräsentative Umfragen und bevölkerungsweite Längsschnittstudien für empirische Analysen verwendet werden, in Kombination mit in jüngerer Vergangenheit entwickelten, anspruchsvollen statistischen Methoden, die zur Aufdeckung kausaler Wirkungspfade entwickelt wurden. Insbesondere ist festzustellen, dass Daten, die zum Zweck der Zustandsbeschreibung generiert wurden, in aller Regel keine Aussagen über geeignete, d.h. wirksame Maßnahmen zur Änderung dieses Zustands ermöglichen.

Im Einklang mit den Ausführungen der Leopoldina zur erhöhten Evidenzbasierung in der politischen Gestaltung finden sich explizite Aussagen zu den Anforderungen, die die Eignung empirischer Informationsquellen als Entscheidungsgrundlage evidenzbasierter Politik determinieren, in der Public-Health-Forschung: Eine evidenzbasierte politische Intervention zur bevölkerungsbezogenen Prävention – nichts anderes ist Regulierung des Glücksspiels, wenn die öffentliche Gesundheit als Zielgröße adressiert wird – erfordert insbesondere „die Ermittlung des präventiven Potenzials auf der Basis repräsentativer nationaler Expositionsdaten und evidenzbasierter Risikoschätzer“ (Knorpp 2013, S. 262). Es muss also wissenschaftlich nachgewiesen werden, dass eine politische Maßnahme die Ursache eines Gesundheitsproblems wirksam beeinflussen kann. Daraus lässt sich ableiten, dass die Instrumente zur Generierung von Datengrundlagen evidenzbasierter Politik so konzipiert und umgesetzt werden müssen, dass sie folgende drei Anforderungen erfüllen: Sie ermöglichen (1) den Nachweis eines kausalen Zusammenhangs zwischen Exposition und Gesundheitsproblem, (2) eine valide Abschätzung des präventiven Potenzials der Intervention und (3) die fortlaufende Evaluation inklusive eines Kriteriums zur Beendigung der Intervention, wenn das erwartete Ausmaß der Prävention durch die Intervention nicht erreicht werden konnte. Die dritte Forderung gründet sich auf die ethischen Prinzipien für Gesundheitsförderung und Public Health, „Do not harm“ (Prinzip Nr. 2) und „Accountability“ (Prinzip Nr. 10) (Tannahill 2008, S. 386).

## 9.2 Negativ-Beispiel: Stimmungsbild Sucht

Der Beauftragte der Bundesregierung für Sucht- und Drogenfragen, Burkhardt Blienert, forderte zum Auftakt der Veranstaltungsreihe „Debatte (ge)SUCHT“ strengere Beschränkungen für Alkohol-, Tabak- und Glücksspielwerbung und beruft sich dabei auf die Bevölkerung: *„Die große Mehrheit der Bevölkerung will keine Alkoholwerbung mehr, sie fordert einen Stopp jeden Sponsorings für Tabakprodukte oder durch Sportwettenanbieter beim Fußball. Das darf die Politik doch nicht länger ignorieren.“* (Blienert 2023a). Dabei berief sich der Bundesdrogenbeauftragte auf das *Stimmungsbild Sucht- und Drogenpolitik: Alkohol, Tabak, Glücksspiel. Einstellungen der Bevölkerung zu Werbe- und Sponsoringregulierungen*. Die Ergebnisse der Befragung wurden in einer siebenseitigen Präsentation zusammengefasst, auf einer der Folien wird das Untersuchungsdesign dargestellt:

Die dargestellten Informationen lassen allerdings viele Fragen offen:

- Wie hoch ist der Anteil der telefonisch bzw. online Befragten?
- Welches Online-Panel wurde genutzt?



Abbildung 9.1: Untersuchungsdesign des Stimmungsbilds Sucht (Blienert 2023b)

- Wurde ein *Nonprobability* oder ein *Probability* Panel genutzt?
- Wie hoch war die *Unit-Nonresponse-Quote*?
- Wie war der *Item-Nonresponse* verteilt?
- Wie wurden die Mailadressen im Bereich Suchthilfe recherchiert?
- Wie wurde das *Oversampling* der Pädagogen erreicht?
- Auf Basis welcher Annahmen wurde die Gewichtung und die Berechnung des Fehlerintervalls berechnet?
- Anhand welcher externen Daten wurde die Gewichtung durchgeführt?

Da diese (und weitere) Fragen nicht beantwortet werden, kann die Qualität der Daten nur in Ansätzen bewertet werden – wie Münnich (2023) feststellte, ist dies bereits an sich ein Merkmal schlechter Datenqualität. Dennoch kann festgestellt werden, dass aufgrund des starken *Oversampling* von Pädagogen (24 % der Befragten) und „Fachkräften Suchthilfe“ starke Verzerrungen in den Daten existieren werden, weiter verstärkt durch die Nutzung eines Online-Panels. Auch die simple Gewichtung ist dem

Anschein nach völlig unzureichend, diese Verzerrungen zu korrigieren (Lohr 2021). Mithin ist von einer Inferenz auf die Bevölkerung klar abzuraten.

Das Stimmungsbild Sucht kann aufgrund mangelnder Transparenz nicht adäquat evaluiert werden, allerdings ist allein aufgrund der fehlenden Transparenz und offensichtlicher Verzerrungen in der Stichprobe von einer Nutzung der Befragungsergebnisse für Inferenzen auf die Bevölkerung abzusehen. Dies gilt natürlich insbesondere für staatliche Akteure (vgl. Kapitel 2), allerdings wurde das Stimmungsbild Sucht vom Beauftragten der Bundesregierung für Sucht- und Drogenfragen selbst in Auftrag gegeben. Insofern befremden die fehlende Transparenz und die offensichtlichen Verzerrungen umso mehr, denn Bürger sollten von der Bundesregierung transparente und korrekte Daten erwarten können. Wenn der Beauftragte der Bundesregierung für Sucht- und Drogenfragen also davon spricht, dass die „*die große Mehrheit der Bevölkerung*“ etwas will bzw. fordert, ist dies – zumindest auf Basis des Stimmungsbilds Sucht – nicht korrekt. Dies bedeutet nicht das Zutreffen des Gegenteils, denn es gilt „Absence of Evidence is not Evidence of Absence“. Das Stimmungsbild Sucht kann dies allerdings aufgrund methodischer Schwächen die Evidenz nicht nachweisen.

### 9.3 Negativ-Beispiel: Studie zu Lockdown und Glücksspiel

Im ihrer „Studie zur Evaluierung der Abwanderungsthese von terrestrischem Glücksspiel während des Covid-19 Lockdowns“ aus der „Zeitschrift für Wett- und Glücksspielrecht“ von 2023 untersuchen Schütze, Kalke und Buth (2023) die Spielgewohnheiten von terrestrischen Spielern im Kontext des ersten Covid-19-Lockdowns und den damit verbundenen (totalen) Einschränkungen des terrestrischen Glücksspiels. Der Artikel beinhaltet selektive Ergebnisse aus einer ausführlicheren Untersuchung der Autoren von 2021, welche von der BZgA gefördert wurde (Buth, Schütze und Kalke 2021).

Die Autoren wenden ausschließlich deskriptive Methoden an, um die Auswirkungen des Lockdowns zu untersuchen, weshalb es sich bestenfalls um eine erste explorative Analyse handeln kann, wie die Autoren selbst in ihrem ausführlichen Bericht schreiben (Buth, Schütze und Kalke 2021, S. 22). Auf die Berechnung von Konfidenzintervallen oder der Betrachtung im Zeitverlauf verzichten sie „*zugunsten einer besseren Lesbarkeit*“ (Buth, Schütze und Kalke 2021, S. 22). An dieser Angabe müssen die Ergebnisse gemessen werden. Kausale Zusammenhänge lassen sich aus der Studie nicht ableiten (siehe auch Abschnitt 6.2 Kausalität), dies wird im Forschungsartikel allerdings nicht erwähnt. Stattdessen schlussfolgern die Autoren eine Notwendigkeit legislativer Maßnahmen aus ihrer lediglich explorativen Studie. Sie stellen die Behauptung auf, die „*im Glücksspielstaatsvertrag enthaltene Prämisse eines*

*natürlichen Glücksspielbedürfnisses (§1 Nr.2 GlüStV), das durch entsprechende gesellschaftliche Angebote befriedigt werden müsse (§ 10 Abs.1 GlüStV), wird durch [ihre] Ergebnisse in Frage gestellt“* (Schütze, Kalke und Buth 2023, S. 28). Die Autoren lassen bei dieser Betrachtung außer Acht, dass sie nicht die Gesamtheit an Glücksspielern untersucht haben, sondern sich von vornherein ausschließlich auf terrestrische Spieler fokussieren – über ein natürliches Glücksspielbedürfnis lässt sich daher, unabhängig von der fragwürdigen Datenqualität, wenig sagen.

Ferner ist zu kritisieren, dass die Autoren im Artikel keine Angabe über im Lockdown hinzugekommene, neue Spieler machen. Unabhängig von einer Evaluation der sog. Abwanderungsthese, wonach Beschränkungen bei terrestrischem Glücksspiel zu einem Ausweichen der Glücksspieler auf Online-Angebote führen, lässt sich somit eine Netto-Abwanderung vom terrestrischen Glücksspiel nicht beurteilen. Daher könnte es, selbst bei einer ausbleibenden Abwanderung zum Online-Glücksspiel während der Beschränkungen, durch neu hinzukommende Spieler eine höhere Gesamtzahl an Glücksspielern geben. Im ausführlichen Bericht wird von 1.073 Spielern gesprochen, die ihr Glücksspiel im Lockdown (wieder) begannen. Diese Zahl stellt einen Anteil von 37,2 % an allen Glücksspielern vor dem Lockdown in der Studie dar. Innerhalb von ungefähr vier Monaten (März bis Juni 2020) hat die Anzahl an Glücksspielern also einen Zuwachs von mehr als einem Drittel zu verzeichnen. Um grundsätzlich einschätzen zu können, ob die Abbruchrate während Covid-19-Restriktionen signifikant unterschiedlich war, fehlt zudem ein Referenzwert zu Abbrechern aus einer Zeit ohne pandemiebedingte Lockdowns. Dies betonen auch die Autoren in ihrem Bericht, allerdings nicht im Artikel.

Darüber hinaus ist die Datenqualität ganz grundsätzlich fragwürdig. Aus Kostengründen wurde in der Studie auf ein Online-Access-Panel (Payback) zurückgegriffen. Dass es sich dabei um eine Gelegenheitsstichprobe (*Convenience-Sample*) handelt, merken die Autoren im Bericht selbst an. Die Kritik an einem solchen *Nonprobability-Sample* wurde in vorigen Kapiteln dargelegt, insbesondere bezüglich des (*Self-*)*Selection-Errors*, der in Online-Panels zustande kommt, da mehrheitlich eben nicht zufällig ausgewählte, sondern freiwillig registrierte Personen an Befragungen in Online-Panels teilnehmen. Die Autoren schreiben selbst: „*Es kann nicht eingeschätzt werden, wie repräsentativ die erhobenen Daten für die Glücksspielenden in Deutschland sind*“ (Schütze, Kalke und Buth 2023, S. 53). Die Autoren rechtfertigen ihre Wahl eines Convenience Samples durch einen angeblichen „*Trend der sich wandelnden empirischen Sozialforschung*“. Doch auch ein angeblicher „Trend“ in der Sozialforschung führt nicht zu einer repräsentativen Stichprobe.

