

Verbrauchergerechtes Scoring



Zitierhinweis für diese Publikation:

SVRV (2018). Verbrauchergerechtes Scoring. *Gutachten des Sachverständigenrats für Verbraucherfragen*.
Berlin: Sachverständigenrat für Verbraucherfragen.

Berlin, Oktober 2018

Gutachten des Sachverständigenrats für Verbraucherfragen

ISSN 2510-0084

Herausgeber:

Sachverständigenrat für Verbraucherfragen

beim Bundesministerium der Justiz und für Verbraucherschutz

Mohrenstraße 37

10117 Berlin

Telefon: +49 (0) 30 18 580-0

Fax: +49 (0) 30 18 580-9525

E-Mail: info@svr-verbraucherfragen.de

Internet: www.svr-verbraucherfragen.de

Gestaltung: Atelier Hauer+Dörfler GmbH, Berlin

Druck: bud

© SVRV 2018

Verbrauchergerechtes Scoring



Danksagung

Es gehört zum Auftrag des Sachverständigenrats für Verbraucherfragen (SVRV), aktuelle Erkenntnisse aus der Wissenschaft und Erfahrungen aus der Praxis in die Erarbeitung seiner Veröffentlichungen einfließen zu lassen. Um diesem Auftrag möglichst umfassend gerecht zu werden, hat der SVRV unterschiedliche Veröffentlichungsreihen geschaffen (Gutachten, Working Paper und Studien im Auftrag des SVRV). Zudem richtet er Fachveranstaltungen aus und führt öffentliche und nicht-öffentliche Gespräche mit Wissenschaftlern, Unternehmensvertretern und Vertretern aus der Zivilgesellschaft. Mithilfe von repräsentativen und nicht-repräsentativen Befragungen gehen Meinungen der Bürgerinnen und Bürger in die Arbeit des SVRV genauso ein wie legitime Interessen der Unternehmen. Nur durch die Unterstützung und Mitwirkung dieser Personen und Institutionen war die Erstellung des vorliegenden Gutachtens möglich.

Der SVRV dankt allen Mitarbeiterinnen und Mitarbeitern der Geschäftsstelle des SVRV für die ausgezeichnete Mitarbeit bei der Erarbeitung des Gutachtens. Wir danken besonders den wissenschaftlichen Mitarbeitern der Geschäftsstelle, Johannes Gerberding, Dr. Christian Groß, Dr. Ariane Keitel und Sarah Sommer sowie Thomas Fischer, dem Leiter der Geschäftsstelle des SVRV, und Stefan Kubat als Vertretung von Herrn Fischer.

Dr. Felix Rebitschek, Jasmin Ghalib und Juri Ritz haben die federführenden Autoren des Gutachtens, Gerd Gigerenzer und Gert G. Wagner, in ihrer Arbeit unterstützt. Auch ihnen gilt der Dank des SVRV.

Das vorliegende Gutachten beruht außerdem in Teilen auf Arbeiten, die in der Working Paper Reihe des SVRV veröffentlicht wurden. Der SVRV dankt den Autorinnen und Autoren der Working Papers „Verbraucher-Scoring aus Sicht des Datenschutzrechts“ und „Dokumentation einer empirischen Pilot-Studie zum Wissen über und zur Bewertung von Verbraucher-Scoring“.

Zudem dankt der SVRV den Verfassern der Studie „Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren“, Prof. Dr. Georg Borges, Dr. Matthias Grabmair, Daniel Krupka, Prof. Burkhard Schäfer, Prof. Dr. Erich Schweighofer, Prof. Dr. Christoph Sorge und Bernhard Waltl von der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V..

Im Vorfeld und im Zuge der Erstellung des vorliegenden Gutachtens haben Mitglieder des SVRV und seiner Geschäftsstelle zahlreiche Gespräche mit Vertretern aus Wissenschaft sowie von Unternehmen und NGOs geführt. Dank gilt (in alphabetischer Reihenfolge): Prof. Dr. Andreas Beyer (Universität zu Köln), der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin), dem Bundesversicherungsamt (BVA), Wolfie Christl (Cracked Labs – Institut für kritische digitale

Kultur), Dacadoo AG, Generali Deutschland AG, Prof. Dr. Justus Haucap (Universität Düsseldorf), Lorena Jaume-Palasi (Algorithm-Watch), Prof. Dr. Genia Kostka (Freie Universität Berlin), infoscore Consumer Data GmbH, Prof. Dr. Sergio Lucia (Technische Universität Berlin), Prof. Dr. Stefan Lessmann (Humboldt Universität zu Berlin), Prof. Dr. Mario Martini (Deutsche Universität für Verwaltungswissenschaften Speyer), Prof. Dr. Steffen Mau (Humboldt-Universität zu Berlin), Prof. Dr. Klaus-Robert Müller (TU Berlin), Walter Palmetshofer (Open Knowledge Foundation e.V.), Prof. Dr. Eberhard Sandschneider (Freie Universität Berlin), SCHUFA Holding AG, Prof. Dr. Wolfgang Schulz (Hans-Bredow Institut und Alexander von Humboldt Institut für Internet und Gesellschaft), Arne Semsrott (Open Knowledge Foundation e.V.) und Sparkassen Direktversicherung AG.

Im Rahmen einer Marktstudie der Marktsegmente Bonitäts-, Kfz- und Gesundheits-Scoring haben sich insgesamt 75 Unternehmen bzw. Krankenversicherungen im Rahmen einer Befragung beteiligt. Der SVRV dankt den beteiligten Unternehmen herzlich für die Mitarbeit, ebenso den Bürgerinnen und Bürger, die bei einer repräsentativen Bevölkerungsbefragung zu Kenntnissen über und zur Akzeptanz von Scoring mitgewirkt haben. Den Survey führte infas Institut für angewandte Sozialwissenschaft unter der Projektleitung von Frau Janina Belz durch. All diesen wichtigen Partnern sei gedankt.

Dr. Philipp Hacker (WZB) und Christine Schäfer (acs plus: data with care) dankt der SVRV nicht nur für die kritische Durchsicht des Gutachten-Entwurfs im Sinne eines unabhängigen Peer Reviews, sondern auch für viele wichtige Anregungen, die in die Endfassung eingeflossen sind.

Abschließend sei darauf hingewiesen, dass die Sprache in diesem Text grundsätzlich geschlechterneutral gemeint ist. Auf eine durchgehende Nennung beider Geschlechter wurde zugunsten der besseren Lesbarkeit verzichtet.



Lucia Reisch



Gerd Gigerenzer



Gert G. Wagner

Executive Summary – Handlungsempfehlungen für verbraucher-gerechtes Scoring

Die folgenden an die Politik gerichteten Empfehlungen für verbrauchergerichtetes Scoring dienen der schnellen Orientierung. Die Begründung für die Auswahl dieser Empfehlungen findet sich in einer Langfassung am Ende des Gutachtens (Kapitel F).

1. Scoring für den Verbraucher verständlich machen

1. Der SVRV empfiehlt den Datenschutzbehörden, die in der DSGVO (vgl. Art. 15 Abs. 1 Buchst. h) festgeschriebenen Verständlichkeitsanforderungen für Scoring und score-basierte Geschäftsprozesse zu operationalisieren. Maßstab für Verständlichkeit sollte der Durchschnittsverbraucher sein. Soweit Scoring eine für den einzelnen Verbraucher nicht mehr nachvollziehbare Komplexität erreichen sollte, wäre sicherzustellen, dass Scoring nicht nur für Aufsichtsbehörden, sondern auch zumindest für Verbraucherorganisationen und Nicht-Regierungsorganisationen nachvollziehbar ist.
2. Scoring-Anbieter sollen den Verbrauchern die für sie wesentlichen Merkmale, auf deren Basis sie gescoret werden, sowie möglichst auch deren Gewichtung auf verständliche und nachvollziehbare Weise offenlegen. Allerdings dürfen Geschäftsgeheimnisse dadurch nicht verletzt werden. Festzustellen, welche Merkmale für Verbraucher wesentlich sind, kann nicht nur Aufgabe des Gesetzgebers sein, sondern sollte auch Aufgabe von Verbraucherorganisationen oder der „Marktwächter“ der Verbraucherzentralen sein. Eine vollständige Offenlegung der Scores und ihrer Eigenschaften gegenüber Aufsichtsbehörden ist in jedem Fall notwendig (vgl. das SVRV-Gutachten Digitale Souveränität).

Ein Teil des SVRV spricht sich für eine weiterreichende Scoring-Transparenz aus. Er ist der Auffassung, dass stets sämtliche in einen Score eingehende Merkmale gegenüber dem Verbraucher offenzu-

legen sind und ihr relatives Gewicht in der Score-Berechnung anzugeben ist. Etwaige Geheimhaltungsinteressen der Scoring-Anbieter und -Nutzer haben insoweit gegenüber dem Auskunftsinteresse der Verbraucher zurückzustehen, wobei das Geschäftsgeheimnis der Entwicklung und Programmierung eines Scores gewahrt bleiben sollte.

3. Ein besseres Verständnis der Verbraucher bezüglich Scoring wird durch die bloße Offenlegung nicht automatisch sichergestellt. Dafür sind verschiedene Maßnahmen notwendig, etwa: beispielhafte Darstellungen von Scores und deren Abstufungen für verschiedene Merkmale; didaktische, visuelle Aufbereitung (etwa durch Verbraucherorganisationen); allgemein Verbesserung der Scoring-bezogenen Kompetenzen der Verbraucher (Skillset). Die Verständlichkeit der Scores durch Verbraucher sollte nicht nur von Fachleuten behauptet, sondern empirisch überprüft werden.
4. Verbraucher haben jetzt schon einen Anspruch (vgl. Art. 13 Abs. 2 lit. f, 15 Abs. 1 Buchst. h DSGVO) darauf, in Textform auf verständliche Weise adressatengerecht darüber informiert zu werden, dass sie gescoret werden. Doch fehlt es an einer Konkretisierung. Unternehmen, Aufsichtsbehörden und Verbraucherorganisationen sollten gemeinsam Standards für Score-Anbieter entwickeln, damit Relevanz und Verständlichkeit gewährleistet werden können. Der SVRV empfiehlt zudem, Verbrauchern mitzuteilen, wie ihr persönlicher Score bezüglich der Verteilung aller Score-Werte in der Bevölkerung zu interpretieren ist (z. B. dass er „im oberen Drittel“ aller Score-Werte liegt).
5. Eine zeitnahe kostenlose Rückmeldung bei größerer Veränderung eines individuellen Scores (z. B. das Abrutschen in eine schlechtere Kategorie) ist zu gewährleisten – zumindest als wählbare Option. Dieser Mitteilung sind Grenzen gesetzt: Um eine Veränderung eines Scorewertes erkennen zu können, muss der Score-Anbieter alte Scorewerte speichern. Dies wird für viele praktische Anwendungen (wie bei Betrugserkennung oder der Bestimmung anzubietender Zahlungsoptionen) nicht der Fall sein. Die Berechnung eines Scores erfolgt

bei Versicherern und Banken anlassbezogen. Damit wird keine Historie aufgebaut und so wird die Veränderung beim nächsten „Event“ nicht klar. Damit kann sich die Forderung nur auf die Institutionen beziehen, die kontinuierlich Daten sammeln, etwa Bonitäts-Scorer und das Fahreignungsregister des Kraftfahrt-Bundesamtes in Flensburg („Verkehrssünderkartei“), das solche Mitteilungen bereits versendet).

2. Scoring-Wissen und Kompetenzen fördern

Wie bereits im SVRV-Gutachten „Digitale Souveränität“ empfohlen wurde, sollten staatliche Einrichtungen, Bildungsträger, NGOs, Verbraucherschutzorganisationen und Verbraucherschutzprojekte

1. Wissen zu Grundaspekten vermitteln, die alle Bereiche des Scorings und seine Verwendung in bestimmten Geschäftsfeldern betreffen. Hierzu sollte die Bundesregierung in ihrer Digitalisierungsstrategie innerhalb der laufenden Legislaturperiode *Informations- und Diskussionsmaterial* erarbeiten, um die Kompetenzen von Verbrauchern, Multiplikatoren und Entscheidungsträgern zu verbessern. Das Grundprinzip und die Qualitätsaspekte von Scoring sowie die Formen von Ungleichbehandlung und deren Ursachen gehören ebenso zum Grundwissen wie die einem Betroffenen zustehenden Rechte.
2. Kompetenzen fördern, um informierte Entscheidungen hinsichtlich der Teilnahme an einem Scoring-Verfahren zu treffen. Hierunter fällt die Kompetenz, Scoring-Angebote sowie Alternativen zu finden, kritisch zu prüfen, zu bewerten (z. B. werden die für den Verbraucher relevanten Informationen offengelegt) und zu nutzen.

3. Diskriminierung prüfen und offenlegen

1. Der SVRV empfiehlt, die Auskunftsansprüche von Verbrauchern zu stärken, wie sie in Art. 15 Abs. 1 Buchst. h DSGVO angelegt sind. Insbesondere sollten Verbraucher (sofern dies durch den Anbieter ermittelbar ist) erfahren können, wie die Scores zwischen verschiedenen Gruppen mit unterschiedlichen geschützten Merkmalen verteilt sind, um so eine mögliche algorithmische Diskriminierung dem Grunde nach belegen zu können.
2. Der SVRV empfiehlt zudem die Stärkung der Aufsichtsbehörden (siehe Handlungsempfehlung 7).
3. Der SVRV spricht sich ferner für ein Verbandsklagerecht für Antidiskriminierungssachverhalte im Bereich des Scorings aus.

4. Telematikfreie Option sicherstellen

1. Der SVRV empfiehlt, die Aufrechterhaltung telematikfreier Optionen im Versicherungsbereich (insbesondere Kfz-Haftpflichtversicherung, Krankenversicherung) gesetzlich zu garantieren. Insbesondere:
2. Versicherungsnehmern, die keinen Telematiktarif nutzen, dürfen keine substanziellen Nachteile im Vergleich zu Versicherungsnehmern mit Telematiktarif entstehen.
3. Telematiktarife sind nach Auffassung der Mehrheit des SVRV so auszugestalten, dass sie sich wirtschaftlich selbst tragen und dadurch nicht indirekt zu Lasten der Versicherungsnehmer gehen, die keinen Telematiktarif nutzen. Da insbesondere im Bereich der Krankenversicherungen Solidaritätsziele eine Rolle spielen, müsste sichergestellt werden,

dass niedrigere Telematiktarife, die nur durch die Entscheidung von überdurchschnittlich gesunden Versicherten für diese Tarife entstehen und nicht die Ausgaben der Versicherer wesentlich reduzieren, nicht zulässig wären.

5. Score-Qualität gewährleisten

1. Der SVRV empfiehlt, ein anspruchsvolles Qualitätsleitbild auf der Basis von Beispielen guter Praxis (Best Practices) zu entwickeln. An bestehende Initiativen zur Qualitätssicherung algorithmischer Prozesse soll angeknüpft werden. Dieses Qualitätsleitbild soll in Zusammenarbeit zwischen Wirtschaft, den Aufsichtsbehörden, den Marktwächtern der Verbraucherzentralen und den Verbraucherorganisationen erarbeitet und aktualisiert werden (Ausarbeitung, Implementierung, Monitoring).
2. Scoring-Anbieter in den sensiblen Bereichen sollten verpflichtet werden, gegenüber den Aufsichtsbehörden darzulegen und im Detail überprüfbar zu machen, wie gut das Verfahren tatsächlich ist. Nur so ist es möglich, die Score-Verfahren auf ihre Verbrauchergerechtigkeit hin zu testen. Diese Verpflichtung soll für Scores, die Verhalten prognostizieren, mittels statistischer Gütemaße (z. B. Falsch-Positiv Rate, Hit-Rate, Gini-Koeffizient, „Area under the ROC“) für die Gesamtpopulation und auch für relevante Teilpopulationen (nach Geschlecht, Alter, Bildung usw.) gelten. Dadurch können auch Diskriminierung erkennbar und fragwürdige Score-Qualitäten sichtbar gemacht werden.
3. Die nicht-adäquate Kennzeichnung des Ziels eines Scoring-Verfahrens ist bereits nach bestehender Rechtslage unzulässig. Zur Aufdeckung falsch gekennzeichnete Scores können neben Aufsichtsbehörden (siehe auch Empfehlung 7) aufgrund ihrer Expertise auch Verbraucherorganisationen oder die Marktwächter der Verbraucherzentralen beitragen.

4. Alle Stellvertreter-Merkmale, wie sie zum Beispiel beim Geo-Scoring verwendet werden, bedürfen einer besonderen Rechtfertigung (Stichwort Kausalzusammenhang) und Kontrolle durch die zuständige Aufsichtsbehörde. Die Verwendung von Proxy-Variablen sollte minimiert werden. Wenn sie eingesetzt werden, sollte ihr Zusammenhang zum Zielmerkmal inhaltlich plausibel begründet sein.

6. Datenqualität sichern

1. Eine angemessene Datenqualität bei der Score-Entwicklung ist zu gewährleisten und den Aufsichtsbehörden zu dokumentieren.
2. Scoring-Anbieter und -Nutzer sollten sich selbst verpflichten, ihre Data Governance und speziell ihr Datenqualitätsmanagement nach Maßgabe des Qualitätsleitbildes zu verbessern.
3. In der Anwendung der Verfahren ist die Korrektheit, Vollständigkeit und Aktualität der Daten sicherzustellen.
4. Zur Überprüfung der Datenbasis durch die Verbraucher selbst hat der SVRV bereits in seinem Gutachten „Digitale Souveränität“ die Option eines Daten-Dashboards skizziert, um ein Verbraucher-zentriertes Datenmanagement zu ermöglichen. Der SVRV bekräftigt seine Empfehlung, eine solche Option zu prüfen. Dazu sollten aktuelle Entwicklungen für ein sicheres Identitätsmanagement über Blockchain-basierte Systeme einbezogen werden, die dem Verbraucher die Möglichkeit geben, sicher und eindeutig die eigenen Identitätsdaten zu verwalten.
5. Der SVRV empfiehlt die Qualität der in relevanten Scoring-Verfahren eingesetzten Daten, insbesondere auch die Entity Recognition, zügig zu erforschen und ggf. zu verbessern, falls nötig durch gesetzliche Vorgaben. Es ist sicherzustellen, dass der beim Anbieter für eine bestimmte Person ermittelte Score tatsächlich der korrekten Person zugeordnet wird. Der Minimierung von Identitäts-

verwechslungen dient die Verpflichtung, dass Anbieter die betreffende Person über die Tatsache informieren, dass sie gescort wird (siehe Handlungsempfehlung 1).

Hier ergibt sich offenkundig ein Konflikt zwischen dem Interesse von Scoring-Anbietern und Nutzern und dem Datenschutz. Der SVRV empfiehlt deswegen, dass die Datenethikkommission der Bundesregierung die Möglichkeiten einer verbesserten Entity Recognition diskutiert und konkrete Empfehlungen erarbeitet.

7. Aufsicht verbessern

1. Der SVRV empfiehlt der Bundesregierung zu prüfen, wie eine Digitalagentur (siehe auch SVRV-Gutachten „Verbraucherrecht 2.0“) als Kompetenzzentrum agieren kann, das den einzelnen Aufsichtsbehörden hilft, ihre Aufgaben wahrzunehmen. Hierbei wäre zum Beispiel die Form eines Bundesinstituts vorstellbar, das methodische Kompetenzen zur Qualitätssicherung bereithält, die auch für nicht-digitale Zwecke eingesetzt werden könnten.
2. Die zuständigen Aufsichtsbehörden sind strukturell und teilweise durch Verbesserungen der Entgelte von Spezialisten, vor allem im Statistik- und Informatik-Bereich, in die Lage zu versetzen, den beschriebenen Aufgaben nachgehen zu können. Die Entwicklung der BaFin (Bundesanstalt für Finanzdienstleistungsaufsicht) in den letzten Jahren kann als Vorbild dienen. Den zuständigen Aufsichtsbehörden sind die für die oben genannten zusätzlichen Aufgaben sowie für das Testen konkreter Scorings notwendigen erheblichen finanziellen Mittel zur Verfügung zu stellen.
3. Zur raschen Umsetzung der vorliegenden Empfehlungen schlägt der SVRV vor, unverzüglich nach Abschluss der Datenethikkommission auf der Ebene der Bundesregierung eine Task Force (etwa beim Bundeskanzleramt) einzurichten, die auf der Basis bereits existierender Verfahren (z. B. der BaFin)

Leitlinien für die Ausarbeitung eines Qualitätsleitbilds vorlegt.

8. Super-Scores verhindern

Der SVRV empfiehlt, die Entwicklung in China und in anderen Ländern, wo mit Super-Scores experimentiert wird, sorgfältig zu verfolgen und zu analysieren. Insbesondere ist ein öffentlicher Diskurs über die sich damit verändernden gesellschaftlichen Werte und Gestaltungsoptionen notwendig.

Die Entwicklung von Super-Scores durch internationale kommerzielle Anbieter kann auch in Deutschland relevant werden. Gesetzgeber und Aufsichtsbehörden sollten sich darauf vorbereiten zu prüfen, ob Maßnahmen ergriffen werden sollten und können, damit Super-Scores nicht auch in Deutschland kommerziell angeboten werden.

Der SVRV empfiehlt, konkret zu prüfen, inwieweit bestehende Instrumente der DSGVO (insbesondere Zweckbindungsgrundsatz und Koppelungsverbot) auch zur Verhinderung von Super-Scores einsetzbar sind.