## 10 | Rezeption von Survey und Begleitmaterial

### Executive Summary

*Der Glücksspiel-Survey 2021 weist zwei eklatante Schwächen auf, die eine sachgerechte Rezeption für Laien (Nicht-Statistiker/-Wissenschaftler) nahezu unmöglich machen: Erstens werden im Survey selbst sowie in der begleitenden Kommunikation (Pressemitteilungen und Aufsätze) unzulässige und undifferenzierte Aussagen getroffen. So werden etwa unzulässigerweise aus Korrelationen Kausalitäten hergeleitet. Zweitens haben sich die Autoren gegen eine Nennung und Diskussion der Limitationen des Surveys und der gewählten Methoden entschieden, was Lesern die Einordnung überhaupt erst ermöglichen würde. Die Darlegung der Einschränkungen und Implikationen eigener Forschungsarbeit insbesondere bei einer an die Öffentlichkeit gerichteten Publikation ist essenziell. Die Verantwortung für die Kommunikation von Studienergebnissen liegt explizit bei den jeweiligen Studienautoren. Zwar sollten Rezipienten wie Medien und Politik Forschungsergebnisse stets kritisch hinterfragen; angesichts ihrer typischerweise weniger ausgeprägten Datenkompetenz wiegt die Verantwortung der Forscher jedoch umso schwerer, da sich Rezipienten aus Politik, Medien und Gesellschaft im Allgemeinen auf die Fundiertheit der Aussagen von Wissenschaftlern verlassen. Die schwerwiegenden Folgen intransparenter, selektiver und unvollständiger Kommunikation seitens der Survey-Autoren in Bezug auf die eigenen Studienergebnisse zeigen sich in zahlreichen Rezeptionen, in denen Falschaussagen der Autoren wiederholt und Ergebnisse des Surveys daher aus wissenschaftlicher Sicht inkorrekt wiedergegeben oder fehlinterpretiert sowie unzulässigerweise auf einzelne Bundesländer projiziert werden. Solche Rezeptionen finden sich in den Medien, in Äußerungen mehrerer exekutiver Akteure und in Gerichtsurteilen.*

Im Gutachten wurden bisher etliche kritikwürdige Punkte und Limitationen des Glücksspiel-Surveys 2021 genannt. Abgesehen von den festgestellten fragwürdigen, da empirisch nicht belegten oder zu undifferenzierten Schlussfolgerungen, sind Limitationen in der empirischen Sozialwissenschaft normal. Kein Befragungsinstrument, kein Forschungsdesign und keine Stichprobe sind perfekt, alle Methoden haben Vor- und Nachteile. Genau deswegen ist es allerdings auch wissenschaftlicher Konsens, diese Vor- und Nachteile (bzw. Limitationen) in Forschungsarbeiten klar zu benennen und zu diskutieren sowie die Ergebnisse entsprechend einzuordnen (Babbie 2015).

Der Glücksspiel-Survey 2021 weist zwei eklatante Schwächen auf, die eine sachgerechte Rezeption für Laien (bzw. Nicht-Statistiker/-Wissenschaftler) nahezu unmöglich machen: Erstens werden im Survey selbst unzulässige und undifferenzierte Aussagen getroffen. Zweitens haben sich die Autoren gegen eine Nennung und Diskussion der Limitationen entschieden, die den Lesern die Einordnung überhaupt erst ermöglichen würden. Dennoch wurde und wird der Glücksspiel-Survey 2021 in der Öffentlichkeit und Politik (breit) rezipiert. Aufgrund der beschriebenen Schwächen sollten fehlinterpretierende Rezeptionen nicht überraschen, diese werden nachfolgend genauer diskutiert.

## 10.1 Pressemeldung der Studienautoren

Zunächst wird die Pressemeldung der Studienautoren zur Veröffentlichung des Glücksspiel-Surveys 2021 (Institut für interdisziplinäre Sucht- und Drogenforschung 2022) analysiert, da mögliche Missverständnisse oder inkorrekte Aussagen innerhalb dieser sich in weiteren Rezeptionen verstärken oder zur Verbreitung falscher Informationen führen können. Positiv ist in der Pressemeldung zunächst der Hinweis zu bewerten, dass der Survey aufgrund veränderter Methodik nur eingeschränkt mit früheren Erhebungen vergleichbar ist.

Ähnliche Hinweise wären auch bezüglich anderer Thematiken wünschenswert gewesen: Hinsichtlich der Prävalenz von Glücksspielstörungen in Deutschland werden keine Unsicherheiten, beispielweise in Form von Konfidenzintervallen, thematisiert. Insbesondere die Tatsache, dass ein erheblicher Anteil der dargestellten Unterschiede durch zufällige Schwankungen erklärt werden kann, bleibt unkommentiert. Dadurch wird es dem Leser der Pressemeldung erschwert, Qualität und Aussagekraft der Ergebnisse zu bewerten – wie ausführlich in Abschnitt 4.1 ausgeführt. Es bleibt offen, warum die Studienautoren zumindest vereinzelt Konfidenzintervalle bei der Auswertung ermitteln (ungeachtet der Tatsache, dass diese Konfidenzintervalle verzerrt sein dürften), diese aber nicht in Einleitung, Fazit und Pressemeldung zum Survey kommunizieren. Aufgrund dieser missverständlichen Angaben in der Pressemeldung

verwundert es nicht, dass derselbe Fehler von verschiedenen Akteuren repliziert wurde, beispielsweise vom Beauftragten der Bundesregierung für Sucht- und Drogenfragen in der BILD (2022).

Ebenso erklärt einer der Studienautoren fälschlicherweise in der Pressemeldung, *„dass das Risiko, glücksspielbedingte Probleme zu entwickeln, sich sowohl hinsichtlich der Glücksspielformen wie auch der Zugangswege [...] unterscheidet“* (Institut für interdisziplinäre Sucht- und Drogenforschung 2022). Zur Untersuchung des Risikos einer Entwicklung glücksspielbezogener Probleme nach Spielform und Zugangsweg ist der Survey, wie unter Abschnitt 6.2 erklärt, qua Design gar nicht geeignet. Denn die Frage nach der Richtung des Effekts (umgekehrte Kausalität) kann nicht mit Querschnittsdaten untersucht bzw. beantwortet werden: Es könnte der Fall sein, dass Poker ein erhöhtes Risiko für eine Glücksspielstörung darstellt. Alternativ könnte es jedoch auch sein, dass Personen, die bereits eine Glücksspielstörung aufweisen, verstärkt dazu tendieren, Poker zu spielen. Dass sich eine solch fundamentale Fehlinterpretation der Ergebnisse durch einen der verantwortlichen Forscher in der Pressemeldung findet, ist sehr bedenklich, erklärt allerdings auch die weitergehenden Fehlrezeptionen des Surveys.

Ansonsten findet die statistische Unsicherheit weiter – wie bereits im Survey selbst – keine Beachtung in der Pressemeldung, wodurch der Anteil an Personen mit Glücksspielstörung nach Spielform fälschlicherweise nach Anteilsgröße geordnet wiedergegeben wird: *„Der jeweilige Anteil von Personen mit einer glücksspielbezogenen Störung ist unter den Spieler\*innen einzelner Spielformen unterschiedlich ausgeprägt. Die höchsten Anteilswerte finden sich unter den Spieler\*innen an Geldspielautomaten in Spielhallen und der Gastronomie mit zusammen 33,4%, gefolgt von Spieler\*innen an Glücksspielautomaten in Spielbanken (31,5%) und den Teilnehmer\*innen an Live-Sportwetten (29,7%).“* (Institut für interdisziplinäre Sucht- und Drogenforschung 2022). Aus statischer Sicht liegt für diese Aussage aufgrund der teils überlappenden Konfidenzintervalle und der nicht-signifikanten Differenzen keine empirische Evidenz vor.

Zudem lässt die Pressemeldung eine wenigstens ansatzweise Nennung der Limitationen, wie sie die Studienautoren beispielsweise im Forschungsartikel mit Glücksspiel-Survey-Daten (Meyer, Kalke und Buth 2023) vorgenommen haben, vermissen. Sowohl zur Verzerrung der Stichprobe als auch zum Anwendungsbereich des DSM-IV/5 Instruments, dessen Eignung im epidemiologischen Kontext zumindest diskutabel ist, verlieren die Studienautoren in der Pressemeldung kein Wort. Dabei ist genau das ein elementarer Aspekt wissenschaftlicher Verantwortung und Transparenz, wie oben bereits ausgeführt. Denn nur dies stärkt das Vertrauen in die Wissenschaft und ermöglicht erst eine fundierte öffentliche Diskussion – nicht jeder Leser der Studie hat die Ressourcen und Fähigkeiten (noch ist es

seine Aufgabe), sich intensiv mit dem Survey auseinanderzusetzen und beispielsweise selbst Signifikanztests durchzuführen.

Insgesamt werden in der Pressemeldung der Studienautoren teils bedenkliche Falschaussagen und aus empirisch-statistischer Sicht nicht haltbare Thesen aufgestellt. Des Weiteren lassen die Studienautoren eine kritische und transparente Einordnung der Ergebnisse unter Berücksichtigung der Limitationen vermissen. Falsche, ungenaue oder missverständliche Rezeptionen des Surveys durch verschiedene Akteure aus Exekutive, Rechtsprechung und Medien sind daher wenig überraschend und liegen im primären Verantwortungsbereich der Autoren des Glücksspiel-Surveys.

## 10.2 Deutscher Lotto- und Totoblock

Auch der Deutsche Lotto- und Totoblock (DLTB) als Förderer des Surveys hat eine Stellungnahme zur Veröffentlichung des Glücksspiel-Surveys 2021 veröffentlicht (Deutscher Lotto- und Totoblock 2022), die von einer gemeinsamen Pressekonferenz der Studienautoren mit den Geschäftsführern der Landeslotteriegesellschaften von Hamburg und Niedersachsen begleitet wurde. In der Stellungnahme sind als Konsequenz der Defizite bei der Kommunikation der Studienergebnisse durch die Studienautoren gleich mehrere inkorrekte bzw. nicht empirisch-statistisch belegte Aussagen zu finden: *„Ebenso bestätigen die Studienergebnisse erneut, dass von Lotterien, wie Lotto 6aus49 und Eurojackpot, ein deutlich geringeres Gefährdungspotenzial ausgeht als von den weiteren, erhobenen Glücksspielformen“*, so die Geschäftsführung der Toto-Lotto Niedersachsen GmbH (Deutscher Lotto- und Totoblock 2022). Dabei ist der Glücksspiel-Survey qua Design ungeeignet, das Gefährdungspotenzial verschiedener Spielformen zu erfassen, siehe Abschnitt 6.1 und 6.2. Vermutlich bezieht sich die Aussage auf die Prävalenz der Glücksspielstörung nach Spielform, allerdings kann daraus nicht geschlossen werden, welcher Faktor den anderen beeinflusst: Führt Lotto seltener zu Glücksspielstörungen? Oder ist Lotto der Einstieg ins risikohafte Glücksspiel? Unabhängig davon, wie plausibel beide Argumente sind, so muss in der Wissenschaft Evidenz für eine Hypothese gefunden werden, bevor diese als „bestätigt“ bezeichnet werden kann. Dies ist zumindest mit dem Glücksspiel-Survey 2021 nicht möglich, da Daten nur zu einem Zeitpunkt erhoben werden – insofern ist die getätigte Aussage nicht korrekt, der vorliegende Survey kann die genannte These gar nicht bestätigen.

Gleiches gilt für die folgende Aussage: *„Darüber hinaus wird durch die Studienergebnisse deutlich, dass es zwar unter den Spielteilnehmern von Glücksspielen mit geringem Gefährdungspotential (Lotterien) einen geringen Anteil von Problem Spielern gibt, bei denen aber – so die Studie – andere Glücks-*

*spielformen die Probleme verursacht haben dürften*“ (Deutscher Lotto- und Totoblock 2022). Der Glücksspiel-Survey liefert keinerlei Evidenz (und kann dies bedingt durch sein Design auch nicht), die diese Aussage stützt. Ebenso präsentiert der Survey keine Daten, die erläutern könnten, warum Lotterie-Spieler eine Glücksspielstörung entwickelt haben, welche spezifischen Glücksspiele prinzipiell zu einer Glücksspielstörung führen könnten, oder ob eine Kombination verschiedener Arten von Spielen zur Ursache für auffälliges Verhalten werden könnte. Dies wird ausführlich in den Abschnitten 6.1, 6.2 ausgeführt.

Auch die folgende statistisch nicht belegte Aussage in der DLTB-Stellungnahme ist die Folge einer zuvor verbreiteten Fehlinterpretation seitens der Survey-Autoren selbst: *„Im Hinblick auf den am 1. Juli 2021 in Kraft getretenen neuen Glücksspielstaatsvertrag (GlüStV 2021) ist die Studie eine wichtige Nullmessung. Die 2021 durch geführte Befragung ist ein solider Ausgangspunkt für die Erfassung von Auswirkungen der neuen gesetzlichen Regelungen, insbesondere die Legalisierung vieler Online-Glücksspielangebote. Inwieweit sich dadurch das Glücksspielverhalten in der Bevölkerung verändert, werden die folgenden Erhebungen für 2023 und 2025 zeigen.“*, so die Geschäftsführung der Lotto Hamburg GmbH (Deutscher Lotto- und Totoblock 2022). Wie im Abschnitt 7.3 erläutert, kann der Glücksspiel-Survey 2021 aufgrund seines Designs keine Nullmessung darstellen, denn der abgefragte Zeitraum beinhaltet sowohl Zeiträume vor als auch nach dem Inkrafttreten des neuen Glücksspielstaatsvertrags (einschließlich des vorangegangenen Umlaufbeschlusses). Zudem können die Konsequenzen der Legalisierung der Online-Glücksspiele nicht erfasst werden, da die Legalität des gespielten Glücksspiels gar nicht erfasst wird (für mehr Informationen, siehe Abschnitt 7.4). Zudem ermöglicht das gewählte Design der Querschnittsbefragungen reine Trendanalysen, für die kausale Zuordnung von Veränderungen zum neuen Glücksspielstaatsvertrag ist der Survey – entgegen der zitierten Aussage – ungeeignet (siehe Abschnitt 6.1).

Die primäre Ursache dieser Weiterverbreitung von nicht korrekten, unzureichend differenzierten und unvollständigen Interpretationen des Glücksspiel-Surveys ist das ursprüngliche Fehlverhalten der Survey-Autoren selbst, die im Survey und in der dazugehörigen Pressemitteilung der Studienautoren selbst ähnliche Aussagen getroffen und die Limitationen des Surveys nicht angemessen diskutiert haben. Rezipienten fachspezifischer Forschungsstudien haben nicht die erforderliche Qualifikation, Studienergebnisse in der gebotenen Tiefe kritisch würdigen und wissenschaftlich bzw. statistisch einordnen zu können. Stattdessen verlassen sich Studienrezipienten nachvollziehbarerweise regelmäßig auf die Transparenz und wissenschaftliche Sorgfalt der verantwortlichen Studienautoren, an der es im vorliegenden Fall mangelt.

### 10.3 Weitere Referenzierung des Surveys durch die Autoren

Die Autoren des Glücksspiel-Surveys haben die Daten des Surveys auch für weitere Forschungsartikel verwendet, unter anderem die Studie „Kenntnis und Akzeptanz von Spielerschutzmaßnahmen bei den Glücksspielenden – Ergebnisse des Glücksspiel-Surveys 2021“ (Buth, Meyer und Kalke 2022). Dieser Artikel hat das Ziel, zu untersuchen, „welche Zusammenhänge zwischen der Intensität glücksspielbezogener Probleme (gemäß DSM-5-Kriterien) und der Kenntnis und Akzeptanz von Spielerschutzmaßnahmen bestehen.“ (Buth, Meyer und Kalke 2023, S. 1). Zur Beantwortung dieser Frage wurde die Bekanntheit und Akzeptanz elf verschiedener Maßnahmen, beispielsweise in Bezug auf das Spielverbot für Minderjährige, abgefragt.

Grundsätzlich ist es möglich, mit Hilfe der Daten aus dem Glücksspiel-Survey 2021 auch Assoziationen zwischen dem Schweregrad von Glücksspielproblemen und der Kenntnis und Akzeptanz von Spielerschutzmaßnahmen zu untersuchen. Er ist allerdings nicht möglich, diesbezüglich kausale Aussagen zu treffen (siehe Abschnitt 6.2). Der Artikel kann also maximal eine explorative Studie sein, die korrelative Zusammenhänge untersucht. Diese Limitation wird von den Autoren nicht klar benannt – dies wäre allerdings gemäß guter wissenschaftlicher Praxis notwendig (Deutsche Forschungsgemeinschaft 2019).

Im Gegensatz zum Glücksspiel-Survey 2021 schätzen die Autoren diesmal statistische Modelle<sup>1</sup>, um den Effekt dritter (möglicherweise) konfundierender Variablen zu erfassen und auszuschließen. Doch auch wenn dies ein Fortschritt zum Glücksspiel-Survey ist, ist die Umsetzung unzureichend und zu simpel. Die Autoren inkludieren nur Geschlecht und Alter in ihre Modelle – obwohl der Survey weitere Variablen (Bildung, Migrationshintergrund, etc.) erhebt, die potenziell einen Effekt auf die Kenntnis und Akzeptanz von Spielschutzmaßnahmen haben können. So bleibt die Möglichkeit bestehen, dass eine Glücksspielstörung gar keinen signifikanten oder zumindest einen kleineren Einfluss auf die untersuchten Variablen hat. Weiterhin präsentieren die Autoren ihre Ergebnisse in einer fragwürdigen Art und Weise: Es werden Tabellen präsentiert, bei denen die im Vergleich zur Referenzkategorie (unproblematische Spieler) signifikant unterschiedliche Kategorie der Spielstörung mit einem oder mehreren Sternchen (Signifikanzniveaus) markiert ist. Bei diesen Kategorien werden dann auch die *Odds Ratio* angegeben, die zur Interpretation des gewählten Modells essenziell sind. Die *Odds Ratio* werden allerdings nur bei signifikanten Unterschieden angegeben, was aus mehreren Gründen problematisch ist:

---

<sup>1</sup>Multinomiale logistische Regression

- **Verzerrung der Ergebnisse:** Das selektive Zeigen von nur signifikanten Ergebnissen kann ein unvollständiges oder irreführendes Bild der tatsächlichen Beziehungen zwischen Variablen geben.
- **Fehlende Informationen über die Größe des Effekts:** Selbst wenn eine Variable statistisch signifikant ist, kann die Größe des Effekts (gemessen an der *Odds Ratio*) immer noch praktisch irrelevant sein.
- **Mangelnde Transparenz:** Ein wichtiger Aspekt wissenschaftlicher Forschung ist Transparenz (siehe Kapitel 8). Wenn Forscher nur signifikante Ergebnisse präsentieren und nicht-signifikante Ergebnisse auslassen, werden dem Leser relevante Informationen vorenthalten.

Zudem werden die *Odds Ratio* von den Autoren zwar in den Tabellen inkludiert, im Artikel aber nie erwähnt oder interpretiert – für den statistisch wenig versierten Leser ist es daher nicht möglich, die Größe des Effekts zu verstehen. Ebenso können selbst statistisch versierte Leser die Effekte ohne vollständige Informationen nicht vollumfänglich interpretieren. Dieses inkohärente Vorgehen der Studienautoren lässt vermuten, dass die Modelle nur aufgrund der Signifikanztests gerechnet wurden. Signifikanztests sollten allerdings immer im Zusammenhang mit den jeweiligen Effektgrößen, hier den *Odds Ratio*, interpretiert werden (Cohen 1988; Cumming u. a. 2012). Das genutzte Modell wird also nicht nur mit zu wenigen Variablen gerechnet, sondern auch gar nicht interpretiert.

Da die Autoren erneut die verwendeten Daten und Materialien nicht veröffentlicht haben, können die Ergebnisse erneut nicht überprüft und repliziert werden – wie in Kapitel 8 ausgeführt, ein klarer Verstoß gegen gute wissenschaftliche Praxis (Deutsche Forschungsgemeinschaft 2019).

Weiterhin überrascht die simple Analyse im vorliegenden Artikel. Bei der Messung der Akzeptanz der Maßnahmen wurden die folgenden Antwortmöglichkeiten vorgegeben:

- finde ich gut
- bin unentschieden
- lehne ich ab
- weiß nicht

Zum einen wäre eine gebräuchlichere und feinere fünfstufige Likert-Skala vorzugswürdig gewesen wäre (vgl. Krosnick und Presser (2010)), da die genutzte dreistufige Skala sehr simpel ist. Zum anderen

ist insbesondere die reine Fokussierung auf die Kategorie „finde ich gut“ problematisch. Denn die Ausblendung der anderen Kategorien liefert ein unvollständiges Bild, da unklar ist, inwieweit die fehlende Zustimmung verteilt ist: Sind die Befragten unentschieden? Lehnen Sie die Maßnahme ab? Oder wissen sie keine Antwort? Jede Antwort hat unterschiedliche Implikationen, diese Informationen werden im vorliegenden Artikel weder kommuniziert noch diskutiert. Die Autoren setzen eine niedrigere Zustimmungsrates mit einer schlechteren Bewertung oder einer geringeren Akzeptanz gleich – das ist falsch, eine niedrige Zustimmungsrates könnte auch das Ergebnis von mehr „weiß nicht“-Antworten sein, da den Befragten beispielsweise Informationen über die Maßnahmen gefehlt haben.

Dies könnte insbesondere der Fall sein, da *„elf verschiedene Maßnahmen abgefragt [wurden], die in Deutschland nach dem neuen Glücksspielstaatsvertrag umzusetzen sind“* (Buth, Meyer und Kalke 2023, S. 4). Die Befragung wurde von August bis Oktober 2021 durchgeführt, der neue Glücksspielstaatsvertrag trat am 1. Juli 2021 in Kraft. Die Fragen bezüglich Glücksspielproblemen wurden allerdings Personen gestellt, *„die angaben, sich in den zurückliegenden 12 Monaten vor der Befragung an mindestens einem Glücksspiel beteiligt zu haben.“* (Buth, Meyer und Kalke 2023, S. 3). Das bedeutet konkret, dass auch Personen, die zum Beispiel im Dezember 2021 das letzte Mal Glücksspiel gespielt haben, bezüglich der Glücksspiel-Maßnahmen, die im Juli 2021 eingeführt wurden, befragt wurden. Dies erkennen die Autoren nur teilweise selbst: *„Bei der Betrachtung und Bewertung der Befunde dieses Beitrags ist zu beachten, dass einige der vorgegebenen Spielerschutzmaßnahmen erst im Zuge der Diskussion um den neuen Glücksspielstaatsvertrag, welcher im Sommer 2021 in Kraft trat, einer breiteren Öffentlichkeit bekannt wurden. Somit ist nicht auszuschließen, dass die Kenntnis dieser Maßnahmen für einen Teil der Befragten nur unvollständig und die inhaltliche Auseinandersetzung mit ihnen gering war.“* (Buth, Meyer und Kalke 2023, S. 8). Die Nennung dieser Limitation ist zwar grundsätzlich positiv zu werten, allerdings fehlt die praktische Umsetzung im Sinne einer Analyse der „weiß-nicht“-Antworten. So ist es für den Leser unmöglich einzuschätzen, ob die Variation in den Zustimmungsrates eine wirkliche Veränderung in den Einstellungen der Befragten bedeutet oder nur Ergebnis höherer „weiß-nicht“-Antworten ist. Dies ist die Konsequenz der mangelnden Transparenz und nicht veröffentlichten Daten.

Da die Daten aus dem Glücksspiel-Survey 2021 stammen, sind die Ergebnisse, wie in Kapitel 4 und 5 ausgeführt, vermutlich stark verzerrt und somit nicht repräsentativ.

Zusammengefasst sollte der vorliegende Artikel nicht für Entscheidungen über regulatorische Entscheidungen herangezogen werden, da in den Daten des Glücksspiel-Surveys starke Verzerrungen vorliegen. Zudem sind die genutzten Modelle zu simpel und werden substantiell nicht interpretiert, was das

Veständnis der Ergebnisse für statistische Laien stark erschwert. Die Daten sind zudem nicht für kausale Aussagen geeignet, die zwar im vorliegenden Artikel nicht getroffen werden, ein entsprechender Hinweis wäre allerdings angebracht gewesen. Weiterhin ist die Fokussierung auf die Zustimmungsrates nicht überzeugend, da unklar ist, inwieweit eine Veränderung in der Zustimmungsrates eine wirkliche Veränderung oder Ergebnis von mehr „weiß-nicht“-Antworten ist. Eine Analyse der „weiß nicht“-Antworten wäre jedoch insbesondere aufgrund des kurzen Zeitraums zwischen der Befragung und den neuen Maßnahmen sowie der Möglichkeit, dass Befragte nur vor den neuen Maßnahmen Glücksspiel gespielt haben, notwendig gewesen.

## 10.4 Jahrbuch Sucht 2023

Das Jahrbuch Sucht von 2023 der Deutschen Hauptstelle für Suchtfragen e.V. beinhaltet ein Kapitel zu Glücksspiel (Meyer 2023), geschrieben von einem der Autoren des Glücksspiel-Surveys 2021.

Grundsätzlich erfolgt im Jahrbuch Sucht eine unkritische Wiedergabe einzelner zentraler Ergebnisse aus dem Glücksspiel-Survey 2021. Es wird dabei nicht auf die durch eine verzerrte Stichprobe bedingte fehlende Repräsentativität der Daten sowie die große Unsicherheit einiger angegebener Werte hingewiesen. Lediglich der notwendige Hinweis auf eingeschränkte Vergleichbarkeit mit den BZgA-Surveys aus den Jahren zuvor wird gegeben. Im Jahrbuch Sucht wird eine Tabelle über die Schweregrade der Glücksspielprobleme in Prozent nach ausgewählten Spielformen gezeigt, deren Daten laut Autor aus dem Glücksspiel-Survey 2021 entnommen wurden (Meyer 2023, S. 103). In dieser aggregierten Form und aufgeteilt in die ausgewählten acht Kategorien wurden die Daten in dem Survey nicht präsentiert. Die Anteile an Glücksspielstörungen für Roulette beispielsweise stammen mutmaßlich aus nicht-veröffentlichten Daten und können nicht nachvollzogen werden. Der Autor weist darauf hin, dass die Kategorien „aufgrund der Differenzierung nur auf geringen Fallzahlen“ (Meyer 2023, S. 103) beruhen, ohne allerdings die Fallzahlen konkret zu nennen sowie Implikationen dieser Tatsache aufzuzählen – schließlich sind diese Ergebnisse mutmaßlich nicht repräsentativ und unterliegen großer Unsicherheit. Die Angabe einer – über die acht gezeigten Glücksspielformen hinausgehenden – Liste wäre zudem an dieser Stelle problemlos möglich gewesen, um ein umfassenderes Bild aller Glücksspielformen abzubilden. Durch den Beitrag im Jahrbuch Sucht 2023 trägt einer der Survey-Autoren abermals zur unkritischen Verbreitung der Ergebnisse des Glücksspiel-Surveys ohne Nennung und Diskussion ihrer Limitationen bei – obwohl ihm als Wissenschaftler die ethische Fragwürdigkeit eines solchen Verhaltens bewusst sein müsste.