Mitglieder und Mitarbeitende des SVRV

Mitglieder des SVRV

Prof. Dr. Lucia Reisch (Vorsitzende)

Professorin für Interkulturelle Konsumforschung und europäische Verbraucherpolitik an der Copenhagen Business School

Dr. Daniela Büchel (stellv. Vorsitzende)

Bereichsvorstand Handel Deutschland REWE Group, Geschäftsführerin REWE Markt GmbH und Penny Markt GmbH

Prof. Dr. Gerd Gigerenzer

Direktor des Harding-Zentrums für Risikokompetenz am Max-Planck-Institut für Bildungsforschung in Berlin

Helga Zander-Hayat

Mitglied der Geschäftsleitung der Verbraucherzentrale Nordrhein-Westfalen

Prof. Dr. Gesche Joost

Professorin für das Fachgebiet Designforschung an der Universität der Künste

Prof. Dr. Hans-Wolfgang Micklitz

Professor für Wirtschaftsrecht am Europäischen Hochschulinstitut in Florenz

Prof. Dr. Andreas Oehler

Professor für Finanzwirtschaft an der Universität Bamberg und Direktor der Forschungsstelle Verbraucherfinanzen und Verbraucherbildung

Prof. Dr. Kirsten Schlegel-Matthies

Professorin für Haushaltswissenschaft an der Universität Paderborn

Prof. Dr. Dr. h.c. Gert G. Wagner

Max Planck Fellow am MPI für Bildungsforschung (MPIB) Berlin, Research Associate beim Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG) Berlin, und Senior Research Fellow bei der Längsschnittstudie Sozio-oekonomisches Panel (SOEP) am DIW Berlin

Mitarbeitende des SVRV

Leiter der Geschäftsstelle:

Thomas Fischer, M.A.

Wissenschaftlicher Stab der Geschäftsstelle:

Johannes Gerberding

Dr. Christian Groß

Dr. Ariane Keitel

Sarah Sommer, M.A.

Inhaltsverzeichnis

A

Zu diesem Gutachten 13

I. Einleitung	14
II. Scores und Scoring	16
III. Ziele des Gutachtens	20
Ziel 1: Informationslage und Wissen über Scoring verbessern	20
Ziel 2: Empirische Basis erweitern und rechtliche Fragen adressieren	21
Ziel 3: Regeln für verbrauchergerechtes Scoring anregen	21

B

Handlungsfelder: Stand der Forschung 25

I. Transparenz und Verständlichkeit	26
1. Transparenz bei prognostischem Scoring	26
2. Transparenz bei verhaltenssteuerndem Scoring	27
3. Sicherung von Transparenz und Verständlichkeit des Scorings als Dauerthema	28
4. Scoring-Transparenz als Spezialfall von Algorithmen-Transparenz	30
5. Transparenz als Bedingung einer gesellschaftlichen Debatte über Scoring	32
II. Diskriminierungsschutz und Gleichbehandlung	34
1. Was ist Diskriminierung?	34
2. Diskriminierung durch Scoring-Input	35
3. Score-Güte und Diskriminierungsverbot	36
4. Unerwünschte Ungleichbehandlungen durch Scoring jenseits der Diskriminierung	39
III. Rechtsdurchsetzung	40
IV. Score-Qualität	41
1. Güte des dem Score zugrundeliegenden Algorithmus	41
2. Der Nutzen neuerer und komplexerer Algorithmen	45
V. Datengrundlage	46
1. Korrektheit, Aktualität und Vollständigkeit	46
2. Verwendung von Proxy-Variablen	47
3. Gewichtung von Merkmalen	48
VI. Konkurrenz zwischen Fairness-Kriterien	50

VII. Verbraucher und Gesellschaft: Erwartungen, Wissen, Kompetenzen und Implikationen	52
1. Erwartungen und Akzeptanz von Verbrauchern im Zusammenhang mit Scoring	52
2. Wissen und Kompetenzen	54
3. Gesellschaftliche Implikationen	57
VIII. Die Gefahr des Super-Scores	61
1. Scoring-Geschäftsmodelle im Ausland	61
2. Datenakkumulation und Datenhandel	65
3. De-Anonymisierung von Daten	68
4. Verknüpfung von Daten zum Super-Score	69



Marktbefragung: Auskunfteien, Kfz-Telematik und Krankenversicherungen

71

I. Einleitung und zentrale Fragestellungen	72
II. Erhebungsdesign	73
1. Anbieterübersicht	74
2. Fragebogen	75
III. Diskussion der Ergebnisse und der sich daraus ergebenden Verbraucherprobleme	76
1. Verbreitung von Scoring in den betrachteten Markt Bereichen	76
2. Transparenz	78
3. Score-Berechnung und statistische Güte	80
4. Verhaltenseffekte	84
5. Diskriminierung	85
6. Verknüpfung von Daten, Einbeziehung neuer Verbrauchermerkmale	87
7. Aufsicht	88

D

Wissen über und Wertungen von Scoring in der Bevölkerung 91

- | | |
|--|-----|
| I. Vorstudie 2017 | 92 |
| II. Repräsentativbefragung 2018 | 93 |
| 1. Analyse der Ergebnisse | 94 |
| 2. Multivariate Regressionsanalysen: Darstellung und Diskussion der Ergebnisse | 106 |
| 3. Gesamtzusammenfassung und Fazit der Ergebnisse der Bevölkerungsbefragung | 109 |

E

Der Rechtsrahmen des Scorings 111

- | | |
|---|-----|
| I. Datenschutzrechtliche Grundlagen | 113 |
| 1. Profiling (Art. 4 Nr. 4 DSGVO) | 113 |
| 2. Automatisierte Einzelfallentscheidung (Art. 22 DSGVO) | 115 |
| 3. Scoring von Wahrscheinlichkeitswerten (§ 31 BDSG) | 118 |
| II. Bereichsspezifische Regelungen | 124 |
| 1. Recht der Allgemeinen Geschäftsbedingungen | 124 |
| 2. Versicherungsvertragsrecht und Versicherungsaufsichtsrecht | 125 |
| 3. Sozialversicherungsrecht der gesetzlichen Krankenversicherung | 128 |
| III. Bausteine eines Regulierungsrechts des Scorings | 129 |
| 1. Regulierung des „Wie“ versus Regulierung des „Ob“ des Scorings | 129 |
| 2. Scoring-Regulierung und Algorithmen-Regulierung | 130 |
| 3. Gewährleistung definierter Score-Güte | 130 |
| 4. Gewährleistung von Transparenz und Verständlichkeit | 132 |
| 5. Gewährleistung des Diskriminierungsverbots | 135 |
| IV. Aufsicht | 138 |

F

Handlungsempfehlungen 139

Anhang	149
I. Liste der Befragungsteilnehmer (Marktbefragung)	150
II. Fragebögen (Marktbefragung)	154
III. Darstellung der Befragungsergebnisse (Marktbefragung)	184
IV. Tabellenband (Marktbefragung)	203
V. Detaillierte Darstellung der Methode und Ergebnisse (Vorstudie und Bevölkerungsbefragung)	248
 Literatur	 273



Zu diesem
Gutachten



I. Einleitung

Das vorliegende Gutachten setzt sich unter dem Stichwort „Scoring“ mit algorithmischen Entscheidungsverfahren mit direktem Verbraucherkontakt¹ auseinander. Es knüpft damit unmittelbar an die Diskussionen der vorangegangenen Gutachten des Sachverständigenrats für Verbraucherfragen – insbesondere „Verbraucherrecht 2.0 (SVRV, 2016) und „Digitale Souveränität (SVRV, 2017a) – an. Dem hier gewählten Thema „Verbraucher-Scoring“ kommt aufgrund seiner Aktualität und zunehmender Verbreitung eine wachsende Bedeutung zu (z. B. Christl & Spiekermann 2016; Mau, 2017). Denn in immer mehr Lebensbereichen werden mit immer komplexeren Verfahren Eigenschaften und Aktivitäten von Verbrauchern analysiert, Prognosen über ihr künftiges Verhalten erstellt, oder sie werden zu bestimmten Score-Wert-verbessernde Verhaltensweisen motiviert. Der Output dieser Analyse ist ein individueller Zahlenwert (der Score), der die Grundlage dafür bieten kann,

- ob und zu welchen Konditionen ein Verbraucher beispielsweise einen Immobilienkredit erhält,
- welche Rückerstattung ein Verbraucher bei guter Fahrweise aus seiner Kfz-Versicherung vergütet bekommt und,
- ob jemand hinreichend viel Prävention für seinen Gesundheitszustand betreibt, sodass er einen Bonus von seiner Krankenversicherung bekommt und vieles mehr.

Dies sind Beispiele aus den drei wichtigen Lebens- und Konsumbereichen Finanzen, Mobilität und Gesundheit, in denen heute Scoring eingesetzt wird. Diese drei Bereiche wurden für das vorliegende Gutachten ausgewählt.

Scoring – vor allem das Bonitäts-Scoring – erfüllt in einer Marktwirtschaft für beide Marktseiten eine wichtige Transparenz- und Vertrauensfunktion, und neue Score-basierte Versicherungsprodukte haben durchaus Vorteile für Verbraucher. Neben solchen positiven Effekten kann Scoring aber auch negative, nicht-intendierte Effekte haben. Während sich der SVRV der Potenziale des modernen Scorings völlig bewusst ist, liegt der Fokus dieses Gutachtens auf möglichen Risiken sowie auf Handlungsoptionen, um diese Risiken zu minimieren. Konkret gehen wir der Frage nach, wie verbrauchergerichtetes Scoring – das es zu definieren gilt – prozedural und inhaltlich aussehen kann, welche Anforderungen aus Sicht der Verbraucherpolitik erfüllt sein müssen, und wie ein solches verbrauchergerichtetes Scoring politisch und institutionell unterstützt werden kann. Diese Überlegungen haben unmittelbare Relevanz für die Regulierung algorithmischer Entscheidungen im Allgemeinen, ebenso für die gesellschaftliche Bewertung und Regulierung künstlicher Intelligenz² und für die Daten-Ethik³.

Die formalisierte Einschätzung von Personen mit Hilfe einer Zahl (Scoring) hat in unserer Kultur eine gewisse Tradition, man denke nur an Schul- und Examensnoten. Die Digitalisierung vervielfacht nun die Möglichkeiten, Personen einzuschätzen, und zugleich die Risiken als auch die Chancen, da formalisiertes Scoring weniger dis-

1 Beispiele für weitere algorithmische Entscheidungsverfahren, die jedoch nicht Gegenstand dieses Gutachtens sind:

- Mit Verbraucherrelevanz z. B.: 1. Personalisierte Gutscheine in Supermärkten und Mikro-Targeting von Online-Shops. Dieses basiert auf Informationen über das Gerät mit dem ein Internet-Nutzer surft sowie wie dem Browserverlauf, um personalisierte Werbung anzuzeigen und personalisierte Preise anzubieten, um Nutzer zum Kauf zu bewegen, die Zahlungsbereitschaft von Kunden auszuschöpfen und Kunden über Vergünstigungen an sich zu binden (Hosell & Schleusener, 2016; Zander-Hayat, Domurath & Gross, 2016; Zander-Hayat, Reisch & Steffen, 2016); 2. Robo Adviser, die bei der Auswahl von Finanzprodukten helfen (Oehler, Horn & Wendt, 2016); 3. Algorithmen-gesteuerte, autonom fahrende Autos. sowie andere, weitgehend autonom agierende Produkte (Putzroboter, Robo-Rasenmäher).
- Ohne direkte Verbraucherrelevanz z. B.: 1. People Analytics (Personalmanagement) (vgl. z. B. aus gewerkschaftlicher Sicht Höller & Wede, 2018), darunter Bewerber-Scoring (sogenanntes Pre-Employment Screening und E-Recruiting; z. B. Christl, 2017); 2. Predictive Policing (z. B. Egbert, 2018; Sommerer, 2017).

2 Zu diesem Problembereich hat der Bundestag am 28. Juni 2018 eine Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche Potenziale“ eingesetzt. Auf der Homepage des Bundestages heißt es dazu: „Die Kommission hat den Auftrag, Handlungsempfehlung im Umgang mit Künstlicher Intelligenz (KI) zu formulieren. Sie soll unverzüglich eingesetzt werden und nach der parlamentarischen Sommerpause 2020 ihren Abschlussbericht mitsamt Handlungsempfehlungen vorlegen.“ Abgerufen am 17. August 2018 von URL <https://www.bundestag.de/dokumente/textarchiv/2018/kw26-de-enquete-kommission-kuenstliche-intelligenz/560330>.

3 Zu diesem Problembereich hat die Bundesregierung eine „Datenethikkommission“ eingesetzt. Auf der Homepage des BMI heißt es dazu: „Der Einsatz von Algorithmen, Künstlicher Intelligenz und digitalen Innovationen birgt große Potenziale. Gleichzeitig stellen sich zahlreiche ethische und rechtliche Fragen. (...) Die Datenethikkommission soll auf der Basis wissenschaftlicher und technischer Expertise ethische Leitlinien für den Schutz des Einzelnen, die Wahrung des gesellschaftlichen Zusammenlebens und die Sicherung des Wohlstands im Informationszeitalter entwickeln. Sie wird der Bundesregierung unter Federführung des Bundesministeriums des Innern, für Bau und Heimat und des Bundesministeriums der Justiz und für Verbraucherschutz bis Sommer 2019 Handlungsempfehlungen geben und Regulierungsmöglichkeiten vorschlagen.“ Abgerufen am 12. August 2018 von URL <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/datenethikkommission-node.html>.

kriminierend sein kann als informelle Entscheidungen durch Einzelne, etwa durch Vermieter oder Arbeitgeber. Zahlreiche maschinelle Handlungen sind eingebettet in komplexe „Entscheidungsarchitekturen“, an denen sowohl menschliche Entscheider als auch Maschinen beteiligt sind. Maschinen nehmen Priorisierungen vor, sortieren und klassifizieren und lenken so die Aufmerksamkeit menschlicher Entscheider. Sie präformieren die Räume, in denen sich menschliche Handlungsautonomie verwirklicht, und strukturieren menschliche Entscheidungsroutinen vor. Menschen treffen Entscheidungen üblicherweise gleichsam nicht vom leeren Blatt aus, sondern fügen mit der Entscheidung einem komplexen sozialen Gewebe einen weiteren Faden hinzu.

Das Potenzial des modernen Scorings kann nur ausgeschöpft werden, wenn eine Reihe von Bedingungen erfüllt sind, die ein verbrauchergerechtes Scoring weitestgehend sicherstellen. So müssen Verwechslungen von Personen ausgeschlossen bzw. minimiert werden – und Widerspruch muss auf einfache Weise effektiv möglich sein. Es darf weder direkt noch indirekt auf Basis von geschützten Merkmalen (wie etwa dem Geschlecht) ungerechtfertigt diskriminiert werden. Verbrauchergerechtes Scoring zur Vorhersage muss belegen, wie gut die Qualität der verwendeten Merkmale sowie die Vorhersagen selbst sind. Solch prädiktive Kraft sollte zudem über verschiedene sozio-ökonomische Gruppen hinweg stabil sein.

Nicht nur hinsichtlich der tatsächlichen Fairness des gewählten Verfahrens sind Transparenz und Verständlichkeit das A und O, wie sie zum Beispiel bei der Studienplatzvergabe aufgrund von Schulnoten gegeben sind. Scorings dürfen die Betroffenen nicht in die Irre führen, indem sie etwa nicht haltbare Versprechen hinsichtlich der Gesundheit machen und Menschen auf Grundlage (reiner) Gruppenzugehörigkeit bewertet werden (weil sie z. B. in einer bestimmten Gegend leben), sondern auch zuverlässige Vorhersagen treffen (etwa eine gewisse Bonität eines Verbrauchers). Sie sollten zudem nicht leichtfertig Schlüsse auf andere Lebensbereiche erlauben (wie dies bei Formen des chinesischen Bürgerscorings systematisch der Fall ist).

Scoring definiert

Scoring ist die Zuordnung eines Zahlenwertes (des Scores) zu einem Menschen zum Zweck der Verhaltensprognose oder Verhaltenssteuerung. Die Bestimmung dieses Zahlenwertes erfolgt in der Regel auf der Grundlage einer breiten Datenbasis durch ein algorithmisches Verfahren („Computerprogramm“).⁴

Vor allem aber muss Scoring für die Betroffenen zu verstehen sein und das ist nicht nur durch Transparenz zu erreichen. Darüber hinaus sollten Verbraucher über die Funktion sowie die Qualitäts- und Diskriminierungspotenziale von Scoring aufgeklärt werden und es sollten kritische Kompetenzen für einen Diskurs gefördert werden. Sind diese Bedingungen erfüllt, auf die wir in unseren Empfehlungen differenziert eingehen, und wird deren Erfüllung systematisch und effektiv vom Staat geprüft und überwacht, spricht aus Verbrauchersicht nichts gegen die Weiterentwicklung digitaler Scoring-Verfahren.

Aufbauend auf dieser Diskussion schlagen wir acht Handlungsempfehlungen vor, deren Umsetzung im Rahmen einer vorausschauenden Verbraucherpolitik in einer digitalisierten Lebens- und Konsumwelt wünschenswert wäre. Sie sind dem Gutachten vorangestellt und werden in Kapitel F im Hinblick auf konkrete Problemlagen begründet. Unsere Handlungsempfehlungen dürften auch für andere Scoring-Anwendungsfelder, wie etwa People Analytics, also indikatorengestütztes Personalmanagement, prüfenswert sein.

⁴ Anders das sogenannte Profiling: Profiling beschreibt den der Bewertung durch ein Scoring-Verfahren vorgelagerten Prozess der Datensammlung und -verknüpfung sowie Mustererkennung über eine Vielzahl an Menschen und den zu diesen Menschen zugehörigen Daten hinweg. Profiling kann also u. a. zum Zwecke eines anschließenden Scorings durchgeführt werden. Zum Profiling als Rechtsbegriff (Art. 4 Nr. 4 DSGVO) s. u. E.I.1.

II. Scores und Scoring

Zu den bekanntesten Scores gehören Bonitäts-Scores, die von privatwirtschaftlichen Auskunftsteilen einzelnen Verbrauchern zugeordnet werden. Auskunftsteile führen eine Vielzahl von Verbraucherdaten zusammen, wie z. B. Informationen über die Kredithistorie eines Verbrauchers, sogenannte Zahlungsstörungen sowie Informationen über die Person selbst. Einige Auskunftsteile beziehen darüber hinaus Daten über das Wohnumfeld (Geo-Scoring) mit ein.⁵ Daraus wird eine Wahrscheinlichkeitsaussage – oder auch: Verhaltensprognose – über die individuelle Kreditwürdigkeit bzw. Rückzahlungswahrscheinlichkeit eines Kredits abgeleitet (z. B. Schröder, Lang, Lerbs & Radev, 2014; Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Die Bonität wird in einem Score abgebildet, der beispielsweise Banken als Entscheidungshilfe bei der Kreditvergabe dient oder Onlinehändlern bei Entscheidungen über Zahlungsweisen, die dem jeweiligen Kunden angeboten werden: Je nach Höhe des Scores, also dem Ausmaß der Kreditwürdigkeit eines (potenziellen) Kunden, bieten Online-Shops Kunden komfortable – aber für den Online-Shop risikobehaftete – Zahlungsweisen wie „Kauf auf Rechnung“ an und schließen andere (potenzielle) Kunden davon aus. Im Online-Handel werden Scores in erster Linie zur Betrugserkennung eingesetzt, um notorisch säumige Zahler von den Kunden ohne Zahlungsstörungen zu unterscheiden. Letztere müssen sonst die durch erstere entstandenen Kosten mitzahlen, die sich trotz Betrugserkennung laut Auskunft des Branchenverbandes bevh (Bundesverband E-Commerce und Versandhandel Deutschland) allein im Jahr 2017 auf mehr als 2,5 Mrd. Euro belaufen.⁶

In der verbraucherpolitischen Debatte wird über Scoring bisher überwiegend im Kontext der Bonitätsprüfung gesprochen (z. B. Bala & Schuldzinski, 2017; Oehler, 2017; Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Vor dem Hintergrund der voranschreitenden digitalen Entwicklung und dem Einsatz von algorithmischen Ent-

scheidungsverfahren in weiteren Geschäftsfeldern, wie etwa Kranken- und Kfz-Haftpflichtversicherungen, sollte der Fokus der verbraucherpolitischen Diskussion heute jedoch auf den neuartigen Formen des Scorings liegen.⁷ Vor allem durch die Fortschritte beim Einsatz der sogenannten schwachen (narrow) Künstlichen Intelligenz (Ashley, 2017; Goodfellow, Bengio & Courville, 2016; Jentzsch, 2018; Nilsson, 2009; Russell & Norvig, 2010; Witten, Frank, Hall & Pal, 2016) sind im Bereich des Machine Learning neue Möglichkeiten der automatisierten Datenauswertung durch Mustererkennung entstanden.

Auf diesen neuartigen, kaum genutzten, aber in ihrer Wirkung noch wenig bekannten und weitgehend von Regulierung unbeachteten Formen des Scorings liegt ein Schwerpunkt dieses Gutachtens. Ein hier ausführlich untersuchtes Beispiel sind die sogenannten Telematik-Tarife (auch Pay-As-You-Drive-Tarife genannt), die bereits heute integraler Bestandteil einiger Versicherungsprodukte im Bereich Mobilität sind: Hierbei zeichnen Versicherungskonzerne (bzw. deren Vertragspartner für Datenauswertungen) beispielsweise mittels Smartphone-App das Fahrverhalten von Versicherten, einschließlich Daten über Zeit und Ort der Fahrt, auf und teilen Fahrer und Versicherung in einem Score-Wert mit, wie sicher gefahren wird. Besonders „gute“ Fahrer, d. h. solche Fahrer, die einen bestimmten Score-Wert überschritten haben, erhalten dann einen Rabatt auf ihren Versicherungstarif. Die Scores dienen hier nicht mehr nur der Prognose, sondern können von den Versicherten auch gezielt zur Veränderung bzw. Verbesserung ihres Verhaltens genutzt werden. Aktuell gibt es in der Kfz-Versicherung keine reinen Telematik-Tarife, bei denen sich der zu zahlende Tarif ausschließlich etwa aus der Höhe des Scores ableiten würde. Stattdessen werden diese immer als Ergänzung zu einem Kfz-Versicherungsvertrag mit konventioneller Prämienkalkulation angeboten.