## 10.5 Exekutive

### 10.5.1 Senator für Inneres der Freien Hansestadt Bremen

Ein besonders prominentes und schwerwiegendes Negativbeispiel für die unkritische und inkorrekte Rezeption des Surveys liefert der Senator für Inneres der Freien Hansestadt Bremen, Ulrich Mäurer, der sich mit den Worten zitieren ließ, „nach neuesten Studien [sind] 1,3 Millionen Menschen in Deutschland glücksspielsüchtig. Weitere 3,25 Millionen Menschen stehen auf der Kippe und zeigen ein riskantes Verhalten.“ (Senator für Inneres der Freien Hansestadt Bremen 2023). Diese Zahlen wurden auf Basis des Glücksspiel-Surveys 2021 sowie den Bevölkerungszahlen für Deutschland berechnet (Statistisches Bundesamt (Destatis) 2023). Neben dem Verdacht auf Geldwäsche führte der Innensenator diese Zahlen als Rechtfertigung der geplanten Schließung von Wettbüros in Bremen im Sommer 2022 an. Wie allerdings ausführlich im Kapitel 4 erläutert, sind die Anteilswerte der Stichprobe aufgrund ihrer Verzerrung nur begrenzt auf die deutsche Bevölkerung übertragbar. Zudem suggerieren die Zahlen eine Genauigkeit, die faktisch in dieser Form nicht existiert – angebrachter wäre (mindestens!) die Angabe der Konfidenzintervalle, die ebendiese Unsicherheit kommuniziert – doch selbst dann würde der vermutlich große *Nonsampling-Error* nicht berücksichtigt. Zudem ist die Nutzung der DSM-5-Kriterien in nicht-klinischem Kontext fragwürdig. So schreiben die Studienautoren in einem auf dem Glücksspiel-Survey basierenden Forschungspapier, das DSM-5 Instrument sei „*less suitable for epidemiological studies as they were originally developed for the clinical context*“ (Meyer, Kalke und Buth 2023). Die Autoren sind sich dieser Limitation also bewusst, erwähnen sie auch im Forschungspapier, aber nicht im Glücksspielsurvey - dabei wäre diese Einschränkung gerade im nicht-wissenschaftlichen Kontext umso wichtiger, um Fehlrezeptionen und -interpretationen aus dem politischen Raum vorzubeugen. Zudem ist, wie unter Abschnitt 7.1 beschrieben, auch die Kategorie der riskanten Spieler fragwürdig, insbesondere in der Form, wie sie vom Bremischen Innensenator interpretiert wird („stehen auf der Kippe“). Solchen ungenauen und falschen Interpretationen hätten die Autoren vorbeugen können, wenn sie die Limitationen des Glücksspiel-Surveys offen und transparent kommuniziert hätten.

### 10.5.2 1. Glücksspielsuchtbericht der Thüringer Landesregierung

Eine ebenso fragwürdige Nutzung der Anteilswerte des Glücksspiel-Surveys 2021 zur Hochrechnung auf absolute Zahlen fand im 1. Glücksspielsuchtbericht der Thüringer Landesregierung (Thüringer Ministerium für Arbeit, Soziales, Gesundheit, Frauen und Familie 2023) statt. Im Bericht wird folgendes

festgestellt: „In Thüringen gibt es ca. 32.400 Personen, die Probleme mit Glücksspielen entwickelt haben. Diese Zahl ist eine Hochrechnung auf der Grundlage der Prävalenzzahl der aktuellen Erhebung von Buth et al. (2022)“ (S. 5). Auch hier ist die trotz der verzerrten Stichprobe und mangelnden Berücksichtigung der Unsicherheit durchgeführte Hochrechnung zu kritisieren, insbesondere da die Unsicherheit aufgrund des Fokus auf Thüringen stark ansteigt. Generell ist unklar, wie genau die Zahl zustande kommt: Wurde die deutschlandweite Prävalenz von 2,3 % oder die (in dem Survey nicht veröffentlichte) Prävalenz für Thüringen genutzt? Welche Bevölkerungszahlen aus welchem Jahr werden als Grundgesamtheit genutzt? Wurden alle Altersklassen miteinbezogen oder nur die 18- bis 70-Jährigen, die im Glücksspiel-Survey befragt wurden? Diese Fragen müssen sich Leser selbst durch eigene Berechnungen beantworten.

Um die im Glücksspiel-Survey 2021 genannte Prävalenz zur Hochrechnung für Thüringen zu verwenden, sollte nur die Thüringer Stichprobe aus dem Survey hochgerechnet werden, um weitere Verzerrungen zu vermeiden – schließlich unterscheidet sich die demografische Struktur der Thüringer Bevölkerung von derjenigen der gesamtdeutschen Bevölkerung. Da die Studienautoren ihre Ergebnisse nicht nach Bundesländern disaggregieren und auch ihre Daten nicht veröffentlicht haben, wurde im Thüringer Bericht mutmaßlich<sup>2</sup> die Gesamtprävalenz genutzt. Wenn dem so wäre, wäre dies eine sehr simple, dafür aber vermutlich ungenaue Schätzung. Wenn aber korrekterweise nur die Stichprobe der Thüringer<sup>3</sup> berücksichtigt worden wäre, wäre die Stichprobe entsprechend kleiner und die Unsicherheit (viel) größer. Nimmt man beispielhaft an, dass die Stichprobe des Glücksspiel-Surveys die Bevölkerungsanteile der 18- bis 70-Jährigen in den Bundesländern<sup>4</sup> akkurat abbildet. Dann wären von den 12.003 18- bis 70-jährigen Befragten des Surveys 2,48 %<sup>5</sup>, also 298, aus Thüringen. Nimmt man weiter an, dass der Anteil an Menschen mit Glücksspielstörung in Thüringen mit dem Bundesdurchschnitt übereinstimmt, dann haben 2,3 % der Thüringer, also 7 Befragte, eine Glücksspielstörung. Entsprechend groß wäre das 95 %-Konfidenzintervall für den Anteil an Menschen mit Glücksspielstörung in Thüringen: 1,1 % bis 4,7 %. Hochgerechnet mit den Bevölkerungszahlen aus 2021 hätten demnach 15.636 bis 66.243 der 18- bis 70-jährigen Thüringer eine Glücksspielstörung – der wahre Wert könnte also doppelt oder auch nur halb so groß wie der im Bericht genannte sein. Insofern ist die Zahl der angeblichen 32.400 Menschen mit Glücksspielstörung irreführend und vermittelt eine Genauigkeit, die statistisch so nicht existiert.

Problematisch ist weiterhin die Feststellung im Glücksspielsuchtbericht, dass „die Zahl der betroffenen

<sup>2</sup>Berechnungen unter dieser Annahme führen zu genannten Zahl.

<sup>3</sup>Die dann nicht mehr nach Bundesland gewichtet ist.

<sup>4</sup>Stand 2021 (Statistisches Bundesamt (Destatis) 2023)

<sup>5</sup>Basis: Bevölkerungsdaten 18- bis 70-Jährige der Bundesländer (Statistisches Bundesamt (Destatis) 2023)

Thüringer:innen von ehemals ca. 11.000 nach oben korrigiert werden“ (S. 5) muss. Die Zahl der 11.000 Menschen mit Glücksspielproblemen basiert auf einer Hochrechnung des BZgA-Glücksspiel-Surveys von 2019, die den deutschlandweiten Anteil der 16- bis 70-Jährigen wahrscheinlich pathologisch Glücksspielenden – es wurde ein anderes Messinstrument genutzt – mit 0,34 % (Konfidenzintervall für Thüringen: 0,02 % – 1,95 %<sup>6</sup>) angab. Die Aussage im Bericht impliziert, dass es eine wirkliche Veränderung gab, die nicht das Produkt einer zufälligen statistischen Schwankung ist. Dies wird im Bericht nicht nachgewiesen. Im Gegenteil, Fishers exakter Test (dessen Anwendung unverzerrte Daten unterstellt) zeigt, dass der Anteil der Thüringer Befragten<sup>7</sup> mit Glücksspielstörung sich im BZgA Survey 2019 und im Glücksspiel-Survey 2021 nicht signifikant unterscheiden ( $p > 0.05$ ) – die vermeintliche Änderung könnte also das Produkt zufälliger statistischer Schwankung sein. Dies überrascht kaum, wenn man sich klarmacht, dass in absoluten (gewichteten) Zahlen in dem BZgA-Survey ein Thüringer Befragter eine Glücksspielstörung hatte, während es im Glücksspiel-Survey 2021 sieben sind. Eine Hochrechnung kann gleichwohl trotz der geringen Fallzahlen durchgeführt werden, allerdings muss die statistische Unsicherheit zwingend mitkommuniziert werden. Zwar weist die Thüringer Landesregierung in einer Fußnote darauf hin, dass es einen Wechsel im Messinstrument gab, schließt daraus aber nicht auf eine mangelnde Vergleichbarkeit. Doch allein dieser methodische Wechsel – die Instrumente legen jeweils einen unterschiedlichen Fokus auf Aspekte einer Glücksspielstörung – könnte den Unterschied in den Anteilen der beiden Stichprobe erklären. Der wahre Wert der Thüringer Bevölkerung kann unterdessen unverändert geblieben sein. Anzumerken ist, dass eine Hochrechnung auf Basis der gesamtdeutschen Prävalenz zwar statistisch signifikante Unterschiede zwischen beiden Untersuchungen ergeben würde. Die nicht quantifizierbare, zusätzliche systematische Verzerrung spricht aber gegen ein solches Vorgehen.

Zusammengefasst hat der Bericht der Thüringer Landesregierung zur Glücksspielsucht die definitive Aussage einer verdreifachten Zahl von Menschen in Thüringen mit problematischem Spielverhalten zwischen 2019 und 2021 getroffen, obwohl die hochgerechnete Zahl auf Basis des Glücksspiel-Surveys 2021 auch schlicht das Produkt einer zufälligen Schwankung sein kann. Diese Falschinformation wurde im Thüringer Kabinett (Süddeutsche Zeitung 2023) und 53 Medien<sup>8</sup> aufgegriffen und unkritisch rezitiert. An diesem Fallbeispiel kann exemplarisch die Dynamik betrachtet werden, die irreführende Kommunikation von Studienergebnissen – angefangen beim Glücksspiel-Survey und seinen Autoren

<sup>6</sup>Unter Annahme einer gleichen Prävalenz wie im durchschnittlichen Bundesgebiet sowie einer Beteiligung an dem Survey entsprechend der Bevölkerungsanteile der Bundesländer für die 18- bis 70-Jährigen.

<sup>7</sup>Für beide Surveys wurde hier einen Anteil an Thüringer Befragten entsprechend der Bevölkerungsanteile der Bundesländer für die 18- bis 70-Jährigen (Glücksspiel-Survey 2021) und 16- bis 70-Jährigen (BZgA 2019) angenommen. Die unterschiedlichen Altersgruppen basieren auf den Berechnungsgrundlagen des jeweiligen Surveys.

<sup>8</sup>Darunter die Süddeutsche Zeitung, Welt Online, Zeit Online und Frankfurter Allgemeine Zeitung Online (alle: „direkt aus dem dpa-Newskanal“)

selbst, über die fragwürdige Rezeption seitens einer Landesregierung bis hin zur Weiterverbreitung durch die Medien – entfalten kann. Zwar können diese weitläufigen Konsequenzen nicht ausschließlich den Autoren des Glücksspielsurvey 2021 angelastet werden, eher mangelt es sowohl bei Medienvertretern als auch innerhalb der Thüringer Landesregierung an Datenkompetenz. Dennoch hätten die Studienautoren solchen Fehlinterpretationen und ungenauen Hochrechnungen vorbeugen können, indem sie die Limitationen ihrer Arbeit (Stichprobenverzerrung, statistische Unsicherheit etc.) — transparent kommuniziert hätten. Denn wenn selbst die Studienautoren die statistische Unsicherheit nicht überprüfen und erwähnen, vermittelt dies dem mit Statistik wenig vertrauten Rezipienten eine Genauigkeit und Korrektheit der Daten (die so nicht existiert), was zu solchen ungenauen Hochrechnungen verleiten kann.