Weitere Beispiele, die in diesem Gutachten differenziert dargelegt werden, finden sich im Bereich der gesundheitlichen Vorsorge: Viele gesetzliche Krankenversi-

⁵ Laut Eigenauskunft gilt dies im Falle der SCHUFA jedoch nur in „wenigen Ausnahmefällen“, und zwar dann, wenn keine anderen Informationen vorliegen. Abgerufen am 17. August 2018 von URL <https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring-schufa/>.

⁶ Abgerufen am 3. September 2018 von URL https://www.bevh.org/nc/veranstaltungen/details/datum/2017/oktober/artikel/betrug-im-e-commerce-und-datenschutz/fe_pw/?cHash=b38d2f5410c449712c2e5c2a4c0b0b1b&sword_list%5B0%5D=betrug.

⁷ Auch für Auskunftsteile gilt, dass „neueste technologische Verfahren wie etwa maschinelles Lernen“, etwa zur Betrugsabwehr, entwickelt werden. Abgerufen am 17. August 2018 von URL <https://schufa-wegbereiter.de/de/digital/innovationen-labor/innovationen-aus-dem-labor.jsp>.

cherungen belohnen ihre Kunden mit einer Gutschrift oder anderen Vergünstigungen, wenn sie im Rahmen eines sogenannten Bonusprogramms durch vorgeblich gesundheitsförderliche Aktivitäten Scoring-Punkte sammeln, indem sie an Präventionsmaßnahmen wie körperlicher Bewegung (aufgezeichnet beispielsweise durch einen Fitness-Tracker), Schutzimpfungen und Gesundheitskursen teilnehmen (Braun & Nürnberg, 2015). Nach einer Stellungnahme des Bundesversicherungsamts (BVA) ist die tatsächliche gesundheitsförderliche Wirkung der jeweiligen Aktivitäten indes kritisch zu hinterfragen (Bundesversicherungsamt, 2018). Den gesetzlichen Krankenkassen sind hier durch die Bestimmungen in § 20 Abs. 3 SGB V enge Grenzen bei der Auswahl von bonifizierbaren Aktivitäten gesetzt. Freier in der Gestaltung von Scoring-basierten Präventions- und Fitnessprogrammen sind sowohl private Krankenversicherungen als auch sonstige Anbieter: So plant das Versicherungsunternehmen Generali für das Jahr 2019, im Bereich der privaten Krankenversicherung das sogenannte Vitality-Programm anzubieten.⁸ Ein Teil des Programms basiert auf einem Gesundheits-Score, welcher sowohl die Teilnahme an Präventionsmaßnahmen als auch das Übermitteln von Vitaldaten mit Gutscheinen bei Partnerunternehmen und günstigeren Versicherungsbeiträgen, je nach erreichtem Score, belohnt: So erhält derjenige, der eine Nichtraucher-Erklärung unterzeichnet, 4.000 Punkte in einem Jahr; derjenige, der bei einem kooperierenden Online-Shop „gesunde Lebensmittel“ (Obst, Gemüse, Fisch) einkauft, erhält bis zu vier Mal so viele Punkte in einem Jahr.⁹

Unklar ist, anhand welcher Kriterien die Gewichtung dieser Einzelaktivitäten geschieht und inwieweit der Score tatsächlich Rückschlüsse auf eine Verbesserung des Gesundheitszustandes zulässt. Generali selbst positioniert sich klar und hebt die wissenschaftliche Fundierung des Vitality-Programms hervor („Vitality ist ein auf wissenschaftlichen Erkenntnissen basierendes Gesundheitsprogramm.“¹⁰) und wirbt vor diesem

Hintergrund auch mit gesundheitsförderlichen Effekten durch Teilnahme am Vitality-Programm („Vitality Mitglieder [haben] niedrigere Gesundheitskosten.“¹¹). Wirkungsstudien, die eine tatsächliche Verbesserung der Gesundheit durch Teilnahme am Vitality-Programm insbesondere für Deutschland belegen (inkl. randomisierter Kontrollgruppenvergleiche), stehen allerdings noch aus, was Kritiker dazu verleitet, Vitality als eine Art Payback-System mit Gesundheitsdaten zu bezeichnen.¹²

Zwar sind neben dem Vitality-Programm ähnliche Scoring-Verfahren im Gesundheitswesen heute noch wenig verbreitet, die zunehmende gesellschaftliche Akzeptanz von Selbstvermessung, u. a. durch sogenannte Wearables – diskutiert unter den Begriffen Quantified Self und Self-Tracking (z. B. Lupton, 2016; Selke, 2014) – legt jedoch nahe, dass sich dies in Zukunft ändern könnte. Anbieter knüpfen bereits daran, dass manche Menschen ihre Körperdaten zu Fitness-Zwecken aufzeichnen und auswerten. Befördert wird dies durch Aussagen wie die des Startup-Unternehmens Dacadoo, wonach die Benutzung von Apps, die den eigenen Gesundheitszustand in einem Zahlenwert ausdrücken, die Gesundheit verbessere.¹³ Eingebettet ist diese Entwicklung in den generellen Trend in der gesundheitlichen Vorsorge mit einer zunehmenden Betonung auf Früherkennung und präventiv gesundheitsförderliche Maßnahmen als Ergänzung zur kurativen Behandlung (z. B. GKV-Spitzenverband, 2017). Eine im Auftrag des Bundesministeriums für Gesundheit durchgeführte Studie kommt zu dem Ergebnis, dass bislang aussagekräftige Untersuchungen fehlen, um eine abschließende Bewertung von positiven Auswirkungen von Fitness-Apps auf die Gesundheit treffen zu können, insbesondere mit Blick auf die Langzeitwirkung der eventuell eingetretenen Effekte (Albrecht, 2016).

Regulatorische Maßnahmen wie das E-Health-Gesetz zum Aufbau einer modernen IT-Infrastruktur im Gesundheitswesen vom 1. Januar 2016 sowie die Locke-

8 Derzeit ist dieses aufgrund regulatorischer Anforderungen nur im Bereich der Erwerbs-/ Berufsunfähigkeitsversicherung und Risikolebensversicherung verfügbar.

9 Abgerufen am 4. September 2018 von URL <https://static01.cosmosdirekt.de/CosmosCAE/S/linkableblob/home/213750.1525232169000/data/Antragsinformationen-zur-Generali-Vitality-Mitgliedschaft-data.pdf>.

10 Abgerufen am 4. September 2018 von URL https://www.generalivitality.de/vmp/so_funktioniert_vitality.

11 Abgerufen am 4. September 2018 von URL https://www.generalivitality.de/vmp/so_funktioniert_vitality.

12 Abgerufen am 4. September 2018 von URL <https://www.sueddeutsche.de/wirtschaft/versicherung-wer-sich-bewegt-zahlt-weniger-1.2920176>.

13 Abgerufen am 19. Juni 2018 von URL <https://info.dacadoo.com/de/unternehmenslosungen/life-health-insurance-solutions/>.

zung des Fernbehandlungsverbots zur Ermöglichung von Telemedizin wie z. B. Videosprechstunden vom 10. Mai 2018 sind zudem Indizien für eine generell stärkere Nutzung von digitalen Dienstleistungen im Gesundheitswesen (siehe auch Gigerenzer, Schlegel-Matthies & Wagner, 2016). Dies ist einerseits ein Gewinn (etwa für diejenigen Verbraucher, die in ländlichen Gebieten mit niedriger (Fach-) Arztdichte wohnen); andererseits eröffnen diese Technologien eine Datenproblematik in bislang nicht gekanntem Ausmaß.

Ein Score, der Auskunft über die eigene Gesundheit gibt, kann durchaus als Mittel zur Stärkung von Verbrauchern betrachtet werden, um die Informationsasymmetrie zwischen Arzt und Patient abzubauen. Fraglich ist jedoch, ob Gleiches für die Informationsasymmetrie zwischen Verbrauchern und Unternehmen gilt: Hier positioniert sich der Deutsche Ethikrat dahingehend, dass diese durch den Einsatz von Big Data eher ausgebaut wird, da Unternehmen so potenziell mehr über ihre Kunden erfahren (Deutscher Ethikrat, 2017; cf. Weichert, 2018). Sowohl Forschung als auch Umsetzung stehen hier jedoch noch ziemlich am Anfang; dabei lohnt es sich, neben den Chancen auch frühzeitig auf – ggf. erst verzögert auftretende – Risiken hinzuweisen, die aufgrund der unmittelbaren Vorteile vom Verbraucher leicht übersehen werden können.

Im Gegensatz zu diesen noch relativ neuen Einsatzbereichen von Scoring wird die verbraucher- und marktpolitische Diskussion um die Rolle des Bonitäts-Scorings im Finanzwesen schon viele Jahrzehnte geführt. Aus der insgesamt guten Studienlage lässt sich kurz über die gesamtgesellschaftlichen Vorzüge des Bonitäts-Scorings zusammenfassen: Die Verwendung von Bonitäts-Scores reduziert Kreditausfälle; sie senken Transaktionskosten und sind daher für die Funktionsfähigkeit von Finanzmärkten wichtig (Schröder et al., 2014). Zudem können Bonitäts-Scores dazu beitragen, eine Informationsasymmetrie zwischen Kreditnehmern und Kreditgebern abzubauen und Kreditrationierung zu verhindern, weil die für eine Kreditvergabe entscheidenden Informationen über den potenziellen Kreditnehmer dem Kreditgeber vorliegen (Schröder et al., 2014). Im Bereich des Online-Shoppings tragen Bonitäts-Scores wesentlich

dazu bei, dass Online-Betrug erkannt und verhindert wird (Bolton & Hand, 2002; Marschall, Morawitzky, Reutter, Schwartz & Baars, 2015). Auf der anderen Seite gibt es berechtigte Sorgen, beispielsweise um die Datensouveränität oder die Diskriminierung bestimmter Gruppen, wie wir in früheren Gutachten ausgeführt haben (z. B. SVRV, 2017a).

Bei den neuen Telematik-Tarifen in der Kfz-Versicherung ist die Lage vergleichbar: Einerseits können sie zu mehr Sicherheit und besserem Verkehrsfluss führen, Informationsasymmetrien können reduziert und Märkte effizienter werden. Versicherungsanbieter werben damit, dass eine kontinuierliche Erfassung und Auswertung von Daten über Geschwindigkeit und Beschleunigung eine umsichtige Fahrweise motiviere und so zu mehr Sicherheit im Straßenverkehr beitrage.¹⁴ Die Auswertung der individuellen Fahrweise kann zudem eine risikogerechtere Bepreisung bzw. Rabattierung ermöglichen (Baecke & Bocca, 2017; Bian, Yang, Zhao & Liang, 2018), wovon besonders junge Autofahrer profitieren können, die sonst sehr hohe Prämien bezahlen müssen. Insgesamt, so wird argumentiert, kann eine (aktuariell) genauere Kostenkalkulation der Kfz-Versicherung realisiert werden (z. B. Baecke & Bocca, 2017; Bitkom, 2014; Kraft & Hering, 2017). Ein weiteres, gesamtgesellschaftlich wünschenswertes Potenzial einer insgesamt umsichtigeren Fahrweise durch Scoring besteht zudem in einer Verringerung von Staus sowie der Reduktion der Umweltverschmutzung (Kraft & Hering, 2017; Litman, 2005). Bedenken werden dahingehend geäußert, dass es aufgrund der permanenten Aufzeichnung und Auswertung der Autofahrt zu einer zunehmenden Überwachung durch kommerzielle Versicherungen kommen kann (z. B. Stiftung Warentest, 2014; Verbraucherzentrale Bayern, 2016). Nicht zuletzt wird häufiger die Kritik vorgebracht, dass Telematik-Angebote nicht nur den Kunden, sondern zuerst auch den Versicherungen selbst dienen: So wird die gesteigerte Möglichkeit der direkten Ansprache von Kunden bspw. durch Push-Benachrichtigungen auf das Smartphone als Mittel der Kundenbindung eingesetzt (z. B. Weidner & Transchel, 2015).

¹⁴ Abgerufen am 19. Juni 2018 von URL <https://www.cosmosdirekt.de/veroeffentlichungen/versicherungstipp-telematik-198254/>.

Um es deutlich zu sagen: Dieses Gutachten stellt nicht in Frage, dass Scoring in der Wirtschaft und Gesellschaft eine wichtige Funktion übernimmt. Dem SVRV geht es vielmehr um die Frage, wie Scoring ausgestaltet wird und werden sollte. Scoring-basierte Geschäftsmodelle sind heute Normalität¹⁵, wenn auch in unterschiedlicher Tiefe implementiert, und können einzelnen Verbrauchern und Märkten insgesamt viele Vorteile bringen. Natürlich bergen sie auch Risiken in sich, die teilweise schon deutlich, teilweise aber nur in Konturen erkennbar sind und aufgrund der rasanten technologischen Entwicklung auf keinen Fall abschließend bewertbar sind. Dieses Gutachten beleuchtet die zentralen Herausforderungen von auf Scoring basierenden Geschäftsmodellen und gibt Empfehlungen, welche verbraucherpolitischen Maßnahmen die Position der Verbraucher stärken können. Im Kern steht dabei die Vorstellung eines verbrauchergerechten Scorings, das es zu beschreiben (vgl. auch Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016) und zu diskutieren gilt:

- Was bedeutet Fairness bei Scoring-basierten Geschäftsmodellen?
- Welche Daten sollten in die Scores einfließen (und welche nicht)?
- Welchen statistischen Gütekriterien sollten Scores genügen?
- Welche Bewertungskriterien sind relevant für verbrauchergerechtes Scoring?
- Was bedeutet Diskriminierung durch Scoring und wo findet sie in welcher Weise statt?
- Welche Elemente von Scores sollten bekannt sein, transparent und verständlich gemacht werden (und welche nicht)?
- Welche Formen der Transparenz und des Monitorings sollte es geben, damit Scoring-Verfahren die Durchsetzung von Verbraucherinteressen stabilisieren bzw. stärken? Reichen die bisherigen Prozesse aus?
- Welche Institutionen bieten sich für die Schaffung von Transparenz und für ein Monitoring an?

Perspektivisch wird im vorliegenden Gutachten auch die Entwicklung von sogenannten Super-Scores betrachtet, also Scores, die sich in der Bewertung von Verbraucherverhalten nicht nur auf einen abgrenzbaren Lebensbereich wie Finanzen, Mobilität und Gesundheit beschränken, sondern übergreifend (Verbraucher-) Verhalten bewerten. Näher betrachtet werden chinesische Pilotprojekte für ein sogenanntes Sozialkreditsystem, mit dem alle Bürger der Volksrepublik ab 2020 einen individuellen Score erhalten sollen, in den Verhaltensweisen aus allen Lebensbereichen eingehen (Kostka, 2018). Obgleich dieses Modell angesichts der unterschiedlichen politischen und Rechtssysteme in dieser Form weder zurzeit noch auf absehbare Zeit auf die westliche Welt übertragbar ist, liefert es doch Diskussionsstoff über das technisch Machbare und gesellschaftlich (nicht) Vertretbare. Ähnlich hat sich hierzu Bundespräsident Dr. Frank-Walter Steinmeier beim Antrittsbesuch in Sachsen-Anhalt an der Leopoldina – Nationale Akademie der Wissenschaften – am 15. Februar 2018 in Halle (Saale) geäußert: “So etwas [wie in China] droht weiß Gott nicht in Deutschland – aber Sie sehen daran, wie wichtig es ist, dass wir gesellschaftliche Auswirkungen von technologischen Entwicklungen intensiv diskutieren.”¹⁶

¹⁵ Vgl. auch Fussnote 1. Hier werden Anwendungen von Scoring genannt, die im Gutachten nicht vertieft werden: Mikro-Targeting von Online-Shops, *robo adviser* für die Auswahl von Finanzprodukten, Bewerber-Scoring, *predictive policing* und insbesondere die individuelle Steuerung sozialer Medien durch deren Anbieter.

¹⁶ Abgerufen am 1. Oktober 2018 von URL <https://www.bundespraesident.de/SharedDocs/Reden/DE/Frank-Walter-Steinmeier/Reden/2018/02/180215-Leopoldina-Sachsen-Anhalt.html>.

III. Ziele des Gutachtens

Die besondere Relevanz des Themas Scoring für Verbraucher, Gesellschaft und Unternehmen ergibt sich aus drei wesentlichen Faktoren: Der zunehmenden Verfügbarkeit von individualisierten Verbraucherdaten, der Verbreitung von wirtschaftlich einsetzbaren Methoden zur Bewältigung dieser Datenmengen und Profilbildung sowie der dadurch wachsenden Anzahl an Anwendungsfällen von Scoring. Hinzu kommt, dass neuartige Scoring-Verfahren nicht mehr nur zur Prognose, sondern zunehmend auch zur Steuerung des Verhaltens von Verbrauchern eingesetzt werden. Konkret verfolgen wir mit dem vorliegenden Gutachten drei (Kern-) Ziele:

Ziel 1: Informationslage und Wissen über Scoring verbessern

Neu mit Blick auf die Datengrundlage ist, dass Daten über Verbraucher nicht mehr nur stichpunktartig abgefragt werden (z. B. Alter, laufende Kreditverträge), sondern weit detailliertere Werte wie Vitaldaten und Fahrweise mithilfe neuer Techniken erfasst werden können. Zudem entwickeln sich die Möglichkeiten der günstigen Speicherung von Daten immer weiter. Das Wachstum von Geschäftsmodellen im Bereich des Handels mit Verbraucherdaten von Unternehmen wie Acxiom und Oracle deutet zusätzlich darauf hin, dass grundsätzlich das Geschäftsinteresse sowie das Know-how vorhanden ist, um – falls dies rechtlich erlaubt würde – umfangreiche Datensätze über Verbraucher zusammen zu tragen, um möglicherweise weit komplexere Scores zu entwickeln, die nicht mehr nur auf einen Lebensbereich wie z. B. Finanzen beschränkt sind, sondern lebensbereichsübergreifend arbeiten. Damit können sie Entscheidungen nicht mehr nur über den Bürger als Verbraucher in einem abgegrenzten Bereich vorbereiten oder gar unmittelbar fällen, sondern die Teilhabe des Bürgers an Wirtschaft und Gesellschaft insgesamt betreffen.

Verfahren, die eine automatisierte Einschätzung darüber erlauben, ob beispielsweise eine Person ein hohes Kreditrisiko darstellt oder ein guter Autofahrer mit geringem Unfallrisiko ist, sind vielfältig: Sie reichen von

einfachen Daumenregeln (auch: Heuristiken) über statistische Schätzverfahren wie die logistische Regression (die in der Praxis des Bonitäts-Scorings Anwendung findet und eine vergleichsweise einfache Form des Machine Learning darstellt) bis hin zu neueren maschinellen und bislang für Scoring nur experimentell eingesetzten Deep Learning Verfahren (etwa auf Basis „neuronaler Netze“), welche weitgehend automatisiert Muster bzw. korrelative Zusammenhänge zwischen einer Vielzahl von (Verbraucher-) Merkmalen in großen Datensätzen verarbeiten.

Aus verbraucherpolitischer Sicht ist klar: Je intransparenter das Verfahren ist, desto schwerer ist für Aufsichtsbehörden die Überprüfung des Verfahrens und zugleich wird damit auch die Nachvollziehbarkeit für die Verbraucher in Frage gestellt. Ob durch den Einsatz maschinellen Lernens auch die Fairness von Scores abnehmen muss oder gesteigert werden kann, wird ausführlich diskutiert.

Dargestellt werden im vorliegenden Gutachten nicht nur der aktuelle Stand verwendeter Algorithmen im Kontext des Verbraucher-Scorings, sondern insbesondere auch Möglichkeiten der individuellen und institutionellen Überprüfung dieser Algorithmen sowie deren Regulierung. Besonderes Gewicht wird auf die Etablierung von Mindeststandards für die (Modell-) Güte von Scoring-Verfahren und Fragen der Verständlichkeit für Verbraucher gelegt.

Insgesamt ermöglicht die zunehmende Verfügbarkeit von Daten und die Verbreitung neuer Methoden die Entwicklung neuartiger Anwendungen von Scoring-Verfahren, die weit über das „klassische“ Bonitäts-Scoring hinausreichen. Exemplarisch betrachtet werden in dem vorliegenden Gutachten dafür die Anwendungsfelder Mobilität (Telematik in der Kfz-Versicherung) und Gesundheit (Punktesysteme bei den Bonusprogrammen der gesetzlichen Krankenversicherungen, erste Ansätze von Gesundheits-Scores bei privaten Krankenversicherungen sowie sonstige Anbieter von Gesundheits-Scores wie Dacadoo). Während das „klassische“ Bonitäts-Scoring auf die Prognose künftigen Verhaltens beschränkt ist, könnten neuartige Scoring-Verfahren, die Bewegungs- und z. T. Vitaldaten auswerten und den resultierenden Score „live“ gegenüber dem Verbraucher kommunizieren, viel stärker auch eine verhaltenssteuernde Wirkung entfalten. Die Frage, ob dies gesamtgesellschaftlich erwünscht ist oder nicht, ist zu diskutieren.

Ziel 2: Empirische Basis erweitern und rechtliche Fragen adressieren

Mit dem vorliegenden Gutachten kommt der SVRV seinem Auftrag nach, eine breite empirische Basis für eine wohlinformierte und zukunftsgerichtete Verbraucherpolitik zu schaffen. In diesem Gutachten steht das dynamische Feld der neuartigen Formen des Scorings im Fokus.

Im Anschluss an eine Darstellung des Stands der Forschung (siehe Kapitel B) legt der SVRV eine umfangreiche Studie der Marktsegmente Finanz-Bonität und Mobilität vor. In diesen Märkten sind Scoring-Methoden bereits seit langem etabliert bzw. haben in den letzten Jahren verstärkt Fuß gefasst (siehe Kapitel C). In dieser Studie wird zudem die Frage adressiert, inwieweit Scoring-Verfahren bereits heute im Segment der gesetzlichen und privaten Krankenversicherungen verankert sind (da dies nur eingeschränkt der Fall ist, verwenden wir hier den Begriff des „Proto-Scorings“). Hierfür wurden insgesamt drei Fragebögen entwickelt und an alle identifizierten Auskunfteien, Kfz-Versicherungen und gesetzlichen sowie privaten Krankenversicherungen geschickt. Die schriftlich beantworteten Fragebögen wurden digitalisiert, die Antworten zum Zweck der Vergleichbarkeit und Aggregation kodiert und analysiert. Schwerpunktmäßig wurde die Durchdringung dieser drei Marktbereiche mit Scoring-Verfahren untersucht, ferner die Frage danach, welche Verbrauchermerkmale üblicherweise erhoben und zur Berechnung der jeweiligen Scores verwendet werden und welche Gütekriterien der dem Scoring zugrundeliegende Algorithmus erfüllt. Zur Vertiefung der Diskussion wurden mit einzelnen Unternehmen und Sachkundigen in den drei Bereichen Hintergrundgespräche geführt.

Darüber hinaus hat der SVRV eine Bevölkerungsbefragung konzipiert und durchführen lassen, um mehr über das Wissen und über die Akzeptanz von etablierten sowie von technisch grundsätzlich realisierbaren, aber noch nicht etablierten Scoring-Verfahren in der Bevölkerung in Deutschland zu erfahren (siehe Kapitel D). In Zusammenarbeit mit dem Markt- und Sozialforschungsinstitut ifas wurde eine Repräsentativbefragung mit der CATI-Methode (Computer Assisted Telephone Interview)

durchgeführt. Die Stichprobe umfasst N = 2.215 Teilnehmer. Die Daten wurden anschließend vom SVRV aufbereitet und ausgewertet.

Neben den empirischen Studien adressiert eine rechtliche Untersuchung (siehe Kapitel E) im Detail die Fragen des Datenschutzes, bereitet die aktuellen Regelungen von Scoring- bzw. Score-ähnlichen Verfahren in den Marktsegmenten Bonität, Kfz und Gesundheit auf und ist um einen Entwurf von Bausteinen eines Regulierungsrechts für Scoring insgesamt ergänzt. Der Teil schließt mit Erwägungen zu Rechtsdurchsetzung und Aufsicht.

Ziel 3: Regeln für verbrauchergerechtes Scoring anregen

Insbesondere möchte der SVRV mit dem vorliegenden Gutachten Kriterien für ein verbrauchergerechtes Scoring zur Diskussion stellen. Scoring ist nach Auffassung des SVRV dann verbrauchergerecht, wenn Scores Verbrauchern verständlich präsentiert werden, Scoring-Wissen und Kompetenzen ausreichend vorhanden sind, Diskriminierung geprüft und offengelegt ist, Telematik-freie Optionen auch in Zukunft ohne substantielle Nachteile verfügbar sind, Score- und Daten-Qualität gewährleistet sind, die Aufsicht über Scoring materiell verbessert wird und Super-Scores wirksam verhindert werden.

Die Geschichte des Scorings

Historisch gesehen ist der Wunsch, Charakteristika, Verhalten und Vorlieben von Individuen möglichst genau zu bewerten und daraus zukünftige Entwicklungen abzuleiten, nichts Neues. Auch in der analogen Welt wurden und werden an bestimmte Merkmale und Verhaltensweisen von Personen Konsequenzen geknüpft sowie ggf. Menschen Zahlenwerte zugeschrieben. Durch die Digitalisierung mit komplexen Algorithmen und breiten Datengrundlagen hat das Thema Scoring heute lediglich eine neue Dynamik bekommen; und manche glauben, dass mit der Entwicklung der Künstlichen Intelligenz (KI) mit sogenannten selbstlernenden Algorithmen auch eine neue Dimension hinzugekommen sei.

Das Bewerten von Leistungen und Fähigkeiten mittels Zahlenwerten oder normierten Worturteilen hat eine lange Tradition. In deutschen Schulen beispielsweise werden Zensuren seit dem 16. Jahrhundert vergeben (Lintorf, 2012) und das nicht nur für Lernleistung, sondern auch – wie in China nun für alle Bürger geplant (Kostka, 2018) – Sozialverhalten. An individuelle Bewertungen von Leistungen und Lernerfolgen sind auch heute noch Konsequenzen wie die Zulassung zu höheren Klassenstufen oder weiterführenden Schularten sowie das Erreichen von Schulabschlüssen (z. B. Abitur) geknüpft. Als Zulassungskriterium zu Hochschulen ist neben anderen Kriterien die Abiturnote ein zentraler Wert. Die einzelnen Schulnoten eines Schülers werden für diese verrechnet, also eine Durchschnittsnote gebildet, wobei einige Noten (z. B. von Hauptfächern) mehr Gewichtung finden als andere. Im Zulassungsverfahren werden dann die Schüler mit den höchsten Abschlussnoten (meist in Kombination mit bisheriger Wartezeit auf einen Studienplatz) angenommen, bis die Kapazitäten erschöpft sind (Numerus clausus). In diesem Kontext wird besonders deutlich, dass in den Abschlussnoten nicht nur eine Bewertung vergangener Leistungen gesehen, sondern ihnen auch eine Vorhersagekraft in Bezug auf zukünftige Leistungen zugeschrieben wird. Eine gute Abiturnote soll signalisieren, dass auch während des Studiums gute Leistungen zu erwarten sind und ein Abschluss wahrscheinlich ist.

Ebenfalls ein Lebensbereich, der traditionell auf der Messung von Leistungen von Menschen beruht: Leistungssport. Aber nicht nur die Leistungen werden ver-

messend, sondern in vielen Disziplinen auch die Menschen selbst, um sie im Interesse möglichst spannender und „fairer“ Wettbewerbe z. B. in bestimmte Gewichtsklassen einzuteilen (etwa beim Boxen und Gewichtheben), ihnen Handicaps aufzuerlegen (beim Golfen) und um Setzlisten zu erstellen (etwa beim Tennis). Schließlich werden Sportler immer wieder anhand ihrer Scores charakterisiert, etwa der erste „Zehn-Komma-Null-Läufer“ Armin Hary¹⁷, der allerdings – präziser elektronisch gestoppt, als es die offizielle Handstoppung konnte – etwa 10,2 Sekunden brauchte (bei grenzwertig starkem Rückenwind) und damit auf eventuelle Messprobleme von Scores jeder Art verweist.

Gerade in Geschäftsbeziehungen, bei denen in hoher Frequenz Verträge mit bis dahin unbekanntem Partnern eingegangen werden und ein gewisser Vertrauensvorschub geleistet werden muss, spielt Risikominimierung eine große Rolle, weshalb in diesem Bereich Scoring seit Jahrzehnten eine hohe Bedeutung zukommt. Aus dem Bedarf, sich über die Zuverlässigkeit und Zahlungsfähigkeit von Geschäftskunden zu informieren, entstanden im 19. Jahrhundert die ersten Auskunfteien. Zu den ersten in Europa gehörten z. B. die Auskunfteien Wys Muller (gegründet 1861), Schimmelpfeng (gegründet 1872) oder Creditreform (gegründet 1879). Diese Unternehmen sammelten wirtschaftlich relevante Informationen zu Personen und Firmen und verkauften diese an Unternehmen und Banken. Seither bildeten Auskunfteien eine wesentliche Grundlage für ein funktionierendes Kreditwesen.

Die ersten Versuche, das Risiko eines Zahlungsausfalls von Personen quantitativ zu berechnen und numerisch darzustellen, wurden in den 1940er Jahren unternommen. Bis dahin bestanden rudimentäre Scoring-Systeme (etwa von Versandhäusern) in einem Kriterienkatalog, mit dessen Hilfe vor der Kreditvergabe eine Reihe von Voraussetzungen auf Erfüllung geprüft und zusammengezählt wurden (Thomas, Crook & Edelman, 2017). Der Mathematiker Davis Durand verwendete 1941 in einem Forschungsprojekt zum ersten Mal die Diskriminanzanalyse, um das Ausfallrisiko von Krediten zu bestimmen (Durand, 1941). Er analysierte Datensätze über bereits vergebene Kredite daraufhin, welches die entscheidenden Faktoren für eine problemlose Rückzahlung oder für

¹⁷ Abgerufen am 1. Oktober 2018 von URL https://de.wikipedia.org/wiki/Armin_Hary.

Rückzahlungsschwierigkeiten gewesen waren, und entwickelte einen Kreditscore. Das erste Unternehmen, das statistische Modelle für die Kreditvergabe kommerziell entwickelte, war *Fair, Isaac and Company* in Kalifornien (heute bekannt als FICO). Seit den 1950er Jahren vertrieb es Scoring-Produkte an Finanzhäuser, Einzelhändler und Versandhäuser (Dixon & Gellmann, 2014).

In den folgenden Jahrzehnten wurden mathematische Fortschritte von Innovationen im Bereich der elektronischen Datenverarbeitung begleitet, was schließlich weitgehend automatisiertes Bonitäts-Scoring ermöglichte. Die Kombination aus Computern und Algorithmen sowie die Erfahrungen von Unternehmen, dass mit Hilfe von Scoring die Ausfallraten ihrer Kredite sowie Betrugsdelikte deutlich reduziert werden konnten, führten zu den Scoring-Produkten der Auskunfteien wie wir sie heute kennen.

Ein weiterer Bereich, in dem solche Formen der Risikoeinschätzung schon lange etabliert sind, ist der Bereich der Versicherungen. Hier kam und kommt Scoring vor allem zum Einsatz, um die Versicherungs- bzw. Beitragssummen individuell für jeden Kunden zu berechnen. Schon in den 1920er und 1930er Jahren entstand in Deutschland das Interesse, die Beitragsberechnung in der Krankheitskostenversicherung auf mathematisch-statistische Grundlagen zu stellen. Mittels sogenannter Morbiditätstabellen wurde festgestellt, dass je nach Geschlecht, Alter und Beruf der Versicherten unterschiedliche Krankheitskosten zu erwarten waren (Wagner-Braun, 2002). Noch heute werden in der privaten Krankenversicherung die Beiträge beim Einstieg individuell berechnet. Auch in der Lebensrisiko- oder Berufsunfähigkeitsversicherung ist dies der Fall. Verbraucher werden anhand einer Kombination einzelner Merkmale (z. B. Alter oder Vorerkrankungen) eingestuft, das Risiko für den Versicherer damit abgewogen und die Beitragsätze werden entsprechend kalkuliert.

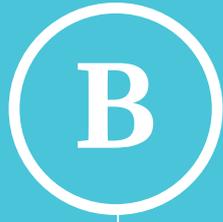
Eine besonders komplexe Prämienkalkulation findet in der Regel in der Kfz-Versicherung statt, in der die Tarife aufgrund einer Vielzahl an Merkmalen an den einzelnen Kunden angepasst werden. Wesentliche Faktoren sind u. a. die Typklasse des Automodells, die Regionalklasse und die Schadensfreiheitsklasse sowie Eigenschaften wie Zahl der Nutzer, Nutzer- und Fahrzeualter, Kilometerfahrleistung und Stellplatz (Gesamtverband der Deutschen Versicherungswirtschaft e.V., 2016). Und noch ein weite-

res Scoring-System kennen Autofahrer: das Fahreignungs-Bewertungssystem des Kraftfahrt-Bundesamtes (die sogenannten „Punkte in Flensburg“). Seit 1974 vermerkt das Amt in einem Register Punkte für Ordnungswidrigkeiten und Straftaten im Straßenverkehr und spricht bei entsprechend hohem Punktestand Ermahnungen aus, ordnet die Teilnahme an Fahreignungsseminaren an oder entzieht die Fahrerlaubnis (Kraftfahrt-Bundesamt, 2017).

Die beschriebenen Formen des Scorings haben eine lange Tradition und existierten ohne Zweifel auch schon in der analogen Welt. Unbestreitbar ist aber auch, dass sich die Durchführung des Scorings mit den technologischen Entwicklungen des digitalen Zeitalters stark verändert hat: So regelt in Frankreich die Studienplatzvergabe seit diesem Jahr ein Scoring-Algorithmus namens *Parcoursup*, der Erfüllung der Zulassungsvoraussetzungen, Wohnort und Präferenzen der Bewerber auswertet (Joeres, 2018). Beim Online-Handel kann die Bonität eines Verbrauchers automatisiert und sekundenschnell berechnet werden, um dann entsprechende Zahlungsmöglichkeiten anzubieten. Und in der Kfz-Versicherungsbranche halten Telematik-Tarife Einzug, in denen Fahrverhalten permanent evaluiert und Beiträge entsprechend des daraus gebildeten Scores angepasst werden.

Algorithmenbasiertes Scoring kommt zudem vermehrt in vielen neuen Bereichen zum Einsatz und bewertet Verbraucher und Verbrauchergruppen auf die vielfältigste Art und Weise und mit den unterschiedlichsten Konsequenzen (Dixon & Gellmann, 2014). Es gibt Scores, welche die Kaufkraft oder die Spendenbereitschaft von Haushalten vorhersagen (Equifax, 2018; Blackbaud, 2014), Scores, die anzeigen, ob Kunden zu anderen Firmen abwandern werden (Versium Analytics Inc., 2018), Scores, die das Ziel haben Schwangerschaften zu erkennen (Duhigg, 2012) oder Scores, die das Energieverhalten messen (Trove, 2018). Partnerbörsen basieren auf Scores, die die Passgenauigkeit von persönlichen Profilen quantifizieren (Carr, 2016).

Es entsteht eine Kultur des Bewertens und Quantifizierens (Mau, 2017). Ob Likes auf Facebook, Anzahl der Follower auf Twitter oder Sterne auf Airbnb – es sind längst nicht mehr nur allein die Algorithmen von Unternehmen, die Verbraucher bewerten und ihnen Zahlenwerte zuschreiben. Scoring ist zur Alltagshandlung geworden.



Handlungsfelder: Stand der Forschung



I. Transparenz und Verständlichkeit

Welches Transparenzniveau im Bereich des Scorings angemessen ist und wie ein solches Transparenzniveau gewährleistet werden kann, diskutieren die allgemeine Öffentlichkeit und die Wissenschaft als offene Frage. Transparenz bedeutet in diesem Zusammenhang die Offenlegung von Informationen gegenüber Verbrauchern durch die Ersteller oder Verwender von Scoring-Verfahren. Mit Überlegungen zu angemessenen Transparenzniveaus stets verbunden ist die Frage danach, wie die transparent gemachten Informationen aufzubereiten und zu strukturieren sind, um ihre tatsächliche Verständlichkeit zu gewährleisten. Hier kann einerseits der Verbraucher, dem eine informierte Teilnahme an Scoring-Verfahren ermöglicht werden soll, den Maßstab abgeben. Andererseits kann es um die Verständlichkeit für Fachleute gehen, denen die Möglichkeit einer kritischen Prüfung von Scoring-Verfahren eröffnet werden soll.

Das Bonitäts-Scoring hat bisher im Zentrum der Transparenz-Debatte gestanden, weil es sich bis in jüngste Zeit hinein um die technisch reifste und am weitesten verbreitete Form des Scorings gehandelt hat (zur Geschichte siehe Beckhusen, 2004). Für die Zukunft ist zu erwarten, dass sich die Diskussion über die Transparenz des Scorings mit der neuerdings intensiv geführten Debatte über die Transparenz von Verfahren algorithmischer Entscheidungsfindung verbinden wird. Der Aspekt der Scoring-Transparenz wird auch deshalb relevant bleiben, weil erst ein ausreichendes Transparenzniveau es Verbrauchern ermöglicht, weitergehende Rechte, zum Beispiel auf Korrektur eines fehlerhaften Score-Wertes, geltend zu machen. Schließlich ist Transparenz eine Bedingung dafür, dass über das Phänomen Scoring überhaupt eine informierte gesellschaftliche Debatte geführt werden kann.

1. Transparenz bei prognostischem Scoring

Scoring-Verfahren werden sowohl zur Prognose von Verhaltensweisen als auch zur Verhaltenssteuerung eingesetzt. Unter dem Gesichtspunkt der Transparenzgewährleistung muss zwischen diesen beiden

Verwendungsabsichten unterschieden werden. Denn Scoring-Verfahren, die auf eine Verhaltensprognose abzielen, wollen üblicherweise nicht reflexiv wirken, das heißt, sie wollen das in den Blick genommene Verhalten nicht selbst beeinflussen. Die Bestimmung eines angemessenen Transparenzniveaus steht damit im Widerstreit gegensätzlicher Interessen, deren Berechtigung nicht von der Hand zu weisen ist und die jeweils auch verfassungsrechtlich unterlegt sind.

Die Person, deren Verhalten Gegenstand einer prognostischen Einschätzung durch ein Scoring-Verfahren ist (im Folgenden auch als die „gescorte Person“ bezeichnet), wird normalerweise ein Interesse daran haben, zu erfahren, dass ein Scoring überhaupt stattfindet. Zweitens wird sie um die Konsequenzen des gebildeten Scores wissen wollen. Weiterhin kann sie daran interessiert sein, zu wissen, auf welcher Datengrundlage die Ermittlung des sie betreffenden Score-Wertes erfolgt, das heißt, welche ihrer Merkmale in die Score-Wert-Ermittlung eingehen. Schließlich kann sie daran interessiert sein, in die innere Funktionsweise des Scoring-Algorithmus Einblick zu bekommen, also insbesondere zu erfahren, welches relative Gewicht jedes erfasste persönliche Merkmal für die Ermittlung des Score-Wertes hat.

Diesen Interessen können auf der anderen Seite Geheimhaltungsinteressen desjenigen, der ein Scoring durchführt (im Folgenden auch als „Scorer“ bezeichnet), oder der Allgemeinheit gegenüberstehen. Ein Geheimhaltungsinteresse des Scorers liegt regelmäßig dann vor, wenn die Prognoseleistung des Scoring-Verfahrens dieses wirtschaftlich wertvoll und damit als Unternehmensgeheimnis schützenswert macht (anerkannt bei BGH, Urteil vom 28. Januar 2014 – VI ZR 156/13 –, BGHZ 200, 38; siehe auch C.III.3). Wenn bekannt wird, wie das algorithmische Verfahren der Score-Berechnung im Einzelnen ausgestaltet ist, verliert es seinen Charakter als Geschäftsgeheimnis und kann von Konkurrenten übernommen werden.

Ein Geheimhaltungsinteresse der Allgemeinheit kann daraus folgen, dass bei Offenlegung des Scoring-Verfahrens dessen Prognosekraft unter bestimmten Umständen (zu diesen sogleich) abnimmt. Dies kann gesellschaftlich unerwünscht sein. Ein Allgemeininteresse an, zum Beispiel, zuverlässigen Bonitätsbewertungen ist nicht von der Hand zu weisen.

Nicht jede Offenlegung begründet das Risiko, dass die Prognosequalität sinkt. Eine Offenlegung ist unschädlich, wenn in den Score nur solche Merkmale eingehen, deren Vorhandensein für die ermittelte Prognosewahrscheinlichkeit tatsächlich ursächlich ist. Dann nimmt die gescorte Person mit einer ihrem Score-Wert dienlichen Verhaltensänderung auf die Wahrscheinlichkeit des prognostizierten Ereignisses tatsächlich Einfluss. Wer sich regelmäßig bewegt, reduziert sein Krankheitsrisiko – man kann deshalb die Entscheidung, regelmäßig Sport zu treiben, nicht als eine „Manipulation“ des Score-Wertes bezeichnen.

Demgegenüber nimmt die Prognosekraft des Scoring-Verfahrens dann ab, wenn die Verhaltensanpassung sich auf Merkmale bezieht, die zwar in der Vergangenheit gute Indikatoren für die durch Scoring zu ermittelnde Wahrscheinlichkeit abgegeben haben, aber die Höhe des Wahrscheinlichkeitswertes nicht beeinflussen (Praxisbeispiele unter B.VIII.1). Niemand senkt sein Krankheitsrisiko dadurch, dass er Sportbekleidung nur kauft, aber nicht benutzt – wer also weiß, dass der Kauf von Turnschuhen als sogenanntes Proxy-Merkmal (siehe näher B.V.2) in einen „Gesundheits-Score“ eingeht, könnte versucht sein, durch bloße Konsumententscheidungen statt durch das Sporttreiben auf seinen Score einzuwirken. Ist die Funktionsweise eines Scoring-Verfahrens offengelegt, kann die gescorte Person die Auswirkungen ihres Verhaltens auf ihren Score-Wert erkennen und ihr Verhalten deshalb „scoredienlich“ anpassen (Bambaucher & Zarsky, 2018).

Die Einflussnahme auf nicht-ursächliche Merkmale wird in der Literatur unter dem Stichwort des „gaming“ eines Scoring-Verfahrens diskutiert (Rona-Tas & Hiss, 2011). Mit dem Ausspruch: „when a measure becomes a target, it ceases to be a good measure“ hat der britische Ökonom Charles Goodhart diese Einsicht in die Selbstreflexivität sozialer Systeme prägnant formuliert. Sie verdient in jeder Diskussion über indikatorengestützte Steuerung Beachtung (Strathern, 1997; Wagner, 2018; Weingart & Wagner, 2015). Ein Score-Gaming zu verhindern, kann im Allgemeininteresse liegen, um die Aussagekraft des prognostischen Scores zu erhalten. Erforderlich ist dann ein gewisses Maß an Intransparenz über die Funktionsweise des Scoring-Verfahrens. Andererseits kann aus der Möglichkeit des „Score-Gaming“ auch gerade die umge-

kehrte Schlussfolgerung gezogen werden: Der richtige Weg, Manipulationsmöglichkeiten abzuwehren, führt nicht über die Aufrechterhaltung von Intransparenz über die in den Score eingehenden Merkmale, sondern über den Ausschluss nicht-ursächlicher Merkmale aus dem Scoring-Verfahren. Dieser Weg mag schwerer zu verwirklichen sein, allerdings ist die erhöhte Fairness eines Abstellens allein auf ursächliche Merkmale nicht von der Hand zu weisen (eingehend Britz, 2008).