### 10.5.3 Beauftragter der Bundesregierung für Sucht- und Drogenfragen

Auch der Beauftragte der Bundesregierung für Sucht- und Drogenfragen, Burkhard Blienert, hat die von den Survey-Autoren vorgenommene Verschiebung des Schwellenwertes für problematisches Spielverhalten übernommen (vgl. Kapitel 7.1), die Daten des Glücksspiel-Surveys 2021 ohne Konfidenzintervalle zur Hochrechnung auf ganz Deutschland genutzt und somit eine Genauigkeit suggeriert, die in Wahrheit nicht existiert (Deutsches Ärzteblatt 2022). Via X (ehemals Twitter) verbreitete er folgende Nachricht: *„Beim #LottoTalk im Fernsehturm: Wenn 8 % der Deutschen ein Problem mit Glücksspiel haben, dann muss etwas passieren. Ich fordere deshalb: Keine Sportwettenwerbung mehr zur Primetime und eine gemeinsame Kraftanstrengung bei der Bekämpfung des illegalen Glücksspiels“* (Beauftragter der Bundesregierung für Sucht- und Drogenfragen 2023).

Die Angabe des Bundesdrogenbeauftragten von 8 % stammt vermutlich aus dem Glücksspielsurvey 2021, der in der Stichprobe einen Anteil von 5,7 % „riskanten Spielern“ und einen Anteil von 2,3 % von Spielern mit einer Glücksspielstörung angibt und diese zu 8 % „Problemspielern“ summiert (Buth, Meyer und Kalke 2022, S. 6). Ungeachtet der vorliegenden Stichprobenverzerrung (siehe Kapitel 4) unterläuft Blienert hier ein Fehler, denn die Angabe von 8 % aus dem Survey bezieht sich nicht auf die deutsche Gesamtbevölkerung und auch nicht auf alle erwachsenen Deutschen, sondern nur auf die deutsche Bevölkerung im Alter von 18-70 Jahren, da nur diese Altersgruppen befragt wurden. Die 8 % sind insofern falsch, denn sie erfordern die Annahme, dass unter 18-Jährige und über 70-Jährige eine Glücksspielstörungs-Prävalenz von ebenfalls 8 % hätten. Angesichts des mangelnden Zugangs der unter 18-Jährigen zum Glücksspiel und der bereits niedrigeren Prävalenz für 56- bis 70-Jährige (0,9 %) ist diese Annahme sehr unrealistisch, geeigneter erscheint eine Schätzung der Prävalenz für alle

Erwachsenen unter Nutzung der Prävalenz der 56- bis 70-Jährigen für die übrigen Erwachsenen.

Der vom Beauftragten der Bundesregierung für Sucht- und Drogenfragen genannte Wert (8 %) überschätzt somit höchstwahrscheinlich die Prävalenz des problematischen Glücksspielverhaltens in Deutschland und kommuniziert daher eine bestenfalls ungenaue und zu hohe Zahl. Seine Fehlinterpretation ist allerdings, wie die anderer Akteure, nachvollziehbar, da die Survey-Autoren die Limitationen ihrer Erhebung nicht transparent diskutiert haben — nicht nur, aber insbesondere auch in Bezug auf genau jene Stellen im Survey, auf die sich der Bundesdrogenbeauftragte offensichtlich bezieht: Weder wird seitens der Survey-Autoren konsistent<sup>9</sup> auf die Einschränkung verwiesen, dass die angegebenen Anteile von Problemspielern sich ausschließlich auf die 18- bis 70-jährige Bevölkerung (und eben nicht die Gesamtbevölkerung) beziehen, noch wird die Stichprobenverzerrung thematisiert, die eine Inferenz auf die Gesamtbevölkerung verbietet. Im Falle des Beauftragten der Bundesregierung für Sucht- und Drogenfragen ist die erfolgte Fehlinterpretation und deren Weiterverbreitung jedoch besonders bedauerlich, da die limitierten und statistisch ungenauen Ergebnisse des Surveys hierdurch qua Amt des Bundesdrogenbeauftragten, der der Bundesregierung zugehörig ist und in der Bevölkerung besondere Glaubwürdigkeit genießt, Aufwertung erfahren haben.

#### 10.5.4 Bundesministerium für Gesundheit

Als Folge der nicht angegebenen und kommunizierten Unsicherheit seitens der Autoren des Surveys wurden zudem viele weitere vermeintliche Unterschiede zwischen Gruppen in der Öffentlichkeit verbreitet, die statistisch nicht signifikant sind. Dass die Unterschiede nicht statistisch signifikant sind, bedeutet, dass basierend auf den vorliegenden Daten nicht mit ausreichender Sicherheit festgestellt werden kann, dass diese Unterschiede tatsächlich in der Gesellschaft existieren und nicht nur durch Zufall in der Stichprobe vorhanden sind. Ein prominentes Beispiel für einen solchen vermeintlichen Gruppenunterschied liefert das Online-Gesundheitsportal des Bundesministeriums für Gesundheit (BMG). Dieses schreibt zur Prävalenz von Glücksspielproblemen: *„Obwohl Glücksspiel erst ab 18 Jahren erlaubt ist, zeigen etwa 0,3 bis 3,0 % der Jugendlichen unter 18 Jahren ein problematisches Glücksspielverhalten. Jungen sind dabei deutlich häufiger betroffen als Mädchen.“* (Bundesministerium für Gesundheit 2022). Lobenswert ist zwar die Referenz des Konfidenzintervalls, allerdings ist der behauptete Geschlechterunterschied, wie ausführlich unter Abschnitt 4.1 erläutert, nicht statistisch signifikant. Basierend auf den Daten des Glücksspiel-Surveys 2021 lässt sich also nicht sagen, ob Jungen oder Mädchen häufiger von Glücksspielproblemen betroffen sind. Die Aussage auf dem BMG-Gesundheitsportal ist demnach

<sup>9</sup>Nicht in Einleitung, Executive Summary und Fazit.

nicht belegt, und das, obwohl die Informationen laut eigener Aussage von der Deutschen Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e.V. geprüft wurden<sup>10</sup>.

### 10.5.5 Landesregierung Nordrhein-Westfalen

In einer Antwort der nordrhein-westfälischen Landesregierung auf eine kleine Anfrage einiger SPD-Abgeordneter<sup>11</sup> im Landtag vom 25. Mai 2023 (LT-Drs. 18/4799) wird auf den Glücksspiel-Survey 2021 zur Einschätzung der Anzahl an Personen mit Glücksspielstörung in Nordrhein-Westfalen Bezug genommen. Hierbei wird der deutschlandweite Anteil an Personen mit Glücksspielstörung (2,3 %) aus dem Survey auf alle 18- bis 70-Jährigen in Nordrhein-Westfalen hochgerechnet, was eine Zahl von „etwa 280.000 Personen“ ergibt. Diese Rechnung ist ausweislich der Kürze der Antwort der Landesregierung und unter der expliziten Einschränkung, lediglich „Orientierung“ bieten zu wollen, nachvollziehbar. Dennoch unterliegt diese Hochrechnung der Glücksspielproblemprävalenz auf die Ebene eines einzelnen Bundeslandes einer durch die kleine Stichprobe induzierten Unsicherheit, die ebenfalls kommuniziert werden müsste (beispielsweise in Form eines Konfidenzintervalls für den Anteil). Die Limitationen einer Hochrechnung der Aggregatdaten auf einzelne Bundesländer wurde bereits ausführlich anhand des Thüringer Glücksspielsuchtberichts erläutert (siehe 10.5.2). Unabhängig davon wird in dieser Antwort der Landesregierung Nordrhein-Westfalens nicht auf die weiteren Limitationen des Surveys eingegangen, wie die niedrige Ausschöpfung und den Umfragenfehler (siehe 4). Mit einer den wissenschaftlichen Standards entsprechenden Diskussion der Schwächen in dem Survey hätte man diesen Verfehlungen bei der Rezeption vermutlich vorbeugen können. Abermals zeigt sich, wie wichtig transparente Kommunikation von wissenschaftlichen Ergebnissen und ihrer Limitationen seitens der verantwortlichen Wissenschaftler ist, um Fehlinterpretationen und -rezeptionen zu vermeiden

## 10.6 Rechtsprechung

Darüber hinaus wurden der Glücksspiel-Survey 2021 und seine statistisch kritikwürdigen Ergebnisinterpretationen seitens der Autoren bereits in mehrere obergerichtlichen Verfahren eingebracht, wodurch der Survey trotz seiner evidenten Intransparenz und nachgewiesener methodischer Fehler und Verzerrungen Eingang in die glücksspielrechtliche Rechtsprechung gefunden hat. Zu nennen sind die

<sup>10</sup>Ähnliche Fehlinterpretationen basierend auf dem Glücksspiel-Survey werden auch auf der Seite des Sächsischen Staatsministeriums für Soziales und Gesellschaftlichen Zusammenhalt vorgenommen (Sächsisches Staatsministerium für Soziales und Gesellschaftlichen Zusammenhalt 2022)

<sup>11</sup>Carolin Kirsch, Elisabeth Müller-Witt, Christina Kampmann und Sven Wolf

beiden folgenden Beschlüsse:

- Verwaltungsgerichtshof München, Beschluss v. 21.03.2023, 23 CS 22.2677 (VGH München 2023),
- Oberverwaltungsgericht Sachsen-Anhalt, Beschluss v. 15.06.2023, 3 M 24/23 (OVG Sachsen-Anhalt 2023).

Abermals zeigt sich die Erforderlichkeit sowohl akkuraten wissenschaftlichen Vorgehens und differenzierter Begleitkommunikation der verantwortlichen Studienautoren als auch funktionierender Qualitätssicherungsmaßnahmen für Studien mit breiter Öffentlichkeitswirkung. Die Autoren des Glücksspiel-Surveys 2021 haben sich dazu entschieden, im Survey auf eine umfangreiche und korrekte Diskussion der Aussagekraft sowie der Limitationen ihrer Arbeit zu verzichten. Das eigentlich obligatorische Peer-Review und eine kritische Diskussion der Forschungsergebnisse in der Fachliteratur, wodurch einer weiteren Verbreitung statistischer Fehlinterpretation der Survey-Ergebnisse vorgebeugt hätte werden können, haben nicht stattgefunden.

## 10.7 Medien

Auch seitens der Medien wurde vielfach auf Basis des Glücksspiel-Surveys fragwürdige Informationen verbreitet: „Am höchsten sei der Anteil der Frauen und Männer mit einer Störung unter den 21- bis 25-Jährigen – er betrage 4,3 Prozent. Bei den 26- bis 35-Jährigen seien es 4,2 Prozent [...]“ (ZEIT ONLINE 2022). Berücksichtigt man allerdings die Unsicherheit, umfasst das 95 %-Konfidenzintervall des Anteils der 21- bis 25-Jährigen (mit den diskutierten Einschränkungen mindestens) 4 % bis 4,7 % und das der 26- bis 35-Jährigen 3,9 % bis 4,6 %. Der Unterschied zwischen beiden Anteilen ist entsprechend nicht statistisch signifikant ( $p > 0.05$ ). Insofern ist die Aussage, „Am höchsten sei der Anteil der Frauen und Männer mit einer Störung unter den 21- bis 25-Jährigen [...]“ aus empirisch-statistischer Sicht falsch. Auch wenn diese Fehlinterpretation den Journalisten zugeordnet werden kann, so liegt doch die Hauptverantwortung für die Kommunikation und Präsentation der Ergebnisse bei den Studienautoren selbst. Da diese sich entschieden hatten, die statistische Unsicherheit der Inferenzen nur äußerst selten (und selektiv) zu kommunizieren, können Fehlinterpretationen und Falschaussagen wie die obige nicht überraschen.

## 11 | Fazit

Der Glücksspiel-Survey 2021 wurde im vorliegenden Gutachten aufgrund seiner (vorgesehenen) Nutzung durch staatliche Stellen an höchsten Qualitätsstandards gemessen. Bei der Festlegung der qualitativen Messlatte ist es dabei im Ergebnis irrelevant, ob diese staatliche und politische Nutzung von den Autoren des Surveys intendiert war<sup>1</sup>, denn alle Umfrageergebnisse sollten so kommuniziert werden, dass die Leser ihre Qualität eigenständig evaluieren können (ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. 2023). Dies ist beim Glücksspiel-Survey 2021 nicht der Fall. Der Survey thematisiert die offensichtlich existierenden Probleme hinsichtlich *Nonresponse* nicht, kommuniziert den Stichprobenfehler nur unzureichend und diskutiert auch bestehende Verzerrungen durch die Nutzung eines Online-Panels nicht. Mehr noch, eine Anfrage des Surveymaterials (Daten, Fragebogen, etc.) im Rahmen dieses Gutachtens wurde von den Studienautoren abgelehnt – dabei sollten die angefragten Materialien nach guter wissenschaftlicher Praxis ohnehin öffentlich zugänglich sein (Deutsche Forschungsgemeinschaft 2019).