2. Transparenz bei verhaltenssteuerndem Scoring

Scoring kann auch ein Instrument der Verhaltenssteuerung sein. Mit diesem Zweck unterlegt, erscheint Transparenz zunächst als eine notwendige Bedingung des wirkungsvollen Einsatzes des Scoring-Verfahrens. Denn ein Anreizsystem kann eine zielgerichtete verhaltenssteuernde Wirkung nur dann entfalten, wenn es den Zusammenhang zwischen dem Verhalten und seiner Bewertung offenlegt. Auf das Scoring bezogen: Wenn es dem Verwender des Scoring-Verfahrens darum geht, zur Verbesserung des Score-Wertes zu motivieren, erscheint es zwingend erforderlich, dass er zumindest offenlegt, dass bestimmte Verhaltensweisen „punketrächtig“ sind.

Allerdings kann die Verhaltenssteuerung durch Scoring-Verfahren auch „weicher“ erfolgen. Dies lässt sich an einem – hypothetischen – Beispiel verdeutlichen. Man stelle sich vor, es wird ein Score-Wert für „gesundheitsförderliches Verhalten“ oder „gutes Autofahren“ ermittelt, dessen Kriterien indessen nicht offengelegt werden. Man wird annehmen dürfen, dass ein solches Scoring-Verfahren Auswirkungen auf das Verhalten der gescorten Person hat. Sie wird versuchen, ihren Score zu verbessern. Nur ist die Richtung dieser Verhaltensänderung unbestimmter, denn die gescorte Person kann nur vermuten, welche Verhaltensweisen als zum Beispiel „gesundheitsförderliches Verhalten“ oder „gutes Autofahren“ vom Scorer bewertet werden und demzufolge zur Verbesserung ihres Scores beitragen. Die gescorte

Person ist also mit der Herausforderung konfrontiert, den ihr unbekanntes Maßstäben des Scoring-Verfahrens zu genügen.

3. Sicherung von Transparenz und Verständlichkeit des Scorings als Dauerthema

Transparenz ist ein Schlüsselinstrument der Verbraucherpolitik. Dementsprechend widmen sich zahlreiche Untersuchungen der Legitimation, der Effektivität und den Grenzen des Transparenzprinzips im Verbraucherschutz (Überblick bei Tamm, 2011, insbes. S. 347 ff.). Allerdings wird die Verknüpfung zwischen Transparenz einerseits und tatsächlicher Informiertheit des Verbrauchers andererseits zunehmend in Zweifel gezogen (Ben-Shahar & Schneider, 2014; Kettner, Thorun & Vetter, 2018; siehe auch B.VIII.2). Aus einem Maximum an Transparenz folgt keineswegs ein Maximum an Verbraucherschutz. Perspektivisch wird deshalb die Absicherung tatsächlicher Verbraucherautonomie ins Zentrum der Diskussion rücken. Die Debatte über die „Algorithmen-Transparenz“ könnte als Katalysator wirken, denn die fehlende Wirksamkeit von Verpflichtungen, die auf die bloße Offenlegung unaufbereiteter Informationen abzielen, ist hier besonders augenfällig. Die Offenlegung des nackten Programmcodes würde jeden Verbraucher überfordern (siehe weiter solgleich unter 4.).

Dessen ungeachtet ist Scoring als ein Datenverarbeitungsvorgang bereits deshalb auf ein gewisses Maß an Transparenz angewiesen, weil nur Transparenz dem Einzelnen ermöglicht, seine subjektiven datenschutzrechtlichen Rechte wahrzunehmen (Bull, 2011). „Die Rechtmäßigkeit von Entscheidungen kann nur prüfen, wer die Datengrundlage, Handlungsabfolge und Gewichtung der Entscheidungskriterien kennt – und versteht.“ (Martini, 2017, S. 1018). Dies betrifft insbesondere die Richtigkeit der individuellen Daten, die zur Bildung von Score-Werten herangezogen werden. Korrekturanprüche (siehe v.a. Art. 16 DSGVO) gehen ins Leere, wo die Fehlerhaftigkeit eines personenbezogenen Datums

nicht bekannt ist (siehe aber zur tatsächlichen Bekanntheit entsprechender Auskunftsrechte B.VII.2).

Die Mittel, mit denen Transparenz im Bereich des Scorings hergestellt werden soll, sind rechtlicher Art. Deshalb hat die wissenschaftliche Diskussion über das dem Scoring angemessene Transparenzniveau ihren Schwerpunkt im Dialog zwischen Gesetzgebung und Rechtsprechung einerseits und der Rechtswissenschaft andererseits. Dabei stand sowohl dem Gesetzgeber als auch der Rechtswissenschaft das Bonitäts-Scoring als besonders regelungsbedürftig vor Augen. Drei Ereignisse strukturieren den Transparenzdiskurs in diesem Bereich.

Eine erste Zäsur ist die Schaffung von scoring-spezifischen datenschutzrechtlichen Bestimmungen im Jahre 2009 (BGBl. I S. 2254). Durch Novellierung des Bundesdatenschutzgesetzes (BDSG) schuf der Gesetzgeber den Scoring-Regelungskomplex aus § 28b BDSG a.F., der Anforderungen an rechtmäßiges Scoring enthielt, und § 28a BDSG a.F., der die Datenübermittlung an Auskunftsteile regelte. Ergänzt wurden diese Regelungen durch eine scoring-spezifische Erweiterung der Auskunftsrechte des Betroffenen in § 34 BDSG a.F. (Heinemann & Wäßle, 2010). Vor dem Erlass dieses Regelungskomplexes hatte sich die Zulässigkeit des Scorings nach den allgemeinen datenschutzrechtlichen Vorschriften bemessen. Die Folge war nicht nur ein erhebliches Maß an Rechtsunsicherheit hinsichtlich der Rechtmäßigkeit des Scorings überhaupt (Petri, 2003; Beckhusen, 2004), die etwa auch in der skeptischen Einschätzung des damaligen Bundesbeauftragten für den Datenschutz anklang (BfDI, 1996, Nr. 31.2.3), sondern auch eine weit verbreitete Klage über unzulängliche Transparenz in Hinblick auf den Einsatz von Scoring-Verfahren und auf ihre konkrete Funktionsweise (Korczak & Wilken, 2008). Wurden dem allgemeinen datenschutzrechtlichen Regelwerk zum Teil weitreichende Auskunftsrechte der Betroffenen entnommen (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, 2005), überwog doch die Diagnose eines bereits in den gesetzlichen Regelungen angelegten Transparenzdefizits (Kloepfer & Kutzschbach, 1998; Möller & Florax, 2003; Petri, 2003; Beckhusen, 2005). Diesen beklagten Transparenzmängeln wollte der Gesetzgeber mit der Schaffung scoring-spezifischer Sondervorschriften abhelfen (s. die Begründung des Regierungsentwurfs des maßgeblichen Gesetzes zur Änderung des Bundes-

datenschutzgesetzes, BT-Drucks. 16/10529, S. 6 u.ö.; Beschlussempfehlung und Bericht des Innenausschusses, BT-Drucks. 16/13219, S. 1 f., 10), wie sich auch aus Darstellungen des Gesetzgebungsprojekts aus der Sicht Beteiligter ergibt (Piltz & Holländer, 2008; Metz, 2009). Aufgrund der damit geänderten rechtlichen Rahmenbedingungen kann an die vor Erlass dieses Gesetzes bereits intensiv geführte Diskussion über Transparenzmängel des Scorings und die dort herausgearbeiteten Transparenzdefizite nicht ohne Weiteres angeschlossen werden, wurden doch vorher offene Fragen durch den Gesetzgeber nunmehr verbindlich entschieden.

Die zweite Zäsur markiert das „SCHUFA-Urteil“ des Bundesgerichtshofs (BGH, Urteil vom 28. Januar 2014 – VI ZR 156/13 –, BGHZ 200, 38). In dieser Entscheidung klärte das Gericht die Reichweite des datenschutzrechtlichen Auskunftsanspruchs nach § 34 Abs. 4 Satz 1 BDSG a.F. Auskunft zu erteilen sei über die in die Bildung des Score-Wertes eingehenden personenbezogenen Datenarten des Auskunftsberechtigten. Keine Auskunft müsse erteilt werden über das Verfahren, nach dem aus diesem personenbezogenen Datenbestand und weiteren Daten ein konkreter Score-Wert ermittelt werde. Insbesondere die Gewichtung, mit der die Daten in die Berechnung eingehen, sei vom Auskunftsanspruch nicht erfasst. Das Scoring-Verfahren genieße als Geschäftsgeheimnis grundrechtlichen Schutz. Das Urteil des Bundesgerichtshofs stieß in der Rechtswissenschaft auf ein ausgeprägtes Interesse; die Rechtsdatenbank Juris weist mehr als ein Dutzend rechtswissenschaftliche Analysen der Entscheidung aus. Sie ergeben ein heterogenes Bild, das von nachdrücklicher Zustimmung (Taeger, 2014) bis zur Kritik (Gärtner, 2014; Schulte am Hülse & Timm, 2014) reicht. Fortan bildete die Entscheidung den maßgeblichen Bezugspunkt der Transparenzdiskussion. So unterziehen die Berichtersteller des für das Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) bzw. später für das Bundesministerium der Justiz und für Verbraucherschutz (BMJV) erstatteten Evaluationsberichts über den 2009 neugeschaffenen Regelungskomplex zum Scoring die Entscheidung einer eingehenden Analyse und Kritik (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Eine verfassungsgerichtliche Korrektur der Entscheidung wird von der Klägerin des Verfahrens derzeit angestrebt. Das Bundesverfas-

sungsgericht hat über ihre Verfassungsbeschwerde gegen das Urteil noch nicht entschieden.

Eine Gesetzgebungsinitiative aus den Reihen der Opposition (BT-Drucks. 18/4864) strebte, letztlich erfolglos, eine Änderung der Rechtslage an. Der Gesetzentwurf sah vor, dass im Bundesdatenschutzgesetz weiterreichende Transparenzanforderungen verankert werden, als der Bundesgerichtshof sie dem Gesetz in der SCHUFA-Entscheidung entnommen hatte. Der Auskunftspflicht unterfallen sollten auch „die verwendeten Einzeldaten, die Gewichtung der verwendeten Daten, die verwendeten Vergleichsgruppen und die Zuordnung der betroffenen Personen zu den Vergleichsgruppen, die in die Berechnung des Wahrscheinlichkeitswerts einfließen“ (a.a.O., S. 4).

Das Inkrafttreten der Datenschutz-Grundverordnung im Mai 2018 markiert die dritte Zäsur in der Transparenzdiskussion. Mit ihr wurde ein unionsrechtlich überformtes nationales Datenschutzrecht abgelöst durch ein unmittelbar anwendbares europäisches Gesetzeswerk. Nicht nur dessen Regelungsinhalt weicht in vielerlei Hinsicht vom früheren Datenschutzrecht ab, auch an die im alten Bundesdatenschutzgesetz etablierte Terminologie und Regelungstechnik und an dessen Gesetzgebungsstil schließt die Datenschutz-Grundverordnung nicht bruchlos an. Es erscheint daher plausibel, dass wegen des Beschlusses der Datenschutz-Grundverordnung der „Beginn einer neuen Zeitrechnung im Datenschutzrecht“ (Schantz, 2016) ausgerufen worden ist. Durchgesetzt ist die Datenschutz-Grundverordnung von Öffnungsklauseln, die dem nationalen Gesetzgeber Gestaltungsspielräume eröffnen. Auf dieser Grundlage ergänzt ein neues Bundesdatenschutzgesetz die Datenschutz-Grundverordnung. Es enthält mit § 31 BDSG n.F. eine scoring-spezifische Sonderbestimmung (siehe näher E.I.3, dort auch zur Unionsrechtskonformität der Vorschrift). Mit dieser Vorschrift zum – so ihre Überschrift – „Schutz des Wirtschaftsverkehrs bei Scoring und Bonitätsauskünften“ wollte der deutsche Gesetzgeber „den materiellen Schutzstandard der §§ 28a und 28b BDSG a.F.“ erhalten, wie die Gesetzesbegründung ausführt (BT-Drucks. 18/11325, S. 101, entspricht BR-Drucks. 110/17, S. 101). Erste juristische Untersuchungen zum Scoring unter der Datenschutz-Grundverordnung gehen nicht von tiefgreifenden Änderungen durch das neue Regelungsregime aus (Taeger, 2016; von Lewinski & Pohl, 2018) und tat-

sächlich dürfen die Beharrungskräfte einer eingespielten Praxis gegenüber legislativen Neuerungen nicht unterschätzt werden. Gleichwohl unterscheidet sich die gesetzgeberische Gestaltung der Transparenzvorgaben und Auskunftsrechte in den Art. 13, 14 und 15 DSGVO tiefgreifend von den §§ 19, 34 BDSG a.F. (s. weiter E.III.4).

Nicht nur in der Wissenschaft, auch in der Öffentlichkeit wird die Transparenz von Scoring-Verfahren diskutiert. Im Februar 2018 starteten die gemeinnützigen Organisationen Open Knowledge Foundation und Algorithm-Watch die Initiative OpenSCHUFA. Zu den erklärten Zielen des Projektes gehört es, den Algorithmus zu „knacken“, mit dem die SCHUFA ihre Bonitäts-Scores ermittelt (OpenSCHUFA, 2018). Sowohl die in die Berechnung des Score-Wertes eingehenden Daten als auch das Verfahren, nach dem aus diesem Datenmaterial ein individueller Score ermittelt wird, will die Initiative herausfinden, indem möglichst viele Personen der Initiative ihren SCHUFA-Score (den sie dort erfragen) und ihre persönlichen Merkmale mitteilen. Zahlreiche Medien berichteten über das Anliegen der Initiative (exemplarisch Erdmann, 2018; Schneider, 2018), die SCHUFA selbst trat ihr kritisch entgegen (SCHUFA Holding AG, 2018a).

4. Scoring-Transparenz als Spezialfall von Algorithmen-Transparenz

Derzeit wird eine intensive Debatte geführt über ein angemessenes Regulierungsregime für „Algorithmen“. Diese Debatte findet nicht nur unter Wissenschaftlern statt, sondern beschäftigt auch intensiv die deutsche Politik. Im Koalitionsvertrag zwischen CDU, CSU und SPD zur 19. Legislaturperiode sind Regulierungsziele für den Bereich algorithmischer Entscheidungen formuliert (CDU/CSU/SPD, 2018, Zeilen 2092 ff.). Auch die Verbraucherpolitik (CDU/CSU/SPD, 2018, Zeilen 6266 ff.) sowie Verbände (Gesamtverband der Deutschen Ver-

sicherungswirtschaft e.V., 2018; Verbraucherzentrale Bundesverband e.V., 2017) und Akteure der Zivilgesellschaft (vgl. etwa die sich unter www.algorithmethik.de und www.algorithmwatch.de präsentierenden Initiativen) haben das Thema der Algorithmenregulierung für sich entdeckt. Der SVRV hat sich grundsätzlich positioniert (SVRV, 2016; SVRV, 2017; SVRV, 2017a; zusammenfassend Micklitz, 2017) und gefordert, dass durch rechtliche Vorgaben sichergestellt werden muss, dass die zugrundeliegenden Parameter bei Algorithmen mit direktem Verbraucherkontakt transparent zu machen und standardisiert einem Kreis von Experten einer Digitalagentur¹⁸ offenzulegen sind (SVRV, 2017a; weiter Gigerenzer, Wagner & Müller, 2018).

Ein Scoring-Algorithmus ist ein bestimmter Typ Algorithmus (Just & Latzer, 2016). Die Diskussion über Scoring-Transparenz lässt sich deshalb als ein Teil der allgemeinen Debatte über die Regulierung von Algorithmen führen. Zwar sind herkömmliche Scoring-Verfahren derzeit bedeutend weniger komplex als die algorithmischen Entscheidungssysteme, die in der Algorithmen-Regulierungsdebatte üblicherweise als Referenzpunkte dienen (SVRV, 2016) und die nicht selten dem Bereich der Künstlichen Intelligenz zuzurechnen sind. Aber auch die gegenwärtig eingesetzten Algorithmen sind für Nicht-Fachleute nicht einfach zu verstehen (siehe weiter B.IV.2 unten). Und nähme die Komplexität der eingesetzten Verfahren zu – hin etwa zu Scoring-Verfahren, die auf Systemen maschinellen Lernens, insbesondere neuronalen Netzen aufbauen (siehe Hurley & Adebayo, 2016) –, würde die Debatte über die „Algorithmen-Transparenz“ auch den Bereich des Verbraucher-Scorings in wachsendem Ausmaß betreffen. Ob es auf breiter Front zu dieser Zunahme der Komplexität von Scoring-Verfahren kommen wird, die sie zu „Black Boxes“ machen würden, ist unsicher, nicht zuletzt aus dem einfachen Grund, dass es eine offene Frage ist, ob solche neuen Scoring-Verfahren den hergebrachten Methoden in einer Weise überlegen sind, die ihre Verwendung wirtschaftlich rechtfertigen würde. So steht der Nachweis aus, dass neuartige algorithmische Entscheidungsverfahren immer „besser“ (etwa mit Blick auf Modellgüte) sind als etablierte Verfahren.

¹⁸ Zu einer solchen Digitalagentur siehe auch Tutt (2017), der für eine zentrale Regulierungsbehörde für Algorithmen nach dem Modell der Food and Drug Administration eintritt und deren Kompetenzen umreißt (S. 105 ff.): „the agency should serve as a centralized expert regulator that develops guidance, standards, and expertise in partnership with industry to strike a balance between innovation and safety.“ (S. 83).

Beispiele hierfür sind „Google Flu Trends“ (zur Prognose von Grippe-Epidemien) und COMPAS (zur Prognose der Rückfallwahrscheinlichkeit von Straftätern), bei denen jeweils die Aussagekraft komplexer Algorithmen widerlegt wurde und stattdessen einfache Regeln eine höhere Prognosekraft zeigten (Dressel, 2018; Lazer, Kennedy, King & Vespignani, 2014).¹⁹

Die Relevanz der Frage, ob und inwieweit Entwicklungen im Bereich der Künstlichen Intelligenz die Praxis des Scorings betreffen werden, ist als eher gering zu veranschlagen. Denn die Schwierigkeiten, Einblick in Prozesse algorithmischer Entscheidungsfindung zu erhalten, betreffen keineswegs nur „moderne“ Verfahren algorithmischer Entscheidungsfindung. Auch ganz konventionelle Algorithmen, die beispielsweise – wie der Bonitäts-Score der SCHUFA – auf multivarianten oder nicht-linearen Regressionsmodellen beruhen, können selbst für Fachleute nicht auf den ersten Blick zu durchschauen sein (Lipton, 2017). Die im Zusammenhang mit Künstlicher Intelligenz vielzitierte Black Box ist also nicht erst in jüngster Vergangenheit zur Herausforderung geworden. Mit nur wenig Übertreibung kann man stattdessen sagen, dass die Black Box die Entwicklung der Softwaretechnik von Anfang an begleitet hat (pointiert Passig, 2017).

Mit der Offenlegung des Quellcodes selbst einfacher Computerprogramme ist den Verbrauchern in aller Regel nicht geholfen (siehe auch B.VII.2). Die Mehrheit der betroffenen Personen sind keine IT-Spezialisten. Und selbst wenn sie es wären: die technische Komplexität der Computersysteme, deren Entscheidungsverhalten „transparent“ gemacht werden soll, macht selbst für die Kontrolle durch Experten eine andere Form von Transparenz erforderlich als die bloße Offenlegung von Programmcodes (Wischmeyer, 2018; Samek, Wiegand & Müller, 2017; Selbst & Powles, 2017; Montavon, Samek & Müller, 2018; Gigerenzer, Wagner & Müller, 2018; Gesellschaft für Informatik, 2018). An diese Einsicht knüpft sich ein derzeit mit großer Intensität verfolgtes Forschungsprogramm, das der Frage nachgeht, wie die Funktionsweise komplexer Algorithmen für Menschen verständlich gemacht werden kann. Für den Bereich der

Künstlichen Intelligenz wird diese Diskussion unter den Schlagworten des *interpretable machine learning* und der „erklärbaren Künstlichen Intelligenz“ (*explainable artificial intelligence*, XAI, siehe Gesellschaft für Informatik, 2018) geführt (Wachter, Mittelstadt & Floridi, 2017; Selbst & Powles, 2017; Selbst & Barocas, 2018).