Die fehlende transparente Kommunikation der Limitationen des Surveys, die nach Münnich (2023) bereits eine niedrige Qualität nach sich zieht, erschwert es ungemein, die Qualität des Glücksspiel-Surveys 2021 einzuschätzen. Viele Charakteristika des Surveys lassen vermuten, dass die Daten deutlich verzerrt sind: So liegt die *Nonresponse*-Quote bei der telefonischen Befragung bei 73 % und bei der Online-Befragung bei 90 %, was höchstwahrscheinlich zu einer starken Verzerrung führt. Dieses Problem wird im Survey allerdings weder diskutiert noch werden diesbezüglich weitere Informationen geliefert, die eine Einschätzung der Ausmaße und Implikationen des Problems ermöglicht hätten. Gleiches gilt für alle weiteren Aspekte des *Total-Survey-Errors* – die vorliegenden Informationen lassen häufig auf mutmaßliche Verzerrungen schließen, doch eine genaue Quantifizierung und Evaluierung dieser ist aufgrund mangelnder Informationen nicht möglich.

---

<sup>1</sup>Sie war es zumindest in Teilen, siehe Abschnitt 2.

Der Glücksspiel-Survey 2021 ist daher eine *Blackbox*, deren Ergebnisse für Dritte größtenteils nicht nachvollzogen und überprüft werden können. Dies widerspricht klar wissenschaftlichen Standards (Deutsche Forschungsgemeinschaft 2019), mindert die Qualität des Surveys und spricht gegen eine Nutzung des Surveys durch staatliche Akteure (Münnich 2023). Denn ohne eine Replizierbarkeit der Surveyergebnisse und einen transparenten Datengenerierungsprozess kann die Präzision und Korrektheit der Ergebnisse nicht sichergestellt werden. Dabei sind diese beiden Aspekte von essenzieller Bedeutung bei staatlichen Entscheidungen, um deren Korrektheit und Unabhängigkeit – beispielsweise in einem Gerichtsverfahren – sicherzustellen. So wurde der Glücksspiel-Survey 2021 unter anderem in zwei Gerichtsentscheidungen genutzt, ungeachtet dessen mangelnder Transparenz und fragwürdiger Präzision. Nun sind Richter keine Survey-Spezialisten, was sie allerdings auch nicht sein müssen, denn die Verantwortung für eine transparente Kommunikation von Studienergebnissen mit all ihren Limitationen liegt klar bei den Studienautoren (ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. 2023). Insofern sind die zahlreichen Fehlinterpretationen des Surveys durch staatliche Stellen mittelbar Folge der Intransparenz und – gemessen an Kriterien, die zur Erreichung der selbstgesetzten Ziele erfüllt sein müssen – einer unzureichenden wissenschaftlichen Qualität des Glücksspiel-Surveys.

Die fehlende Transparenz bei den Limitationen des Surveys erschwert zudem die Evaluation der Präzision der Ergebnisse. Aufgrund der erheblichen *Nonresponse*-Quote, der Nutzung eines *Nonprobability Samples* und der unzureichenden Gewichtung ist von einer starken Verzerrung der Daten auszugehen, sodass eine niedrige Präzision höchstwahrscheinlich anzunehmen ist. Durch die Inklusion des Online-Panels sollten beispielsweise die Anteilswerte für Spieler mit Glücksspielstörung aufgrund des *Selection-Errors* höher sein, als sie es in Wahrheit in der Bevölkerung sind (Sturgis und Kuha 2022). In welche Richtung sich die übrigen Verzerrungen auswirken, ist aufgrund fehlender Informationen seitens der Autoren nicht einzuschätzen. Die Präzision der Ergebnisse des Glücksspiel-Surveys wird allerdings aufgrund all dieser Verzerrungen niedrig sein und daher nicht dem hohen Qualitätsstandard entsprechen, der für die Ziele des Surveys notwendig ist.

Wegen seiner evidenten Intransparenz und der nachgewiesenen methodischen Fehler und Verzerrungen ist von einer Nutzung des Glücksspiel-Surveys 2021 durch staatliche Stellen im Zusammenhang mit gesellschaftspolitischen Bewertungen und erst recht im Zusammenhang mit glücksspielrechtlicher Regulierung dringend abzuraten. Er liefert keine valide und präzise Datenbasis in Bezug auf das Glücksspielverhalten in Deutschland. Damit soll dem Survey nicht grundsätzlich sein Wert für die Forschung abgesprochen werden<sup>2</sup>, für eine Berücksichtigung bei Entscheidungen mit Konsequenzen

<sup>2</sup>Eine explorative Analyse ist beispielsweise denkbar, da der Survey der Generierung zahlreicher Hypothesen dienen

für die breite Gesellschaft ist die Belastbarkeit der Survey-Ergebnisse allerdings nicht ausreichend. Doch auch die „Wissenschaftliche Studie zur Vorbereitung der Evaluierung der Sechsten Verordnung zur Änderung der Spielverordnung“<sup>3</sup> (Bühringer u. a. 2023), verwendet die Ergebnisse des Glücksspiel-Surveys aufgrund mangelnder Transparenz nur eingeschränkt – ein weiterer deutlicher Hinweis, dass der Glücksspiel-Survey nicht unkritisch rezipiert werden sollte. Dies wird insbesondere im Vergleich zu den vorherigen (BZgA-)Surveys deutlich, die sowohl hinsichtlich Transparenz als auch Präzision qualitativ höherwertige Ergebnisse lieferten.

Ebenso sollte von kausalen Aussagen bezüglich der Effektivität bestehender Regulatorik, der Gefährlichkeit einzelner Spielformen und den Wirkungen des neuen Glücksspielstaatsvertrag auf Basis der Daten des Surveys abgesehen werden, da dieser dafür aus methodischer Sicht schlichtweg nicht geeignet ist – obwohl die Studienautoren anderslautende Aussagen treffen.

Für zukünftige Surveys ist grundsätzlich eine Längsschnittstudie zu empfehlen, die auf *Probability Sampling* mit wohldefiniertem *Sampling-Frame* basiert<sup>4</sup>, ähnliches forderten bereits Banz (2019) und Hayer u. a. (2022). Eine solche Studie sollte von Beginn an die Erhebung verschiedener Hilfsvariablen zur Korrektur der *Nonresponse* und weiterer Verzerrungen einplanen, die Gewichtung selbst muss sich, wie auch die übrigen genutzten Methoden, an aktueller methodischer Literatur orientieren. Ebenso ist ein *Pretest* des Fragebogens für Screening-Zwecke aufgrund der sensiblen Fragen als auch eine explizite Abfrage der Legalität des genutzten Glücksspiels dringendst zu empfehlen. Selbstverständlich sollten auch Empfehlungen guter wissenschaftlicher Praxis (Deutsche Forschungsgemeinschaft 2019) vollumfänglich umgesetzt werden, also Limitationen transparent diskutiert und die Daten sowie alle relevanten Informationen (Fragebogen, Feldbericht, Code) öffentlich zur Verfügung gestellt werden. Nur so kann die Qualität der erhobenen Daten unabhängig evaluiert und sichergestellt werden.

Sollten die Daten den notwendigen hohen Qualitätsstandards entsprechen, könnten sie anschließend auch im Rahmen evidenzbasierter Politik verwendet werden. Als Positivbeispiel für die datenbasierte Unterstützung der Politik durch einen effektiven Qualitätssicherungsmechanismus sei an dieser Stelle auf Großbritannien verwiesen. Dort wird die Nutzung des „Code of Practice“ der *UK Statistics Authority* nicht nur allen Institutionen des öffentlichen Sektors dringend empfohlen, sondern in einer Übersicht sind zudem staatliche und nicht-staatliche Stellen mit ihrer jeweiligen Selbstverpflichtung zur Orientierung an statistischen Qualitätsstandards aufgeführt. Mehrere offizielle Beratungsgremien können in Anspruch genommen werden, um die Qualität von Daten und statistischen Ergebnissen

kann.

<sup>3</sup>Die vom Bundesministerium für Wirtschaft und Klimaschutz in Auftrag gegebenen wurde.

<sup>4</sup>Vorbild könnte hier beispielsweise die schwedische Längsschnittstudie Swelogs sein.

sicherzustellen.

Es sei im Folgenden zitiert (UK Statistics Authority 2023a):

*„You should*

- *Be ethical and honest in using any data*
- *Consult professional analysts to assure the appropriate use of evidence*
- *Use firm evidence that measures what is needed*
- *Present information accurately, clearly and impartially*

*But you should not*

- *Share unpublished data and statistics without authorised pre-release access*
- *Selectively quote favourable data*
- *Use other people’s data without checking their reliability first“*

Ein ethischer Umgang mit Daten, der in der Politik selbstverständlich sein sollte, orientiert sich an diesen Grundsätzen, die das Bild von „Statistics for the public good“ zum Ziel haben. Ein wichtiges Instrument ist in diesem Kontext die Möglichkeit einer Selbst-Evaluation von Datenerhebungen und Studien, die vom Datenethik-Team der UK Statistics Authority entwickelt und angeboten wird (UK Statistics Authority 2023b).

An dieser Stelle sei abschließend der Hinweis gegeben: Dieses Gutachten versteht sich ausdrücklich als Plädoyer für eine Regulierung, die auf der Basis qualitativ hochwertiger Daten und entsprechend gewonnener Evidenz agiert. Nur so kann langfristig ein Grundpfeiler der Demokratie, nämlich das Vertrauen in evidenzbasierte Politik, gestärkt werden.

# Bibliografie

- Abbott, Max, Ulla Romild und Rachel Volberg (2018). „The prevalence, incidence, and gender and age-specific incidence of problem gambling: Results of the Swedish longitudinal gambling study (Swelogs)“. In: *Addiction* 113.4, S. 699–707.
- Adams, Michael und Ingo Fiedler (2018). „§ 7. Die Sozialen Kosten Von Glücksspielen“. In: *Ökonomie, Recht, Sucht*. Hrsg. von Ihno Gebhardt und Stefan Korte. Berlin, Boston: De Gruyter, S. 133–144. DOI: doi:10.1515/9783110259216-007.
- ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (2023). *ADM fordert zum wiederholten Male, umfänglich und korrekt über die Entstehung und Aussagekraft von Umfrageergebnissen zu informieren*. URL: <https://www.adm-ev.de/2023/06/19/adm-fordert-zum-wiederholten-male-umfaenglich-und-korrekt-ueber-die-entstehung-und-aussagekraft-von-umfrageergebnisse-n-zu-informieren/> (besucht am 20.06.2023).
- Agresti, Alan und Brent A Coull (1998). „Approximate is better than “exact” for interval estimation of binomial proportions“. In: *The American Statistician* 52.2, S. 119–126.
- Arbeitsgemeinschaft ADM-Telefonstichproben (2016). *Forschungsprojekt Dual-Frame-2*. URL: <https://www.adm-ev.de/wp-content/uploads/2021/02/Dual-Frame-2.pdf> (besucht am 19.07.2023).
- Ausschuss für das Europäische Statistische System (2017). *Verhaltenskodex für europäische Statistiken*. URL: <https://ec.europa.eu/eurostat/documents/4031688/9394019/KS-02-18-142-DE-N.pdf/27ca19ca-e349-45f8-bbd4-4d78a33601ae?t=1542709797000> (besucht am 04.07.2023).
- Babbie, Earl R (2015). *The practice of social research*. 14th edition. Boston: Cengage learning.
- Baker, Reg, Stephen J Blumberg u. a. (2010). „Research synthesis: AAPOR report on online panels“. In: *Public Opinion Quarterly* 74.4, S. 711–781.
- Baker, Reg, J Michael Brick u. a. (2013). „Summary report of the AAPOR task force on non-probability sampling“. In: *Journal of survey statistics and methodology* 1.2, S. 90–143.