Ein gegenstandsangemessenes Verständnis von Transparenz wird auf die Einbettung algorithmischer Entscheidungssysteme in arbeitsteilige Begründungs- und Kontrollarchitekturen abzielen (Wischmeyer, 2018). Transparent ist ein System algorithmischer Entscheidungsfindung nicht dann, wenn es in allen Einzelheiten offen vor dem Betrachter liegt (was bei neuronalen Netzen durchaus auch schwierig wäre, vgl. etwa Ribeiro, Singh & Guestrin, 2016; Burrell, 2016; Alber, Lapuschkin & Seegerer, 2018), sondern wenn es sich seinen Benutzern zu erklären vermag: „Lässt sich nachträglich begründen oder sichtbar machen, wie der Weg zum Ergebnis aussah?“ (Passig, 2017, S. 25). Deshalb ist es nicht notwendig, dass Transparenz „volles Verständnis“ für Scoring herstellt. Es reicht aus, Möglichkeiten zu schaffen, auch unter den Bedingungen unvollständiger Transparenz und unvollständigen Verständnisses Erkenntnisse über die Funktionsweise eines Algorithmus zu erlangen, das sogenannte *black box tinkering* (Perel & Elkin-Koren, 2017; Wachter, Mittelstadt & Floridi, 2017). Der Vorschlag einer Transparenz-Schnittstelle (Gigerenzer, Wagner & Müller, 2018) knüpft hieran ebenso an wie die Vorschläge der Gesellschaft für Informatik, das Testen von Algorithmen als Regulierungsinstrument stark zu machen (Gesellschaft für Informatik, 2018, mit der Empfehlung zur Schaffung eines „Rechts auf Durchführung von Tests“). Hierzu werden die Eingaben systematisch variiert und die Ausgaben ausgewertet. Dies könnte z. B. eine datenschutzrechtliche Aufsichtsbehörde im Rahmen von sogenannten Datenschutzüberprüfungen („audits“) nach Art. 58 Abs. 1 Buchstabe b DSGVO verlangen und ggf. selbst durchführen. Was im Inneren der Black Box im Detail vor sich geht, erkennt man zwar auf diese Weise nicht notwendigerweise, aber man kann hinreichende Erkenntnisse über die *relevanten* Funktionsweisen des Algorithmus gewinnen. Dieses *testing* folgt im Übrigen der Logik der Stiftung Warentest,

19 Zur Problematik des Einsatzes von Algorithmen in der amerikanischen Justiz siehe Kehl, Guo & Kessler (2017). Die möglichen Schlussfolgerungen für das Verbraucherrecht sind bislang nicht wirklich untersucht.

soweit diese für ihre Tests nicht Baupläne und Rezepte studiert, sondern aus dem Gebrauch eines Produktes auf seine relevante Qualitäten schließt.

Auch von den Entwicklern und Anwendern von Scoring-Systemen wird inzwischen gesehen, dass die Nachvollziehbarkeit ihrer Verfahren wichtig ist. Francesca Rossi von IBM formulierte in einem Interview mit einer Tageszeitung: „Neben *deep learning* gibt es Systeme wie *decision trees*, die besser nachzuvollziehen sind, aber leider nicht ganz so akkurat. Wir müssen also herausfinden, was uns wichtiger ist: das akurate Ergebnis oder die Nachvollziehbarkeit des Ansatzes.“ (Rossi, 2018).²⁰

5. Transparenz als Bedingung einer gesellschaftlichen Debatte über Scoring

In der Verwendung von Scores liegt das Risiko, gesellschaftlich nur unzureichend diskutierte Wertungen mit dem Anschein der Objektivität zu versehen und dadurch der Kritik zu entheben (grundlegend Porter, 1995; s. weiter etwa Heintz, 2007). Diese Kritik der Quantifizierung und des „gesellschaftlichen Zahlengebrauchs“ (Vormbusch, 2012, S. 37) ist inzwischen zum Gegenstand einer Vielzahl von Studien gemacht worden, die ein produktives Forschungsfeld begründen. Etablierte Referenzgebiete für solche Analysen sind die Wirtschaftspolitik (Weingart & Wagner, 2015; Wagner, 2018; Schlaudt, 2018) ebenso wie diverse Felder der Bildungs-, Gesundheits- und Sozialpolitik (Muller, 2018) und insbesondere das Handeln internationaler Organisationen, soweit sich dieses maßgeblich auf Indikatoren, Rankings, Grenzwerte usw. stützt (Davis & Fisher, 2012; Rottenburg, Merry, Park & Mugler, 2015; Merry, Davis & Kingsbury, 2015; Merry, 2016).

Anhand zahlreicher Beispiele und Fallstudien haben diese Analysen aufgezeigt, dass, wo immer eine Entscheidung zu treffen ist, an eine Zahl deutlich leichtgängiger angeschlossen werden kann als an eine vielschichtige, differenzierte und unter Umständen auch ambivalente Beurteilung eines Sachverhalts oder – bezogen auf das Scoring – einer Person. Damit reduzieren Scores die Komplexität qualitativer Urteile in der beinahe größtmöglichen Art und Weise. Das macht die Heranziehung von Scores zur Entscheidungsfindung sehr attraktiv, zumal wenn diese Entscheidungen automatisiert, schnell und massenhaft getroffen werden (müssen). Allerdings werden die Wertungen, die bei der Entwicklung von Scoring-Verfahren getroffen werden müssen, oft nicht in einer Weise diskutiert, die der späteren sozialen Verwendung dieser Verfahren angemessen ist. Welche Kriterien mit welchem Gewicht in die Ermittlung eines „Punktwertes für gesundheitsförderliches Verhalten“ in einem Krankenversicherungs-Bonusprogramm eingehen, liegt üblicherweise außerhalb des Aufmerksamkeitsfeldes der Öffentlichkeit (siehe auch C.III.2). Ob beispielsweise lediglich solche Aktivitäten positive Berücksichtigung finden sollen, die dem eigenen Gesundheitszustand förderlich sind, oder auch solche, mit denen die Funktionsfähigkeit des Gesundheitssystems unterstützt werden soll (zum Beispiel Blutspende, Knochenmarkstypisierung) oder gar solche, die ohne Gesundheitsbezug als sozial wertvoll gelten (zum Beispiel ehrenamtliches Engagement), ist eine diskutabile Frage. Das Fehlen öffentlicher Diskussionen über die Wertungen, die bei der Schaffung eines Score-Verfahrens zu treffen sind, könnte man als ein Mangel an Politisierung bezeichnen, d. h. als eine Durchsetzung normativer Wertungen mit erheblichen sozialen Folgen ohne einen vorbereitenden und begleitenden gesellschaftlichen Diskurs.

²⁰ Zu der Frage, ob dieser Trade-Off in der Realität des Scorings überhaupt besteht, siehe etwa Hand, 2006, der ihn mit guten Argumenten bestreitet. Hier geht es nur darum, dass die Praxis den Belang der Verständlichkeit offenkundig anerkennt. Insoweit ist es auch kein Gegeneinwand, dass *random forests*, also eine Vielzahl von *decision trees* (Gesellschaft für Informatik, 2018), nicht notwendigerweise leichter interpretierbar sind (Groll, Ley, Schauburger & Van Eetvelde, 2018) und das Problem der Verständlichkeit deshalb durch die Methodenumstellung nicht verschwindet oder auch nur notwendigerweise verringert wird.

Unzureichende Transparenz kann zudem gesellschaftlichen Fehlvorstellungen darüber Vorschub leisten, was ein Score eigentlich aussagt (siehe auch C.III.3). Zum Beispiel kann ein Kfz-Score als Grundlage für einen Telematik-Tarif so gebildet werden, dass er nicht nur auf das beeinflussbare Fahrverhalten abstellt, also etwa die Umsicht, mit der der Fahrer bremst und beschleunigt oder Geschwindigkeitsbegrenzungen beachtet, sondern auch fahrleistungsindifferente, aber unfallwahrscheinlichkeitsrelevante Merkmale erfassen. Dies können zum Beispiel das Verhältnis von Stadtfahrten und Landfahrten (denn in der Stadt ist die Unfallwahrscheinlichkeit höher) oder von Nachtfahrten und Tagesfahrten sein (denn nachts zu fahren erhöht die Wahrscheinlichkeit eines Unfalls). Entsteht bei dem gescorten Autofahrer oder der Öffentlichkeit der Eindruck, der Score-Wert drücke vor allem die Fähigkeiten des Fahrers aus, fallen die reale Bedeutung des Score-Wertes und seine soziale Verwendung auseinander. Transparenz in dieser Hinsicht zielt somit darauf, die Bedeutung von Scores realistisch einzuschätzen und Scores nur dieser Bedeutung entsprechend einzusetzen.

II. Diskriminierungsschutz und Gleichbehandlung

Scoring-Verfahren führen zu von Person zu Person unterschiedlichen Scores. Darin liegt gerade ihr Sinn: Scores markieren Unterschiede und Scoring-Verfahren wollen differenzieren. Jede marktwirtschaftliche demokratische Wirtschafts- und Rechtsordnung baut auf der Möglichkeit auf, privatautonom Unterscheidungen zu treffen. Prinzipiell steht es jedem Unternehmer frei, ob er mit einem Verbraucher einen Vertrag abschließen will oder nicht. Eine Verpflichtung zum Vertragsschluss gibt es nur in wenigen Ausnahmefällen. Umgekehrt gewinnt der Diskriminierungsschutz in der Rechtsordnung immer weiter an Boden. Scoring-Verfahren bewegen sich genau in diesem Spannungsverhältnis von unternehmerischer Freiheit und gesellschaftlichen Wertvorstellungen, das immer wieder neu austariert werden muss.

1. Was ist Diskriminierung?

Das Phänomen Diskriminierung wird in diesem Gutachten weit verstanden. Es erfasst Handlungen und Strukturen, die dazu führen, dass die Träger bestimmter Merkmale²¹ (etwa Frauen, Homosexuelle oder Personen „fremder“ ethnischer Herkunft) im gesellschaftlichen Leben benachteiligt werden. Dafür, welche Merkmale insoweit relevant sind, liefert § 1 des Allgemeinen Gleichbehandlungsgesetzes (AGG) Orientierungspunkte. Die Vorschrift bestimmt: „Ziel des Gesetzes ist, Benachteiligungen aus Gründen der Rasse oder wegen der ethnischen Herkunft, des Geschlechts, der Religion oder Weltanschauung, einer Behinderung, des Alters oder der sexuellen Identität zu verhindern oder zu beseitigen.“ In anderen Vorschriften wird der Kreis der sogenannten Diskriminierungsmerkmale²² abweichend bestimmt, wiewohl natürlich zahlreiche Überschneidungen beste-

hen. Die jeweils maßgeblichen Merkmale, nach denen zu unterscheiden unerwünscht oder sogar verboten ist, sind das Ergebnis von gesellschaftlichen Aushandlungsprozessen, von Einsicht in Strukturen sozialer Exklusion und in historisches Unrecht sowie, letzten Endes, von zivilisatorischem Fortschritt (Fritzsche, 2017). Sie müssen nicht für alle Lebensbereiche und sozialen Situationen identisch bestimmt werden und sie sind der gesetzgeberischen Anpassung und Fortentwicklung zugänglich.

Eine solche Bestimmung des Phänomens Diskriminierung ist von zwei konkurrierenden Bedeutungen des Begriffs abzugrenzen: Einerseits ist nicht schlechthin jede Unterscheidung zwischen Personen eine „Diskriminierung“, wie sie Gegenstand dieses Gutachtens ist. Ein solches Begriffsverständnis (vgl. Adomeit, 2002; Picker, 2008) hätte zur Folge, dass unter „Diskriminierung“ eine Vielzahl gesellschaftlicher Vorgänge zu fassen wären, die nicht problematisiert werden müssen und keine politische Handlungsnotwendigkeit auslösen, bis hin zu der Unterscheidung des Gastwirts zwischen zahlungswilliger und nicht zahlungswilliger Kundschaft. Auch manchen statistischen Fachbegriffen wie etwa dem der Diskriminanzanalyse liegt ein wertneutrales Verständnis des Begriffs Diskriminieren – das Unterscheiden – zugrunde. Andererseits soll mit dem Begriff der Diskriminierung auch nicht gemeint sein, dass die Unterscheidung nach dem Diskriminierungskriterium kategorisch gesellschaftlich missbilligt oder gar rechtlich verboten ist. Eine „Diskriminierung“ soll stattdessen jede Ungleichbehandlung nach einem Kriterium bezeichnen, dessen Heranziehung zu Unterscheidungszwecken als besonders legitimationsbedürftig bewertet wird. Der bloße Verweis auf das freie Belieben desjenigen, der die Unterscheidung vornimmt, reicht zur Rechtfertigung der Ungleichbehandlung nicht aus.²³

21 Zur Klarstellung: Es geht nicht darum, dass die genannten Merkmale „objektiv“ vorliegen und der diskriminierten Person gewissermaßen inhärent sind.

Diskriminierungsverbote schützen vor Unterscheidungen aufgrund bloß zugeschriebener Merkmale (Schiek, 2000; für das AGG siehe BT-Drucks. 16/1780, S. 30 f.). Damit ist etwa verständlich, dass sich ein Diskriminierungsverbot auf das Merkmal „Rasse“ beziehen kann, die eine soziale, aber keine anthropologische Kategorie darstellt.

22 Es herrscht eine große begriffliche Vielfalt, die verwirren kann. Man spricht nicht nur von „Diskriminierungsmerkmalen“ (Pärl, 2017, S. 106 ff.), sondern auch von „verbotenen Merkmalen“ (Schramm, 2013, S. 7), wiewohl verboten selbstverständlich nicht das Merkmal ist, sondern unter Umständen die an das Merkmal anknüpfende Diskriminierungshandlung. Gleichbedeutend spricht man auch von „geschützten Merkmalen“ (Schramm, 2013, S. 3 und passim); in dieser Terminologie klingt an, dass Zweck des Diskriminierungsverbots der Schutz von Trägern bestimmter Merkmale ist.

23 Aus juristischer Perspektive: Mit dem Verweis auf die Privatautonomie, verstanden als „das Prinzip der Selbstgestaltung der Rechtsverhältnisse durch den einzelnen nach seinem Willen“ (Flume, 1965, S. 1) und als „Anerkennung der ‚Selbstherrlichkeit‘ des einzelnen“ (a.a.O., S. 6), ist die vom Entscheider vorgenommene Ungleichbehandlung nicht abschließend legitimiert.

2. Diskriminierung durch Scoring-Input

Scoring-Verfahren ist die Gefahr der Diskriminierung inhärent. Scoring-Verfahren knüpfen an eine Vielzahl von Merkmalen an, um einen personenbezogenen Score zu ermitteln. Wenn zu diesen erfassten und für die Score-Bildung herangezogenen Merkmalen auch geschützte Merkmale gehören, liegt es auf der Hand, dass das Scoring-Verfahren diskriminierend wirkt. Man kann hier von einer unmittelbaren – oder, gleichsinnig: direkten – Diskriminierung durch Scoring sprechen. Zum Beispiel diskriminiert ein Scoring-Verfahren unmittelbar, wenn das Merkmal Geschlecht oder eine bestimmte ethnische Gruppenzugehörigkeit als Input in die Bildung des Score-Wertes eingeht. Das Vorliegen einer solchen unmittelbaren Diskriminierung durch Scoring ist vergleichsweise einfach zu erfassen. Sofern man die Tatsache ermitteln kann (siehe oben I.1 und 2 zur Transparenz des Scoring-Verfahrens), dass die Information über ein Diskriminierungsmerkmal als „Input“ in die Berechnung des Scores eingeht und gerade dies zu einer Schlechterstellung der Merkmalsträger führt, ist die Diskriminierung gezeitigt (siehe auch Hacker, 2018).

Diskriminierungsgefahren können sich im Bereich des Scorings aber nicht nur dadurch ergeben, dass ein Scoring-Verfahren direkt an ein geschütztes Merkmal anknüpft. Im Gegenteil, dies dürfte ein vergleichsweise seltenes Phänomen sein. Sozial relevanter ist die Gefahr einer mittelbaren – oder, gleichsinnig: indirekten – Diskriminierung durch Scoring-Verfahren (anschaulich Hofstetter, 2016). Hier nimmt das Scoring-Verfahren Merkmale als Input, die zwar für sich genommen harmlos sind, aber mit geschützten Merkmalen in einem statistischen Zusammenhang stehen. Zum Beispiel kann man an die Körpergröße, an das Konsumverhalten oder an Freizeitgewohnheiten anknüpfen, um mittelbar das Merkmal „Geschlecht“ zu erfassen. Anschauungsmaterial zu diesem Phänomen bilden jene Studien, die aufzeigen, wie man etwa die ethnische Zugehörigkeit oder die

sexuelle Orientierung aus der schmalen Datenbasis einer paar Handvoll „Facebook-Likes“ mit einiger Wahrscheinlichkeit ermitteln kann (Kosinski, Stillwell & Graepel, 2013; siehe auch B.VIII.2). Es liegt also auf der Hand, dass die unmittelbare Diskriminierung das Problem diskriminierender Scoring-Verfahren nicht ausschöpft.

Spätestens an dieser Stelle wird die Diskriminierungsproblematik außerordentlich komplex. Denn es wird kaum Merkmale geben, von denen sich von vornherein sicher annehmen ließe, dass sie mit dem Vorliegen eines Diskriminierungsmerkmals nicht in Zusammenhang stehen. Vielmehr werden zahlreiche Merkmale, die in einen Score eingehen, teils positiv und teils negativ, teils stärker und teils schwächer mit Diskriminierungsmerkmalen korrelieren. Zum Beispiel liegt die Annahme nicht fern, dass es eher Männer sind, die ein riskantes Fahrverhalten an den Tag legen, das wiederum zu einem vergleichsweise schlechteren „Kfz-Score“ führt. Die Anknüpfung an das neutrale Merkmal Fahrverhalten ist damit unter dem Gesichtspunkt des Diskriminierungsschutzes zumindest aufmerksamkeitsbedürftig (zur Weiterführung des Beispiels siehe unten 3.2).

Tatsächlich ist das Problem sogar noch komplexer. Denn problematisch unter dem Gesichtspunkt des Diskriminierungsschutzes ist nämlich auch die Anknüpfung an ein Merkmal, das zwar mit einzelnen Diskriminierungsmerkmalen (Geschlecht, ethnische Herkunft, ...) in keinem statistisch signifikanten Zusammenhang steht, wohl aber mit Kombinationen verschiedener Diskriminierungsmerkmale (Personen eines bestimmten Geschlechts und gleichzeitig einer bestimmten ethnischen Herkunft). Mit der Intersektionalitätsforschung hat sich eine wissenschaftliche Fachdisziplin etabliert, die sich mit der Situation von Personen beschäftigt, die mehrere geschützte Merkmale in sich vereinen (Meyer, 2017; Chege, 2012). Die Untersuchungen in diesem Bereich unternehmen es anhand zahlreicher Beispielen, den Nachweis zu führen, dass solche Kombinationen geschützter Merkmale eigenständig konturierte Diskriminierungskategorien bilden, die nicht auf die „einfacheren“ Diskriminierungsmerkmale reduzierbar sind.²⁴

²⁴ Das geltende Antidiskriminierungsrecht bildet diesen Umstand (noch) nicht ab. Der Gerichtshof der Europäischen Union, die maßgebliche Instanz für die Auslegung des – in erheblichem Ausmaß europäisch harmonisierten – Antidiskriminierungsrechts, erkennt die eigenständige intersektionelle Diskriminierung nicht an, s. EuGH, Urteil vom 24. November 2016 – Rs. C-443/15 –, Parris [ECLI: EU:C:2016:897], Rn. 80 bis 82). Angesichts der Kritik und der intensiven Diskussion um die Einordnung der intersektionellen Diskriminierung in Wissenschaft und Praxis dürfte mit weiteren Verfahren vor dem EuGH zu rechnen sein.

3. Score-Güte und Diskriminierungsverbot

Eine Analyse diskriminierenden Scorings, die den Input des Scoring-Verfahrens in den Blick nimmt, betrifft verhaltenssteuernde wie verhaltensprognostische Scoring-Verfahren. Dagegen tritt der Aspekt der Score-Güte nur bei verhaltensprognostischen Scoring-Verfahren auf den Plan. Solche Scoring-Verfahren wollen eine bestimmte Prognoseleistung erbringen. Diese Leistung können sie besser oder schlechter erfüllen. Hohe Score-Güte und Diskriminierungsverbot stehen zueinander in keinem eindeutigen Verhältnis. Als Anforderungen an ein faires Scoring-Verfahren können sie Hand in Hand gehen, sie können aber auch in einem Spannungsverhältnis zueinander stehen.

3.1 Score-Güte und Diskriminierungsverbot im Gleichklang

Hand in Hand gehen die normativen Ziele eines allgemeinen Diskriminierungsverbotes und der Gütegewährleistung in besonders anschaulicher Weise dort, wo die Anknüpfung an ein geschütztes Merkmal lediglich Ausdruck von überkommenen Vorlieben und Abneigungen ist. Für eine gute Prognoseleistung eines Scoring-Verfahrens tragen solche sogenannten „präferenzbedingten Diskriminierungen“ (Gardner, 1998; Block, 2018, § 3 AGG Rn. 10 ff.) nichts bei. Im Gegenteil: Hier verstellen dem Scoring-Nutzer seine Vorurteile den Blick darauf, dass ein leistungsfähigeres Scoring-Verfahren möglich wäre. An einem hypothetischen Beispiel illustriert: Wer als Schöpfer eines Scoring-Verfahrens zur Bestimmung der Unfallwahrscheinlichkeit beim Autofahren meint, dass „Frauen ohnehin schlechter Auto fahren“ und deshalb an das Geschlecht einen Score-Malus knüpft, diskriminiert nicht nur, sondern gestaltet auch sein Scoring-Verfahren schlechter, als es ihm möglich wäre. Denn ohne dass sich hierfür empirische Gründe anführen ließen, wird das Scoring-Verfahren Autofahrerinnen eine höhere Unfallwahrscheinlichkeit zuordnen als Autofahrern. Das Scoring-Verfahren ist mangelhaft. Diese Karikatur zeigt bereits, dass die praktisch drängenden Probleme der Diskriminierung durch Scoring heute woanders liegen.

Denn deutlich relevanter ist die „diskriminierende Fehlerhaftigkeit“ von Scoring-Verfahren. Gemeint ist damit, dass die Güte des Scoring-Verfahrens nicht für alle gescoreten Personen gleich gut oder schlecht ist, sondern bestimmte Personengruppen besonders intensiv von Mängeln des Scoring-Verfahrens betroffen sind. Man stelle sich nur einen Toaster vor, der häufiger in Flammen aufgeht, wenn er von Frauen bedient wird, als wenn er von Männern bedient wird – was bisher eine absurde Vorstellung war, wird in der Welt komplexer prognostischer Scoring-Algorithmen zu einem realen Problem. Als ein personenbezogenes „Produkt“ können Scoring-Verfahren für unterschiedliche Personen unterschiedlich gut funktionieren. Gewisse Parallelen weist die Diskussion darüber auf, ob bestimmte Medikamente aufgrund der sozialen Bedingungen, unter denen Pharmaunternehmen sie erforschen, entwickeln und testen – zum Beispiel nur an männlichen Probanden –, vermeidbare Qualitätsunterschiede für Männer und Frauen aufweisen (dazu Nieber, 2014).

Die Diskussion über „Big-Data-Diskriminierung“ vermittelt wichtige Impulse über das Zusammenspiel von Gütegewährleistung und Diskriminierungsverbot (Barocas & Selbst, 2016). In der Diskussion wird herausgearbeitet, wie die mangelhafte Konstruktion von Big-Data-Anwendungen diskriminierende Effekte zeigen kann. Eine Zusammenfassung aus jüngster Zeit (Executive Office of the President of the United States, 2016) gliedert die Ursachen für diskriminierende Effekte in einerseits solche, die in der Datenbasis für Big-Data-Anwendungen liegen, und andererseits solche, die sich aus dem Design der Algorithmen ergeben, die diese Datenbasis verarbeiten (siehe auch Hacker, 2018).

Ein Beispiel ist etwa ein automatisiertes System zur Bewerberauswahl, das stets Männer ohne Migrationshintergrund als besonders geeignet für eine Stelle auszeichnet – weil ein aus der Unternehmenshistorie gewonnener Datenbestand, auf dessen Grundlage das System seine Entscheidungskriterien „gelernt“ hat, Männer ohne Migrationshintergrund als besonders erfolgreich auszeichnet (frühes, vieldiskutiertes Beispiel bei Lowry & Macpherson, 1988). Dass sich in diesem Datenbestand überkommene diskriminierende Einstellungen, eine „unintentional perpetuation and promotion of historical biases“ (Executive Office of the President of the United States, 2016, S. 8) niederschlagen, ist hier die naheliegendste Erklärung.

3.2 Score-Güte und Diskriminierungsverbot im Spannungsverhältnis

Die Gewährleistung einer hohen Score-Güte und der Schutz vor Diskriminierung gehen nicht notwendigerweise und immer Hand in Hand. Ein bestimmtes Merkmal kann prognostische Kraft haben und zugleich kann es entweder selbst ein geschütztes Merkmal sein oder mit dem Vorliegen eines geschützten Merkmals korrelieren. Bereits genannt wurde das Beispiel des „riskanten Fahrverhaltens“, das – so sei hier unterstellt – vor allem Männer an den Tag legen. Gleichzeitig ist es evident, dass für die Ermittlung eines Scores, der die Unfallwahrscheinlichkeit prognostizieren will, die Anknüpfung an riskantes Fahrverhalten die Prognoseleistung erheblich verbessert. Zwischen der prognostischen Leistung des Scoring-Verfahrens und dem Belang des Diskriminierungsschutzes kann es deshalb zu Zielkonflikten kommen.

Dieses Dilemma ist anzuerkennen. Die Probleme defizitärer Score-Güte und diskriminierender Scoring-Verfahren dürfen nicht gleichgesetzt werden. Die hier erforderlichen Abwägungen sind gesellschaftlich zu diskutieren. Diese Diskussion kann kontrovers verlaufen, sie muss es aber nicht. Im Falle der negativen Berücksichtigung riskanten Fahrverhaltens dürfte das Ergebnis der Abwägung auf der Hand liegen: Sicherlich würde sich niemand auf den Standpunkt stellen, dass eine Berücksichtigung riskanten Fahrverhaltens für die Ermittlung eines Scores ausgeschlossen sein soll, nur weil dadurch Männer als die riskanteren Fahrer intensiver betroffen sind als Frauen als die vorsichtigeren Fahrer. Dahinter steht die legitime gesellschaftliche Erwartung, dass Männer in der Lage sein sollten, ihr Fahrverhalten zu ändern. Für andere Merkmale – wie Geschlecht oder ethnische Herkunft – wären entsprechende Erwartungen demgegenüber als illegitim bis unerfüllbar zurückzuweisen.

Deutlich zu kurz gegriffen wäre es, wollte man unter Hinweis auf die prognostische Leistungsfähigkeit eines Scoring-Verfahrens diskriminierende Effekte ohne weiteres in Kauf nehmen. Damit würde man verkennen, dass eine gute prognostische Aussagekraft gerade auch das Ergebnis früherer gruppenbezogener Diskriminierungen sein kann (Britz, 2008). Dieser Zustand würde sich durch die Anerkennung unmittelbar diskriminierender prognostischer Scoring-Verfahren gerade verfestigen (Block, 2018, § 3 AGG Rn. 15). An einem Beispiel lässt sich dieser Zusammenhang verdeutlichen: Es erscheint möglich,

dass eine bestimmte ethnische Zugehörigkeit prognostische Kraft besitzt, die Ausfallwahrscheinlichkeit eines Kredits zu bestimmen. Nicht von vornherein verwerfen kann aber man die Vermutung, dass sich hierin der Umstand widerspiegelt, dass sich die Angehörigen dieser ethnischen Gruppe aufgrund vorgängiger Diskriminierungen generell in einer ökonomisch schwächeren Position befinden. Akzeptiert man, dass Scoring-Verfahren, mit denen über die Kreditvergabe entschieden wird, an das Merkmal der ethnischen Herkunft anknüpfen dürfen, verfestigen sich solche Lagen sozialer Benachteiligung (Kim, 2017).

3.3 Das Beispiel des „Redlining“

Die Verwendung von Anschriftendaten zum Zweck des Bonitäts-Scorings bietet Anschauungsmaterial, um das Spannungsverhältnis von prognostischer Güte und Diskriminierungsverbot von Scoring-Verfahren zu verdeutlichen (siehe auch B.V.2). § 31 Abs. 1 Nr. 3 BDSG bestimmt, dass prognostisches Scoring nur zulässig ist, wenn „für die Berechnung des Wahrscheinlichkeitswerts nicht ausschließlich Anschriftendaten genutzt wurden“. Ausgesprochen wird mit der Vorschrift ein Verbot, sogenanntes „Redlining“ zu betreiben. „Der Begriff beschreibt die Praxis, Bereiche einer Karte mit einer roten Linie einzufassen, in denen z. B. eine Bank keine Hypothek oder nur zu schlechteren Konditionen vergibt. Beim Redlining kann also alleine die Auswertung der Anschrift des Betroffenen zu einer negativen Kreditentscheidung, zu schlechteren Konditionen oder zu strengeren Zahlungsbedingungen (z. B. Vorkasse) führen“ (Hammersen & Eisenried, 2014, S. 343).

Mitunter wird das Verbot des Redlining damit gerechtfertigt, dass schlicht die prognostische Kraft von Anschriftendaten in Abrede gestellt wird: „Solange man nicht weiß, wer im Penthaus wohnt und wer im Souterrain, bleiben Vorhersagen unscharf“ (von Lewinski, 2018, § 31 BDSG Rn. 42.1). Soweit das stimmt, wirft das Verbot der ausschließlichen Verwendung von Anschriftendaten keine besonderen Schwierigkeiten auf – der Gesetzgeber hat es hier dem Scoring-Unternehmen nämlich „nur“ untersagt, sein Scoring-Verfahren maßgeblich auf eine Datenkategorie zu stützen, die zu einer guten Prognoseleistung ohnehin nichts beizutragen vermag.

Bloß wird man es eben nicht von vornherein ausschließen können, dass die Anschrift einer Person ein Datum ist, das eine ausreichende Prognose über die Kreditausfallwahrscheinlichkeit erlaubt. Das Verbot des Redlining hat deshalb eine genuine Fairness-Dimension und dient nicht allein der Qualitätsgewährleistung. Das Verbot zielt darauf ab, die Dynamik einer vermuteten sich selbst erfüllenden Prophezeiung zu durchbrechen: „Schlechte Risikoeinstufungen führen zu schlechteren Konditionen und höheren finanziellen Belastungen, die wiederum eher ein Scheitern der Erfüllung von Zahlungsverpflichtungen bewirken.“ (Korczak & Wilken, 2008, S. 24; siehe auch Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014).

Das Redlining könnte durchaus beispielhaften Charakter haben. Jede Merkmalsverwendung zum Zweck des prognostischen Scorings kann geeignet sein, für die Merkmalsträger eine Dynamik in Gang zu setzen, die zur Verfestigung benachteiligender sozialer Lagen führt und die es deshalb zu verhindern gilt (Kim, 2017).

3.4 Gutes Scoring im Spannungsverhältnis zum Diskriminierungsverbot

Ein Scoring-Verfahren soll nicht nur eine hohe Güte aufweisen. Diese Güte soll auch allen gescorten Personen(gruppen) gleichermaßen zugutekommen, es soll „diskriminierungsfrei gut“ sein. Mit dieser Forderung ist die Brücke geschlagen zu der lebhaften Diskussion, die – ganz vorwiegend von Informatikern – unter dem Label der „algorithmischen Fairness“ geführt wird (Gesellschaft für Informatik, 2018). Im *algorithmic-fairness*-Diskurs werden die Fairnessanforderungen, die man an (unter anderem) prognostische Scoring-Verfahren stellen kann, mathematisch formalisiert. Inzwischen hat die Diskussion eine Vielzahl an Kriterien („Fairnessmaße“) entwickelt, anhand deren man bewerten kann, inwieweit man ein Scoring-Verfahren als fair bezeichnen darf (näher unten B.VI.). Die Diskussion analysiert, wo diese

Fairnessmaße in Widerspruch zueinander stehen und nicht gleichzeitig erfüllt werden können (zur Einführung siehe Zweig & Krafft, 2018).²⁵ Die Frage, welche Fairnessmaße wichtiger sind als andere, woran sich dann die Entwickler von prognostischen Scoring-Verfahren zu halten haben, kann nicht anhand mathematischer Kriterien entschieden werden. Die Entscheidungen über Fairness-Prioritäten, in der Literatur auch „tragische Entscheidungen“ genannt (Calabresi & Bobbitt, 1978), muss die Gesellschaft diskutieren und der Gesetzgeber muss sie gegebenenfalls verbindlich treffen.

²⁵ Beispielhaft für dieses Dilemma steht der auch in Deutschland viel diskutierte COMPAS-Algorithmus, der in der US-amerikanischen Strafjustiz verwendet wird. COMPAS gibt eine Prognose ab, wie hoch das Risiko ist, dass ein Straftäter in den nächsten zwei Jahren erneut straffällig wird (Moll, 2016). Dabei werden 137 Fragen über eine Person beantwortet (mit Informationen aus direkter Befragung oder aus Polizeiakten), die Antworten mit Datenbanken über bereits verurteilte Straftäter abgeglichen und ein Risikoquotient errechnet. 2016 zeigte die New Yorker Stiftung ProPublica in einer aufwendig angelegten Studie, dass der verwendete Algorithmus gegenüber Afroamerikanern in der Hinsicht diskriminierend ist, dass er sie deutlich häufiger fälschlicherweise für rückfallgefährdet erachtet als Weiße (also als *false positives* klassifiziert), gemessen an ihrer tatsächlichen Rückfallquote (Angwin, Larson, Mattu & Kirchner, 2016). Mehrere Wissenschaftler traten dem Diskriminierungsvorwurf mit einem anderem Fairnesskriterium entgegen: Für Straftäter mit demselben Risikoscore ist die Wahrscheinlichkeit, erneut straffällig zu werden, dieselbe, egal ob sie Afroamerikaner oder Weiße sind (Angwin & Larson, 2016).

4. Unerwünschte Ungleichbehandlungen durch Scoring jenseits der Diskriminierung

Zu unterscheiden sind Diskriminierungen von anderen „unerwünschten Ungleichbehandlungen“. Damit sind Unterscheidungen gemeint, die vom Verbraucher als unfair oder sachfremd wahrgenommen werden und die daher nach seinem Empfinden nicht zum Kriterium für Entscheidungen gemacht werden dürfen. Das geltende Recht wird an dieser Stelle regelmäßig kein Problem erkennen können: Die Privatrechtsordnung wird ganz überwiegend so gedeutet, dass Entscheidungen, die Private treffen, heteronom vorgegebenen Rationalitätsmaßstäben nur ausnahmsweise genügen müssen. Nicht zu vernachlässigen ist zwar der Umstand, dass das Privatrecht von zahlreichen sachgebietsspezifischen Gleichbehandlungsgeboten – vom Arbeitsrecht über das Kapitalmarktrecht bis zum Wettbewerbsrecht – durchsetzt ist (siehe eingehend Grünberger, 2013). „Ohne vernünftigen Grund“ anders behandelt worden zu sein als jemand anderes, kann das Privatrecht aber regelmäßig nicht als Problem erfassen, weil es für Entscheidungen nicht „objektive“ Rationalitätsmaßstäbe, sondern den Willen der beteiligten Parteien für entscheidend erachtet. Privatautonomie bedeutet gerade die Freiheit des Entscheiders, für seine Entscheidungsmaßstäbe gegenüber der Rechtsordnung nicht rechenschaftspflichtig zu sein, mögen andere diese Maßstäbe auch als idiosynkratisch, unvernünftig oder moralisch anstößig bewerten. In den Worten des Bundesverfassungsgerichts: Es gibt kein Rechtsprinzip, „wonach die Rechtsbeziehungen zwischen Privaten von diesen prinzipiell gleichheitsgerecht zu gestalten wären. Grundsätzlich gehört es zur Freiheit jeder Person, nach eigenen Präferenzen darüber zu bestimmen, mit wem sie unter welchen Bedingungen Verträge abschließen will.“ (BVerfG, Beschluss des Ersten Senats vom 11. April 2018 – 1 BvR 3080/09 –, 1. Leitsatz). Es ist eine offene Frage, ob diese Sichtweise in Ansehung ubiquitärer Datenerhebung und Persönlichkeitsprofilbildung der Anpassung und Fortentwicklung bedarf.

Denn diese Sichtweise kann mit Fairnesserwartungen der Verbraucher in Konflikt geraten. Welche Handymarke ein Verbraucher verwendet, sollte aus seiner Sicht für den Preis, den er für eine mit diesem Handy gebuchte Hotelübernachtung zahlen soll, nichts zu tun haben (Townley, Morrison & Yeung, 2017; Zander-Hayat, Reisch & Steffen, 2016; Zander-Hayat, Domurath & Gross, 2016). Welchen Inhalten er in sozialen Netzwerken ein „like“ gibt, welche er teilt und kommentiert, sollte für die Beurteilung seiner Bonität außer Betracht bleiben. Auf den Bereich des Scorings bezogen, bedeutet das: Gehen in die Ermittlung eines Score-Wertes Daten ein, die von der gescorten Person oder der Allgemeinheit als sachfremd wahrgenommen werden, dann ist damit zumindest die Frage aufgeworfen, ob dies regulatorisch hinzunehmen oder für Abhilfe zu sorgen ist.

Hierbei handelt es sich aber nicht um die Frage von Diskriminierung im eingangs des Kapitels dargelegten Sinne, weil es an einer Unterscheidung nach „klassischen“ Diskriminierungsmerkmalen fehlt: Man wird die Benutzer von Apple-Produkten nicht als eine typischerweise benachteiligte Gruppe beschreiben können. Vielmehr wird die Frage zu adressieren sein, wo gesamtgesellschaftlich legitimierte Grenzen der Ungleichbehandlung in der digitalen Welt im Allgemeinen und beim Scoring im Besonderen bestehen, die nicht durch bestehendes Antidiskriminierungsrecht erfasst werden.

III. Rechtsdurchsetzung

In Recht gegossene Fairnessanforderungen an Scoring lassen sich nur verwirklichen, wenn neben zivilrechtlichen Rechtsdurchsetzungsinstrumenten auch effektive behördliche Kontrollstrukturen bestehen. Bereits in seinem Gutachten Verbraucherrecht 2.0 hat der SVRV eingehend dargelegt, dass es untunlich ist, die Durchsetzung des Verbraucherrechts – insbesondere das Lauterkeitsrecht, das Recht der Allgemeinen Geschäftsbedingungen und das Antidiskriminierungsrecht – allein in die Hände der Verbraucher zu legen (SVRV, 2016). Für diesen Befund lassen sich mehrere Gründe anführen (siehe auch Podszun, Busch & Henning-Bodewig, 2018). In gerichtlichen Verfahren befinden sich Verbraucher typischerweise in einer Position faktischer Unterlegenheit gegenüber einer Unternehmer-Prozesspartei. Gründe hierfür sind zum Beispiel größere Budgetrestriktionen und geringere Rechtskenntnis von Verbrauchern, aber auch eine geringere Prozesserfahrung (Fries, 2016). Dazu kommt, dass die Streitwerte in Verbraucherrechtsstreitigkeiten üblicherweise gering sind. Deshalb haben Verbraucher oft nur ein geringes wirtschaftliches Interesse an dem Gegenstand des Streites und zögern, die mit einer individuellen Rechtsdurchsetzung verbundenen Mühen und Risiken auf sich zu nehmen. In dieses Bild fügt sich bruchlos ein, dass spektakuläre gerichtliche Erfolge einzelner Verbraucher – man denke nur an den „Fall Schrems“ zur Durchsetzung der Garantien des unionsrechtlichen Datenschutzrechts (EuGH, Urteil [GK] vom 6. Oktober 2015 – Rs. C-362/14 – [ECLI:EU:C:2015:650], s. dazu SVRV, 2016, S. 60, 70) – klar akzentuiert auf eine Prozessführung im öffentlichen Interesse zurückzuführen sind, nicht auf die Kosten-Nutzen-Abwägung eines auf seinen individuellen Vorteil bedachten Verbrauchers.

Auch unter systemischen Gesichtspunkten erscheint ein individualzentriertes „Rechtsdurchsetzungsregime“ (Hellgardt, 2016, S. 549 ff.; 560 ff.) als unbefriedigend (SVRV, 2016). Die erhebliche Dauer zwischen dem erstmaligen Auftreten eines Problems im Verbraucheralltag und seiner gerichtlichen Lösung, der transnationale Charakter zahlreicher Rechtskonflikte in der digitalen Welt (dazu eingehend Calliess, 2006) und die Entterritorialisierung des Rechts im globalen Internet sprechen dagegen, die Rechtsverfolgung vornehmlich dem betroffenen Verbraucher selbst aufzugeben. Weiterhin kann der Verbraucher stets nur einen schmalen, individualisierten Ausschnitt unternehmerischer Tätigkeit erken-

nen und bewerten. Zum Beispiel lassen sich Werbung, Vertragsangebote, Preise und Vertragskonditionen durch Algorithmen individualisieren. Ob dieser Individualisierung eine verbotene Diskriminierung zugrunde liegt, kann kaum verlässlich festgestellt werden, solange man den Vorgang aus der Perspektive des einzelnen Verbrauchers betrachtet (siehe auch unten E.III.5).

Private Verbraucherschutzverbände teilen die genannten Schwächen des Individualverbrauchers nicht. Doch auch sie sind zur Rechtsdurchsetzung als alleinige Institution nicht in der Lage. Ihnen fehlen die Kompetenzen staatlicher Einrichtungen, die gegenüber Unternehmen die Offenlegung von Informationen verlangen können, denen unter Umständen weitreichende Ermittlungs- und Eingriffsbefugnisse zustehen und die in der Europäischen Union in transnationale Behördennetzwerke eingebettet sind, was einen Austausch von Informationen und Erfahrungen begünstigt (SVRV, 2016).

Die Durchsetzung rechtlicher Anforderungen an faires Scoring ist deshalb eine Aufgabe, die nicht allein dem Verbraucher aufzubürden, sondern auch staatlichen Stellen anzuvertrauen ist. Auf der Grundlage dieses Befundes gelangen Fragen des Zuständigkeitszuschnitts, der Organisation und der Personalausstattung der hierfür in Betracht kommenden staatlichen Stellen in das Blickfeld. Der SVRV hat – für die Durchsetzung einer Vielzahl von Verbraucherrechten in der digitalen Welt – ausländische Vorbilder aus den USA, Großbritannien und den Niederlanden dargestellt (SVRV, 2016) und die Fachkompetenzen und die Expertise umrissen, die eine deutsche Einrichtung aufbauen und pflegen müsste, um effektive Rechtsdurchsetzung betreiben zu können. Für die Errichtung einer solchen „Digitalagentur“ (SVRV, 2016; siehe auch BMWi/BMJV, 2015; BT-Drucks. 19/1982, S. 8) sieht der SVRV einen dringenden politischen Handlungsbedarf und hat Lösungsoptionen vorgeschlagen.

IV. Score-Qualität

1. Güte des dem Score zugrundeliegenden Algorithmus

Es besteht Dissens zwischen den gesetzlich vorgegebenen Anforderungen an ein Scoring-Verfahren, die auf den weitgehend unbestimmten Rechtsbegriffen in § 31 Abs. 1 Nr. 2 BDSG beruhen (siehe näher E.III.3), im Vergleich zu wissenschaftlich diskutierten Standards in der empirischen Sozialforschung. Aus rechtlicher Sicht ist bemerkenswert, dass die Rechtsprechung bislang nur wenig Gelegenheit hatte, den unbestimmten Rechtsbegriff zu konkretisieren. Verbraucher sind offensichtlich kaum oder nur sehr selten bereit, ihre Ansprüche vor Gericht durchzusetzen. Ohne einschlägige Urteile fehlt es aber an der notwendigen argumentativen Untermauerung der offen formulierten gesetzlichen Vorgaben. Unabhängig davon, mit welchen statistischen Verfahren Score-Anbieter in der Praxis arbeiten und welche Qualitätsregeln diese für sich selbst aufgestellt haben, muss festgestellt werden, dass allein durch die genannte Bestimmung eine hinreichende Güte des Scorings im Hinblick auf Fairness und Qualität nicht gewährleistet ist.²⁶ Das gilt allein schon deswegen, weil es an höchstgerichtlicher Rechtsprechung fehlt, die helfen würde zu verstehen, was mit wissenschaftlich anerkannten Standards in concreto gemeint ist.