- Banz, Markus (2019). *Glücksspielverhalten und Glücksspielsucht in Deutschland. Ergebnisse des Surveys 2019 und Trends*. Köln: BZgA - Bundeszentrale für gesundheitliche Aufklärung. DOI: 10.17623/BZGA:225-GS-SY19-1.0.
- Banz, Markus und Tilman Becker (2019). „Glücksspielsucht in Deutschland: Häufigkeit und Bedeutung bei den einzelnen Glücksspielformen“. In: *Zeitschrift für Wett- und Glücksspielrecht*, S. 212–223.
- Beauftragter der Bundesregierung für Sucht- und Drogenfragen (25. Jan. 2023). *Beim LottoTalk im Fernsehturm: Wenn 8% der Deutschen ein Problem mit Glücksspiel haben, dann muss etwas passieren. Ich fordere deshalb: Keine Sportwettenwerbung mehr zur Primetime und eine gemeinsame Kraftanstrengung bei der Bekämpfung des illegalen Glücksspiels!* URL: [https://twitter.com/BdB\\_SD\\_Blienert/status/1618325599931731970?cxt=HHwWhMDQreKjufUsAAAA](https://twitter.com/BdB_SD_Blienert/status/1618325599931731970?cxt=HHwWhMDQreKjufUsAAAA) (besucht am 21.06.2023).
- Bethlehem, Jelke (2016). „Solving the nonresponse problem with sample matching?“ In: *Social Science Computer Review* 34.1, S. 59–77.
- Biemer, Paul P (2010). „Total survey error: Design, implementation, and evaluation“. In: *Public opinion quarterly* 74.5, S. 817–848.
- BILD (2022). *Sportwetten und Bier gehören nicht auf die Stadionbande*. URL: <https://www.bild.de/sport/fussball/fussball/bild-interview-zum-aktionstag-gegen-gluecksspielsucht-sportwetten-und-bier-gehoe-81454642.bild.html> (besucht am 19.06.2023).
- Blienert, Burkhard (2023a). *Alkohol, Tabak und Glücksspielwerbung benötigen viel engere Leitplanken*. URL: <https://www.bundesdrogenbeauftragter.de/presse/detail/alkohol-tabak-und-gluecksspielwerbung-benoetigen-viel-engere-leitplanken/> (besucht am 31.07.2023).
- (2023b). *Stimmungsbild Sucht- und Drogenpolitik: Alkohol, Tabak, Glücksspiel. Einstellungen der Bevölkerung zu Werbe- und Sponsoringregulierungen*. URL: [https://www.bundesdrogenbeauftragter.de/assets/Presse/2023/Stimmungsbild\\_Debatte\\_geSUCHT.pdf](https://www.bundesdrogenbeauftragter.de/assets/Presse/2023/Stimmungsbild_Debatte_geSUCHT.pdf) (besucht am 31.07.2023).
- Blom, Annelies G, Christina Gathmann und Ulrich Krieger (2015). „Setting up an online panel representative of the general population: The German Internet Panel“. In: *Field methods* 27.4, S. 391–408.
- Bosnjak, Michael u. a. (2013). „Sample composition discrepancies in different stages of a probability-based online panel“. In: *Field Methods* 25.4, S. 339–360.
- Brooker, Ivy S, Ian P Clara und Brian J Cox (2009). „The Canadian Problem Gambling Index: Factor structure and associations with psychopathology in a nationally representative sample.“ In: *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 41.2, S. 109.

- Brown, Lawrence D, T Tony Cai und Anirban DasGupta (2001). „Interval estimation for a binomial proportion“. In: *Statistical science* 16.2, S. 101–133.
- Browne, Matthew, Nancy Greer u. a. (2017). „A population-level metric for gambling-related harm“. In: *International Gambling Studies* 17.2, S. 163–175. DOI: 10.1080/14459795.2017.1304973.
- Browne, Matthew und Matthew Rockloff (2017). „The dangers of conflating gambling-related harm with disordered gambling. Commentary on: Prevention paradox logic and problem gambling (DeLafabro King, 2017)“. In: *Journal of Behavioral Addictions* 6.3, S. 317–320. DOI: 10.1556/2006.6.2017.059.
- Buchwald, C. (2011). „Hätten Sie Zeit für ein telefonisches Interview?“ In: *Junges altes Sachsen-Anhalt. Ein Bundesland im Blick der Sozialforschung*. Hrsg. von I. Wiekert und S. Winge. Halle: Universitätsverlag Halle-Wittenberg, S. 195–218.
- Bühringer, G. u. a. (2023). *Wissenschaftliche Studie zur Vorbereitung der Evaluierung der Sechsten Verordnung zur Änderung der Spielverordnung*. URL: %5Curl%7Bhttps://www.bmwk.de/Redaktion/DE/Textsammlungen/Mittelstand/gewerberecht.html%7D (besucht am 07.07.2023).
- Bundesministerium für Gesundheit (2022). *Glücksspielsucht (pathologisches Glücksspielen)*. URL: https://gesund.bund.de/gluecksspielsucht (besucht am 19.06.2023).
- Bundeszentrale für gesundheitliche Aufklärung (2017). *Gambling Behaviour and Pathological Gambling in Germany 2013*. GESIS Data Archive, Cologne. ZA6679 Data file Version 1.0.0. DOI: 10.4232/1.12748.
- Buth, S. und H. Stöver (2008). „Glücksspielteilnahme und Glücksspielprobleme in Deutschland: Ergebnisse einer bundesweiten Repräsentativbefragung“. In: *Suchttherapie* 9.1, S. 3–11. DOI: 10.1055/s-2008-1042440.
- Buth, Sven, Gerhard Meyer und Jens Kalke (2022). *Glücksspielteilnahme und glücksspielbezogene Probleme in der Bevölkerung–Ergebnisse des Glücksspiel-Survey 2021*. Hamburg: Institut für interdisziplinäre Sucht- und Drogenforschung (ISD).
- (2023). „Kenntnis und Akzeptanz von Spielerschutzmaßnahmen bei den Glücksspielenden–Ergebnisse des Glücksspiel-Surveys 2021“. In: *Suchttherapie*.
- Buth, Sven, Christian Schütze und Jens Kalke (2021). *Auswirkungen der Schließung von terrestrischen Glücksspielangeboten aufgrund eines pandemiebedingten Lockdowns auf das Glücksspielverhalten*. Hamburg: Institut für interdisziplinäre Sucht- und Drogenforschung (ISD).
- Calado, Filipa und Mark D Griffiths (2016). „Problem gambling worldwide: An update and systematic review of empirical research (2000–2015)“. In: *Journal of behavioral addictions* 5.4, S. 592–613.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2. Aufl. Routledge. DOI: 10.4324/9780203771587.

- Couper, Mick P (2000). „Web surveys: A review of issues and approaches“. In: *The Public Opinion Quarterly* 64.4, S. 464–494.
- Cumming, Geoff u. a. (2012). „The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis“. In: *Australian Journal of Psychology* 64.3, S. 138–146.
- Delfabbro, Paul und Daniel King (2017). „Prevention paradox logic and problem gambling: Does low-risk gambling impose a greater burden of harm than high-risk gambling?“ In: *Journal of Behavioral Addictions* 6.2, S. 163–167. DOI: 10.1556/2006.6.2017.022.
- Delfabbro, Paul H., Gambling Research Australia. u. a. (2010). *A review of Australian gambling research : implications for inter-jurisdictional public policy and regulation*. 2nd update version. Dept. of Justice [Melbourne].
- Deutsche Forschungsgemeinschaft (2019). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. DOI: 10.5281/zenodo.6472827.
- Deutscher Lotto- und Totoblock (2022). *Stellungnahme des Deutschen Lotto- und Totoblocks zum Glücksspiel-Survey 2021*. URL: [https://www.lottoindeutschland.de/assets/ctf/assets/z8xt1vqfl2wu/7AwlVJOQloX3mIXRC4eWbx/e92c770a7f01dcb9a26fbe633b8db833/2022-03-14\\_PK\\_Gluecksspiel-Survey\\_2021\\_Presseinformation.pdf](https://www.lottoindeutschland.de/assets/ctf/assets/z8xt1vqfl2wu/7AwlVJOQloX3mIXRC4eWbx/e92c770a7f01dcb9a26fbe633b8db833/2022-03-14_PK_Gluecksspiel-Survey_2021_Presseinformation.pdf) (besucht am 19.06.2023).
- Deutsches Ärzteblatt (2022). *Bundesdrogenbeauftragter: Sportwettensucht boomt*. URL: <https://www.aerzteblatt.de/nachrichten/139332/Bundesdrogenbeauftragter-Sportwettensucht-boomt> (besucht am 01.12.2022).
- (2023). *Suchtbeauftragter will Sportwettenwerbung erst nach 23 Uhr*. URL: <https://www.aerzteblatt.de/nachrichten/145210/Suchtbeauftragter-will-Sportwettenwerbung-erst-nach-23-Uhr> (besucht am 11.08.2023).
- Dutwin, David, Micheline Blum u. a. (2018). *Spam flagging and call blocking and its impact on survey research*. American Association for Public Opinion Research.
- Dutwin, David und Trent D Buskirk (2017). „Apples to oranges or gala versus golden delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples“. In: *Public Opinion Quarterly* 81.S1, S. 1–9.
- Dutwin, David und Paul Lavrakas (2016). „Trends in telephone outcomes, 2008-2015“. In: *Survey Practice* 9.3.
- Enderle, Tobias, Ralf Münnich und Christian Bruch (2013). „On the impact of response patterns on survey estimates from access panels“. In: *Survey Research Methods*. Bd. 7. 2, S. 91–101.
- Eurostat (2020). *European Statistical System handbook for quality and metadata reports: 2020 edition*. Publications Office of the European Union. DOI: 10.2785/666412.

- Fahimi, Mansour u. a. (2015). „Scientific surveys based on incomplete sampling frames and high rates of nonresponse“. In: *Survey Practice* 8.6, S. 1–12.
- Falk Kiefer, Hans-Jürgen Rumpf und (2011). „DSM-5: Die Aufhebung der Unterscheidung von Abhängigkeit und Missbrauch und die Öffnung für Verhaltenssüchte“. In: *Sucht* 57.1, S. 45–48. DOI: 10.1024/0939-5911.a000072.
- Ferris, J. und H. Wynne (2001). *The Canadian Problem Gambling Index: Final report*. Ottawa. URL: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F00772-000>.
- Fisher, Ronald A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver und Boyd.
- Gangl, Markus und Thomas A. DiPrete (2004). *Kausalanalyse durch Matchingverfahren*. Berlin.
- GapFish (2021). *PANELBOOK 2021*. URL: <https://gapfish.com/wp-content/uploads/2021/04/Gapfish-Panelbook-2021-DE.pdf> (besucht am 11. 07. 2023).
- Hayer, Tobias u. a. (2022). „Handlungsempfehlungen zur Evaluation des Glücksspielstaatsvertrags“. In: *SUCHT* 68.4, S. 237–245.
- Hodgins, David C, Jonathan N Stea und Jon E Grant (2011). „Gambling disorders“. In: *The Lancet* 378.9806, S. 1874–1884.
- Institut für interdisziplinäre Sucht- und Drogenforschung (14. März 2022). *Pressemitteilung: ISD und Universität Bremen legen den Glücksspielsurvey 2021 vor!* Zugriff am: 19.06.2023. URL: [https://www.lottoindeutschland.de/assets/ctf/assets/z8xt1v9fl2wu/3XZJLf8y1cuMdDw9YRFtEX/c202628455b60427d95ecfc41101849f/2022-03-14\\_Pressemeldung\\_ISD.pdf](https://www.lottoindeutschland.de/assets/ctf/assets/z8xt1v9fl2wu/3XZJLf8y1cuMdDw9YRFtEX/c202628455b60427d95ecfc41101849f/2022-03-14_Pressemeldung_ISD.pdf).
- Knorpp, Leonie Silvia (2013). „Evidence-based Public Health Nutrition: Entwicklung eines Konzepts für einen systematischen und standardisierten Ansatz zur Anwendung in Deutschland“. Diss. Justus-Liebig-Universität Gießen. URL: <http://geb.uni-giessen.de/geb/volltexte/2014/11111>.
- Krosnick, Jon A. und Stanley Presser (2010). „Question and Questionnaire Design“. In: *Handbook of Survey Research*. Hrsg. von Peter V Marsden und James D Wright. Second. Bingley, UK: Emerald Group Publishing.
- Kühnel, Steffen und André Dingelstedt (2022). „Kausalität“. In: *Handbuch Methoden der empirischen Sozialforschung*. Springer Fachmedien Wiesbaden, S. 749–762. DOI: 10.1007/978-3-658-37985-8\_46.
- Langham, Erika u. a. (2016). „Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms“. In: *BMC Public Health* 16.1. DOI: 10.1186/s12889-016-2747-0. URL: <https://doi.org/10.1186/s12889-016-2747-0>.
- Latvala, Tiina u. a. (2021). „18-Item Version of the Short Gambling Harm Screen (SGHS-18): Validation of Screen for Assessing Gambling-Related Harm among Finnish Population“. In: *International Journal of Environmental Research and Public Health* 18.21, S. 11552. DOI: 10.3390/ijerph182111552.

- Lavrakas, Paul John u. a. (2022). „Comparing Probability-Based Surveys and Nonprobability Online Panel Surveys in Australia: A Total Survey Error Perspective“. In: *Survey Research Methods*. Bd. 16, 2, S. 241–266.
- Lenau, Susanne u. a. (2021). *Methods for sampling and inference with non-probability samples Deliverable D11.8*. Leuven: InGRID-2 project 730998 - H2020.
- Lesieur, H. R. und S. B. Blume (Sep. 1987). „The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers“. In: *The American Journal of Psychiatry* 144.9, S. 1184–1188. DOI: 10.1176/ajp.144.9.1184.
- Li, En u. a. (2016). „Breaking Bad: Comparing Gambling Harms Among Gamblers and Affected Others“. In: *Journal of Gambling Studies* 33.1, S. 223–248. DOI: 10.1007/s10899-016-9632-8.
- Lohr, Sharon L (2021). *Sampling: design and analysis*. Boca Raton, Florida: CRC press.
- Lorant, Vincent u. a. (2007). „Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census“. In: *International Journal of Epidemiology* 36.6, S. 1292–1299.
- MacInnis, Bo u. a. (2018). „The accuracy of measurements with probability and nonprobability survey samples: Replication and extension“. In: *Public Opinion Quarterly* 82.4, S. 707–744.
- Manrique, A. u. a. (2023). *Pathologisches Glücksspielen*. URL: [https://www.bas-muenchen.de/wp-content/uploads/Pathologisches-Glueckspiel\\_2023\\_v2\\_230721.pdf](https://www.bas-muenchen.de/wp-content/uploads/Pathologisches-Glueckspiel_2023_v2_230721.pdf) (besucht am 17.08.2023).
- Mauz, Elvira u. a. (2023). „Time trends in mental health indicators in Germany’s adult population before and during the COVID-19 pandemic“. In: *Frontiers in public health* 11.
- McPhee, Cameron u. a. (2022). *Data Quality Metrics for Online Samples: Considerations for Study Design and Analysis*. American Association for Public Opinion Research.
- Meinfelder, Florian (2014). „Multiple Imputation: An attempt to retell the evolutionary process“. In: *AStA Wirtschafts-und Sozialstatistisches Archiv* 8, S. 249–267.
- Mercer, Andrew, Arnold Lau und Courtney Kennedy (2018). *For weighting online opt-in samples, what matters most?* Pew Research Center.
- Meyer, Christian, Anja Bischof u. a. (2015). „The “Pathological Gambling and Epidemiology”(PAGE) study program: Design and fieldwork“. In: *International Journal of Methods in Psychiatric Research* 24.1, S. 11–31.
- Meyer, Gerhard (2023). „Glücksspiel – Zahlen und Fakten“. In: *DHS Jahrbuch Sucht 2023*. Hrsg. von Deutsche Hauptstelle für Suchtfragen e.V. (DHS). Lengerich: Pabst Science Publishers. Kap. 2.4, S. 89–110.
- Meyer, Gerhard, Jens Kalke und Sven Buth (2023). „Problem gambling in Germany: results of a mixed-mode population survey in 2021“. In: *International Gambling Studies*, S. 1–18.

- Münnich, Ralf (2023). „Discussion of “Probability vs. Nonprobability Sampling: From the Birth of Survey Sampling to the Present Day” by Graham Kalton“. In: *Statistics in Transition. New Series* 24.3, S. 39–41.
- Nationale Akademie der Wissenschaften Leopoldina (2021). *Was ist evidenzbasierte Politikgestaltung?* URL: <https://www.leopoldina.org/themen/evidenzbasierte-politikgestaltung/politikgestaltung-22/> (besucht am 14. 08. 2023).
- Nationale Akademie der Wissenschaften Leopoldina, acatech – Deutsche Akademie der Technikwissenschaften und Union der deutschen Akademien der Wissenschaften (2015). *Wissenschaftliche und gesellschaftspolitische Bedeutung bevölkerungsweiter Längsschnittstudien*. Halle (Saale). URL: [https://www.leopoldina.org/uploads/tx\\_leopublication/2016\\_Stellungnahme\\_Laengsschnittstudien.pdf](https://www.leopoldina.org/uploads/tx_leopublication/2016_Stellungnahme_Laengsschnittstudien.pdf) (besucht am 16. 08. 2023).
- Nature (2023). *Reporting standards and availability of data, materials, code and protocols*. Zugriff am: 12.07.2023. URL: <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards> (besucht am 12. 07. 2023).
- Opp, Karl-Dieter (2010). „Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik“. In: *Handbuch der sozialwissenschaftlichen Datenanalyse*. VS Verlag für Sozialwissenschaften, S. 9–38. DOI: 10.1007/978-3-531-92038-2\_2.
- Orth, B. und C. Merkel (2022). *Der Substanzkonsum Jugendlicher und junger Erwachsener in Deutschland. Ergebnisse des Alkoholsurveys 2021 zu Alkohol, Rauchen, Cannabis und Trends*. Köln: Bundeszentrale für gesundheitliche Aufklärung. DOI: <https://doi.org/10.17623/BZGA:Q3-ALKSY21-DE-1.0>.
- Pffor, Klaus und Jette Schröder (2015). „Warum Panelstudien?“ In: *GESIS Survey Guidelines*. URL: [https://doi.org/10.15465/gesis-sg\\_008](https://doi.org/10.15465/gesis-sg_008).
- Pickering, Dylan und Alex Blaszczynski (2021). „Paid online convenience samples in gambling studies: Questionable data quality“. In: *International Gambling Studies* 21.3, S. 516–536.
- Plan International (2023). *Spannungsfeld Männlichkeit*. URL: [https://www.plan.de/fileadmin/website/04.\\_Aktuelles/Umfragen\\_und\\_Berichte/Spannungsfeld\\_Maennlichkeit/Plan-Umfrage\\_Maennlichkeit-A4-2023-NEU-online\\_2.pdf](https://www.plan.de/fileadmin/website/04._Aktuelles/Umfragen_und_Berichte/Spannungsfeld_Maennlichkeit/Plan-Umfrage_Maennlichkeit-A4-2023-NEU-online_2.pdf) (besucht am 11. 07. 2023).
- Russell, Alex MT u. a. (2022). „Are any samples representative or unbiased? Reply to Pickering and Blaszczynski“. In: *International Gambling Studies* 22.1, S. 102–113.
- Sächsisches Staatsministerium für Soziales und Gesellschaftlichen Zusammenhalt (2022). *Daten und Fakten*. URL: <https://www.gesunde.sachsen.de/daten-und-fakten-5453.html#ftn-3> (besucht am 19. 06. 2023).

- Särndal, Carl-Erik und Sixten Lundström (2005). *Estimation in surveys with nonresponse*. Hoboken, NJ: John Wiley & Sons.
- Schnell, Rainer (10. Aug. 2023). „Wie man sich eine Studie backt“. In: *Die ZEIT*. URL: <https://www.zeit.de/2023/34/rainer-schnell-meinungsumfragen-sozialforschung> (besucht am 12. 08. 2023).
- Scholes, Shaun u. a. (2008). „Understanding non-response to the British Gambling Prevalence Survey 2007“. In: *Prepared for the Gambling Commission by the National Centre for Social Research*. Available at <http://www.gamblingcommission.gov.uk/pdf/understanding%20non-responses%20to%20bgps> 20, S. 2007.
- Schütze, Christian, Jens Kalke und Sven Buth (2023). „Lockdown und der Wechsel ins Online-Glücksspiel. Das Glücksspielverhalten von Kasino- und Automatenspielern sowie Sportwettern in der Phase des ersten Covid-19-Lockdowns in Deutschland.“ In: *Zeitschrift für Wett- und Glücksspielrecht* 18.1, S. 22–29.
- Senator für Inneres der Freien Hansestadt Bremen (2023). *Gemeinsam gegen Sportwettenwerbung: Senator Mäurer empfängt Sucht- und Drogenbeauftragten der Bundesregierung Blienert und Glücksspielforscher Dr. Hayer*. URL: <https://www.senatspressestelle.bremen.de/pressemitteilungen/gemeinsam-gegen-sportwettenwerbung-senator-maeurer-empfaengt-sucht-und-drogenbeauftragten-der-bundesregierung-blienert-und-gluecksspielforscher-dr-hayer-418074> (besucht am 13. 06. 2023).
- Sleczka, Pawel u. a. (2015). „DSM-5 criteria for gambling disorder: Underlying structure and applicability to specific groups of gamblers“. In: *Journal of behavioral addictions* 4.4, S. 226–235. DOI: 10.1556/2006.4.2015.035. URL: <https://europepmc.org/articles/PMC4712756>.
- Statistisches Bundesamt (Destatis) (2023). *Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre*. URL: <https://www-genesis.destatis.de/genesis/online?operation=statistic&levelindex=0&levelid=1687347046035&code=12411#abreadcrumb> (besucht am 13. 06. 2023).
- Sturgis, P und J Kuha (2022). „How survey mode affects estimates of the prevalence of gambling harm: A multisurvey study“. In: *Public Health* 204, S. 63–69.
- Sturgis, Patrick (2020). *An assessment of the accuracy of survey estimates of the prevalence of problem gambling in the United Kingdom*. URL: <https://www.begambleaware.org/sites/default/files/2020-12/an-assessment-of-the-accuracy-of-survey-estimates-of-the-prevalence-of-problem-gambling-in-the-united-kingdom.pdf> (besucht am 13. 07. 2023).
- Süddeutsche Zeitung (2023). *Bericht: 32.400 Menschen haben Glücksspiel-Suchtprobleme*. URL: <https://www.sueddeutsche.de/politik/regierung-erfurt-bericht-32-400-menschen-haben-gluecksspiel-suchtprobleme-dpa.urn-newsml-dpa-com-20090101-230620-99-122340> (besucht am 20. 06. 2023).
- Syvertsen, André u. a. (2023). „Validation of the 7-Item Domain-General Gambling Harm Scale (DGHS-7)“. In: *Addictive Behaviors Reports* 17, S. 100499. DOI: 10.1016/j.abrep.2023.100499.

- Tagesschau (2023). *Aufregung um Umfrage zur Männlichkeit*. URL: <https://www.tagesschau.de/faktenfinder/umfrage-maennlichkeit-plan-100.html> (besucht am 11.07.2023).
- Tannahill, Andrew (2008). „Beyond evidence–to ethics: A decision-making framework for health promotion, public health and health improvement“. In: *Health Promotion International* 23.4, S. 380–390. DOI: 10.1093/heapro/dan032.
- Tedersoo, Leho u. a. (2021). „Data sharing practices and data availability upon request differ across scientific disciplines“. In: *Scientific Data* 8.1. DOI: 10.1038/s41597-021-00981-0.
- Thüringer Ministerium für Arbeit, Soziales, Gesundheit, Frauen und Familie (2023). 1. *Glücksspielsuchtbericht der Thüringer Landesregierung*. URL: [https://www.tmasgff.de/fileadmin/user\\_upload/Allgemein/Publikationen/Erster\\_Gluecksspielsuchtbericht.pdf](https://www.tmasgff.de/fileadmin/user_upload/Allgemein/Publikationen/Erster_Gluecksspielsuchtbericht.pdf) (besucht am 23.06.2023).
- Toto-Lotto Niedersachsen GmbH (2023). *Keno Spielregeln*. URL: <https://www.lotto.de/keno/spielregeln> (besucht am 01.07.2023).
- Tourangeau, Roger, Lance J Rips und Kenneth Rasinski (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, Roger und Ting Yan (2007). „Sensitive questions in surveys.“ In: *Psychological Bulletin* 133.5, S. 859.
- UK Statistics Authority (2023a). *Code of Practice for Statistics. For organisations that produce official statistics*. URL: <https://code.statisticsauthority.gov.uk/for-organisations-that-produce-official-statistics/> (besucht am 01.08.2023).
- (2023b). *Data Ethics*. URL: <https://uksa.statisticsauthority.gov.uk/what-we-do/data-ethics/> (besucht am 01.08.2023).
- United Nations (1954). *Handbook of Statistical Organization*. New York.
- Valliant, Richard (2020). „Comparing alternatives for estimation from nonprobability samples“. In: *Journal of Survey Statistics and Methodology* 8.2, S. 231–263.
- Yeager, David S u. a. (2011). „Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples“. In: *Public opinion quarterly* 75.4, S. 709–747.
- ZEIT ONLINE (2022). *Bremer Sportwettbüros droht die Schließung*. URL: <https://www.zeit.de/news/2022-07/27/maeurer-plant-schliessung-von-sportwettbueros-in-bremen> (besucht am 27.07.2023).
- Ziemann, Mark, Yotam Eren und Assam El-Osta (2016). „Gene name errors are widespread in the scientific literature“. In: *Genome biology* 17.1, S. 1–3.