Beispielhaft für „wissenschaftlich anerkannte“ Standards, die für Veröffentlichungen von empirischen Ergebnissen im Rahmen eines Peer-Review Verfahrens in der empirischen Sozialwissenschaft empfohlen werden, seien genannt: Kenntlichmachung oder Übersendung aller verwendeten Merkmale bzw. eine Zusammenfassung dieser (Durchschnitt, Standardabweichung etc.) im Sinne der Nachvollziehbarkeit der Schätzergebnisse, transparente Herleitung der Ergebnisse und Darstellung dieser (einschließlich der Verbrauchermerkmale und des Zielmerkmals, Signifikanz der Verbrauchermerkmale, Einfluss der Verbrauchermerkmale auf das Zielmerkmal, Güte des Schätzmodells), Begutachtung von Ergebnissen und idealerweise des gesamten Schätzverfahrens durch qualifizierte externe Gutachter.

Mit Blick auf die Berechnung von Scores stehen üblicherweise folgende maschinellen Lernverfahren zur Verfügung: logistische Regression, Cluster-bildende Methoden (z. B. kMeans), Entscheidungsbäume, Ensemble-Methoden (z. B. Boosting, Random Forest) und testhalber auch Deep Learning, etwa mit neuronalen Netzen (Lessmann, Baesens, Seow & Thomas, 2015; Thomas, Crook & Edelman, 2017).

Im Bereich des Bonitäts-Scorings ist bekannt, dass üblicherweise die logistische Regression verwendet wird und neuere Methoden wie z. B. neuronale Netze zwar getestet, jedoch nicht operativ eingesetzt werden (Schröder et al., 2014; Thomas et al., 2017; Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014).

²⁶ Abgerufen am 16. Mai 2018 von URL <https://www.br.de/nachrichten/luecken-im-pruefsystem-von-auskunfteien-100.html>.

Exkurs: Auszug aus einem stark simplifizierten Algorithmus²⁷

Ein Algorithmus ist keine „Magie“, sondern beschreibt lediglich den Prozess einer maschinellen Datenanalyse, die auf etablierten – und oft simplen (wie im Falle der logistischen Regression) – Schätzverfahren aus der Statistik beruht (James, Witten, Hastie & Tibshirani, 2013; Jentsch, 2018). Auch hochgradig nichtlineare Algorithmen, so auch neuronale Netze, lassen sich auf ihre Struktur hin untersuchen und sie können immer besser approximativ als Entscheidungsbäume oder Regressionsgleichungen dargestellt werden (vgl. zum Beispiel Montavon, Samek & Müller, 2018). Insofern ist Streit darüber müßig, ob neuronale Netze immer – oder nur für spezielle Fälle – verständlich erklärt werden können (vgl. Abschnitt B.1.4 oben).

Die logistische Regressionsanalyse testet, ob ein Zusammenhang zwischen mehreren unabhängigen Variablen (hier: Verbrauchermerkmale) und einer binären (d. h. 0 oder 1) abhängigen Variable (hier: Zielmerkmal) besteht. Genauer: Ob die Verbrauchermerkmale einen Einfluss auf die Wahrscheinlichkeit haben, dass das Zielmerkmal den Wert 1 annimmt bzw. wie stark der Einfluss der individuellen Ausprägungen der Verbrauchermerkmale darauf ist (ausgedrückt durch den Regressionskoeffizienten).

Untersucht werden kann hiermit beispielsweise die Frage:

„Was beeinflusst die Wahrscheinlichkeit, dass ein Verbraucher seinen Kredit nicht zurückzahlen wird (Y= Kreditausfallrisiko)?“

Die Anzahl der Girokonten eines Verbrauchers (X_1), die Höhe der laufenden Kredite (X_2), oder die Adresse des Verbrauchers (X_3)?“

Technisch lässt sich dieses ausdrücken durch die Formel:

$$P(Y = 1|X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}$$

Aus der Anwendung der logistischen Regression lassen sich ableiten:

- Die Signifikanz des Einflusses eines Verbrauchermerkmals auf das Zielmerkmal (z. B. berechnet mittels des sogenannten Wald-Tests). Auf dieser Basis wird entschieden, ob ein Verbrauchermerkmal überhaupt in die Score-Berechnung mit einbezogen wird.
- Die Größe des Einflusses eines Verbrauchermerkmals auf das Zielmerkmal (ausgedrückt durch den Regressionskoeffizienten). Hieraus leitet sich in der Regel die Gewichtung der Verbrauchermerkmale in der Score-Berechnung ab.
- Die Güte des gesamten Modells, welche auch als Trennschärfe bezeichnet wird (ausgedrückt z. B. durch den sogenannten Gini-Koeffizienten).

²⁷ Siehe beispielsweise Auer & Rottmann (2015) und Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe (2014).

Grundsätzlich ist anzumerken, dass schwer durchschaubare „künstliche Intelligenz“ bislang – zumindest in Deutschland – für Scoring keine Rolle spielt. Das liegt wahrscheinlich auch daran, dass die Möglichkeiten künstlicher Intelligenz jenseits der Bild- und Spracherkennung weit weniger eindrucksvoll sind als dies von den Entwicklern von KI behauptet wird. Es liegt eine ausführliche Literatur vor, die zeigt, dass weniger die Scoring-Methode für die Güte des Scorings eine Rolle spielt als die (fehlende) Qualität der eingehenden Daten (einschliesslich der Entity Recognition, also der Identifizierung von Personen, vgl. dazu einen anekdotischen Bericht von Seibt (2018) und Abschnitt V unten) und die Stabilität des gescorten Prozesses (vgl. Hand, 2005; Hand, 2006; Verbeke, Dejaeger, Martens, Hur & Baesens, 2012; und – trotz eines gewissen Optimismus hinsichtlich der Leistungsfähigkeit von KI – Lessmann et al., 2015). Verbeke et al. (2012) kommen deswegen zu der Schlussfolgerung, dass einfach nachvollziehbare Entscheidungsbäume komplizierten Verfahren vorgezogen werden könnten, nicht zuletzt weil bestimmte Entscheidungsbäume bei vergleichbarer Vorhersagegüte (z. B. Phillips, Neth, Woike & Gaissmaier, 2017) mit typischerweise wenigen Merkmalen auskommen (Jenny, Pachur, Williams, Becker & Margraf, 2013; für die analytische Begründung siehe Brighton & Gigerenzer, 2015). Neben Investitionen in neue Schätzverfahren sollte demnach umso mehr in die Messqualität bereits zum Einsatz kommender Verfahren investiert werden.

Die faktische Verwendung vergleichsweise einfacher und nachvollziehbarer Scoring-Algorithmen bedeutet im Übrigen nicht zwingend, dass die eingesetzten Algorithmen transparent gemacht würden (was aus wissenschaftlicher Sicht unschwer möglich wäre).

Welche Algorithmen im Bereich der Kfz-Telematik Einsatz finden, ist aus öffentlich zugänglichen Quellen nicht ersichtlich (die Frage danach wird in Kapitel C aufgegriffen). Bei der Berechnung von Gesundheits-Scores im Bereich der Krankenversicherungen finden komplexe algorithmische Entscheidungsverfahren kaum Anwen-

dung – so basiert die Berechnung der Gesamtpunktzahl von Bonuspunkten im Bereich der gesetzlichen Krankenversicherung auf einer simplen Addition einzelner Bonuspunkte. Andere Anbieter des zweiten Gesundheitsmarkts wie Vitality²⁸ oder Dacadoo²⁹ verfügen über ausdifferenziertere Bepunktungssysteme.

Die beschriebenen Modelle dienen dazu, mithilfe einer Vielzahl von Verbrauchermerkmalen wie Wohnort, Kreditverträgen etc. (im Falle des Bonitäts-Scorings), Brems- und Beschleunigungsverhalten, Fahrtzeit und -ort etc. (im Falle der Kfz-Telematik) Verbraucherverhalten bzw. mögliche Konsequenzen von Verbraucherverhalten wie die Kreditausfallwahrscheinlichkeit bzw. die Unfallwahrscheinlichkeit (hier: Zielvariablen) zu prognostizieren.

Wie „gut“ ein Algorithmus im Einsatz sein kann, hängt dabei von der Wahl des Algorithmus selbst ab, dem verwendeten Zielmerkmal, der Wahl der Verbrauchermerkmale sowie weiteren Modellparametern (z. B. der relativen Gewichtung von zu negativen und zu positiven Prognosen). Ein Verfahren lässt sich dann als sehr gut qualifizieren, wenn es Verbraucher hinsichtlich ihres Risiko-Verhaltens überwiegend korrekt klassifiziert, z. B. in „gute“ und „schlechte“ Risiken hinsichtlich des Kreditausfallrisikos oder der Unfallwahrscheinlichkeit.³⁰

Im Kontext der Score-Berechnung stehen drei Kategorien von Gütemaßen zur Verfügung: Maße wie der Gini-Koeffizient bzw. Area Under The Curve (AUC), welche die Trennschärfe (Discriminatory Power) eines Modells beschreiben;³¹ Maße wie der Brier-Score, welche die Genauigkeit der Prognosekraft eines Modells beschreiben; sowie Classification Error (z. B. Lessmann et al., 2015).

Aus verbraucherpolitischer Perspektive besteht das Problem dabei im Wesentlichen darin, dass Gütemaße weder von Score-Anbietern berichtet werden, noch gibt es gesetzliche Mindestvorgaben für die Güte von Scoring-Verfahren. Die Minimal-Anforderung des Gesetzgebers (§ 31 Abs. 1 Nr. 2 BDSG) ist bereits dann erfüllt,

28 Abgerufen am 24. Mai 2018 von URL https://www.generalivitality.de/vmp/punkte_und_status/tipps_zum_punktesammeln.

29 Abgerufen am 24. Mai 2018 von URL <https://www.dacadoo.com/?lang=de>.

30 Einschränkung ist zu sagen, dass es keine komplett fehlerfreien Methoden und Entscheidungen gibt. Dies gilt für Algorithmen-basierte Entscheidungen wie für menschliche Klassifizierungen, wie sie z. B. von Kreditsachbearbeitern von Banken vorgenommen werden.

31 Für ein perfektes Vorhersageinstrument hätte der Gini-Koeffizient den Wert 1,00. Ein Wert von 0,00 bedeutet, dass nur zufällig vorhersagt werden kann.

wenn eines der eingangs beschriebenen algorithmischen Entscheidungsverfahren überhaupt angewandt wird – eine aus Sicht der empirischen Sozialforschung denkbar schwache Qualitätsanforderung (siehe näher unter E.III.3).

Ein weiteres mögliches Problem, welches sich aus einer bestimmten Auswahl von Verbrauchermerkmalen für die Berechnung von Scores ergibt, ist die sogenannte Multikollinearität von Daten. Je nachdem, welche Merkmale zur Berechnung eines Scores herangezogen werden, kann nicht ausgeschlossen werden, dass einzelne Merkmale stark mit anderen Merkmalen korrelieren – im Falle eines Bonitäts-Scores könnte dies beispielsweise für die Anzahl von Girokonten und die Verschuldung eines Verbrauchers gelten (hypothetisches Szenario: „Ein Verbraucher mit vielen Girokonten ist der Tendenz nach stärker verschuldet.“). Korrelieren diese beiden Merkmale untereinander, so kann die Bestimmung des tatsächlichen Beitrags eines Merkmals zur statistischen Prognose des Zielmerkmals erschwert werden. Mit anderen Worten: Da eine Unterscheidung beider Merkmale aus statistischer Sicht schwer fällt, ordnet der Algorithmus Veränderungen des Scores, die z. B. eigentlich auf Veränderungen in der Anzahl von Girokonten zurückzuführen sind, Veränderungen in der Verschuldung zu oder umgekehrt (siehe auch Auer & Rottmann, 2015). Theoretisch kann es durch Multikollinearität sogar dazu kommen, dass die Einflussrichtung eines Merkmals auf den Score-Wert im statistischen Modell umgekehrt wird: Obwohl der Einfluss eines Merkmals auf die Kreditwürdigkeit zum Beispiel positiv ist, würde in diesem Fall durch den Algorithmus ein negativer Zusammenhang geschätzt (Schröder et al., 2014).

Ein weiterer Aspekt mit Blick auf die Güte eines Algorithmus ist die mögliche Verschlechterung der Prognosekraft eines Algorithmus aufgrund struktureller Veränderungen, sogenannter exogener Effekte. So ist anzunehmen, dass aufgrund fortwährender gesellschaftlicher Veränderungen mit wachsenden Zeitabständen die Stabilität der in Scoring-Modellen ermittelten statistischen Zusammenhänge abnimmt (Hand & Henley, 1997). Die Prognosekraft eines statistisch berechneten Scorewerts kann sich demnach im Zeitverlauf ändern. Auch können wiederkehrende konjunkturelle Schwankungen Einfluss auf die Verlässlichkeit von Bonitäts-Scores ausüben (Schröder et al., 2014). Im Gesundheitsbereich könnte eine strukturelle Verschlechterung von Gesundheits-Scores durch den demografischen Wandel und den sich im Durchschnitt verschlechternden Gesundheitszustand der Bevölkerung stattfinden. Scoring-Verfahren müssen also in regelmäßigen Abständen gegen die Einflüsse exogener Effekte abgesichert werden. Zudem können in alten Daten überkommene soziale Normen enkodiert sein (z. B. die Prävalenz von männlichen Chefärzten), die durch Modelle, die auf nicht aktualisierten Daten aufbauen, (unwillentlich) perpetuiert werden (siehe etwa Lowry & Macpherson, 1988).

2. Der Nutzen neuerer und komplexerer Algorithmen

Score-Anbietern ist es freigestellt, jederzeit Änderungen innerhalb eines bestehenden algorithmischen Entscheidungsverfahrens vorzunehmen (z. B. durch Austausch verwendeter Daten oder Anpassung der Gewichtungsparemeter) oder das bestehende Entscheidungsverfahren insgesamt zu wechseln, wie z. B. durch den Wechsel von einer logistischen Regression zu einem neueren Machine Learning Verfahren (vgl. auch Abschnitt B.I.4 oben).

Im Bereich der Kfz-Versicherung hält sich ein Anbieter beispielsweise vor, Änderungen der Gewichtung in Abhängigkeit neuer Ergebnisse aus Unfallforschung oder dem absolvierten Pilotprojekt vorzunehmen³² – mit der Konsequenz, dass sich die „Spielregeln“ des Scores (z. B. Anpassung von für den Score relevanter Daten bzw. deren Gewichtung zueinander) ändern und Verbraucher möglicherweise ihr Verhalten an ein neues Bewertungsschema anpassen müssen (sofern ihnen dieses überhaupt bekannt ist). Anpassung des Algorithmus sind grundsätzlich insbesondere dann begrüßenswert, wenn sie zur Steigerung der Score-Güte beitragen: So hat eine aktuelle Analyse des Fahrverhaltens von Teilnehmern des Smart Driver Programms der HUK Coburg ergeben, dass ein „deutlicher Zusammenhang zwischen hoher Geschwindigkeitsüberschreitung und Unfallhäufigkeit“ erst bei einer Überschreitung der erlaubten Fahrgeschwindigkeit von >30 km/h besteht. Vermutet wurde zuvor eine geringere Schwelle der Geschwindigkeitsüberschreitung.³³

Mit Bezug auf den Wechsel zu neuartigen algorithmischen Entscheidungsverfahren wie z. B. neuronalen Netzen wird kritisiert, dass diese selbst für ihre Entwickler einen „Blackbox-Charakter“ haben und – anders als verhältnismäßig simple Schätzverfahren wie die logistische Regression – schwerer nachvollziehbar seien, da die Analyse des Codes nicht weiterhilft (16. TB Hess LReg. 2003, 21; 17. TB Hess LReg. 2005, 10.; cf. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Es gibt jedoch inzwischen Verfahren die Black Box durch systematisches Testen aufzuhellen (Gigerenzer, Wagner & Müller, 2018).

Problematisch sind neue Verfahren nicht an sich, sondern die Leistungsfähigkeit der Aufsichtsbehörden, wie etwa der Landesdatenschutzbeauftragten und des Bundesversicherungsamts, die bislang – anders als die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) – kaum über die technische und personelle Ausstattung verfügen, um algorithmische Entscheidungsverfahren materiell prüfen zu können.

³² Abgerufen am 15. Mai 2018 von URL <https://www.sparkassen-direkt.de/telematik/faq/>.

³³ Schriftliche Stellungnahme von HUK Coburg vom 10. September 2018 (liegt dem SVRV vor).

V. Datengrundlage

1. Korrektheit, Aktualität und Vollständigkeit

Fehler in der Datengrundlage – sowohl bei der Erstellung eines Scores wie bei dessen operativen Anwendung – können sich unmittelbar auf die Güte eines Algorithmus auswirken. Im Extremfall ist der Score, der einem Verbraucher zugesprochen wird, schlichtweg falsch, weil Personen verwechselt werden. Auf Qualitätsprobleme der Daten bei der Entwicklung eines Scores gehen wir in den folgende Abschnitten näher ein. Hier sollen zuerst die wenig erforschten Fehlerquellen bei der Anwendung eines Scores diskutiert werden.

Mögliche Fehlerquellen, die sich aus der Literatur sowie Verbraucherbeschwerden ableiten lassen, sind bei der Anwendung von Bonitäts-Scores fehlende Daten, veraltete Daten und Verwechslung von Personen etc. (z. B. Schröder et al., 2014). Im Bereich der Kfz-Telematik wird berichtet, dass Daten entweder falsch aufgezeichnet wurden bzw. sich Fehler in den Daten durch eine Verknüpfung mit fehlerhaften Daten ergaben. So wurde Verbrauchern ein zu geringer Score aufgrund veralteten Kartenmaterials zugewiesen (siehe hierzu Kapitel C): In diesen Fällen wurde eine Geschwindigkeitsüberschreitung aufgezeichnet, die tatsächlich keine war: Der Fahrer fuhr mit erlaubter Geschwindigkeit – jedoch in einem Bereich, in dem sich einige Zeit zuvor eine Baustelle befand, mit entsprechend reduzierter maximaler Fahrtgeschwindigkeit.

Bezüglich der Qualität der sogenannten Entity Recognition (also der Zuordnung der richtigen Daten zu einer Person), die immer wieder im Mittelpunkt anekdotischer Evidenz steht, über die in der Presse berichtet wird (vgl. z. B. Seibt, 2018), ist wissenschaftlich wenig bekannt. Verbeke et al. (2012) heben aus Sicht der Ent-

wickler von Algorithmen die Bedeutung der Datenqualität allgemein hervor und messen einer besseren Datenqualität eine höhere Priorität zu als ausgefeilteren Algorithmen zur Berechnung von Scores.³⁴ Britz (2008) berichtet, wenn auch ohne Beleg im Einzelnen, von der Bedeutung schlichter Eingabefehler.

Aus Datenschutzgründen werden persönliche Informationen so wenig wie möglich erfasst. Dies führt zu Herausforderungen in der eindeutigen Zuordnung zwischen Mensch und Datensatz. Dies gilt erschwert für nach Deutschland zugewanderte Menschen (z. B. aus Ländern wie Bulgarien, Russland, Indien, Thailand), deren Namen aus den Personaldokumenten in die lateinische Schrift transkribiert werden muss. Und es ist z. B. naheliegend, dass es bei Umzügen an einen neuen Wohnort zu vermehrten Problemen bei der Entity Recognition kommen kann, nämlich dann wenn ein Score-Anbieter ggf. zunächst einen neuen Datensatz anlegen muss, der aufgrund einer möglicherweise fehlenden Datenverknüpfung keine Informationen über die Bonitäts-Historie einer Person enthält. Falls dann die Bonität dieser Person zunächst durch Geo-Scoring approximiert werden muss, kann es sein, dass dieser Person dadurch ein ungünstigerer Score zugewiesen wird (siehe auch B.V.2.).

Die Herausforderungen sind insbesondere Anbietern von Bonitäts-Scores bekannt und werden, wie im Falle der SCHUFA, durch ein großes Team von Spezialisten täglich bearbeitet. Die Zahl der Nachfragen und Beschwerden bei der SCHUFA ist auch außerordentlich gering – und Nachfragen und Beschwerden gehen sicher nicht nur wegen fehlerhafter Entity Recognition ein. Insofern kann das Problem der Entity Recognition quantitativ nicht allzu groß sein – es trägt jedoch in jedem Einzelfall dazu bei, dass Scoring bei vielen einen schlechten Ruf hat.³⁵

34 Bei Techniken maschinellen Lernens sind Fehler in den Trainingsdaten zudem besonders misslich, da diese über ein inkorrektes Modell alle Prognosen verzerren, auch wenn die jeweiligen Inputdaten, mit denen das Modell gefüttert wird, korrekt sind.

35 Einen weiteren Eindruck bekommt man, wenn man beispielsweise die Zahl der Fälle beim Privatkunden ServiceCenter der SCHUFA betrachtet. Im Tätigkeitsbericht des Ombudsmannes wird dazu berichtet: „Täglich treffen [...] etwa 1.000 neue Fragen, Anmerkungen und auch Beschwerden von Verbrauchern zu den im SCHUFA-Datenbestand gespeicherten Informationen ein. In der Abteilung [...] dreht sich vieles um die Richtigkeit der Daten. Vier Teams mit zusammen 75 Sachbearbeitern – darunter Juristen, Rechtsanwaltsgehilfen, Notariatsfachangestellte und Bankkaufleute – prüfen, ob die Hinweise und Anmerkungen der Kunden zutreffend sind und Einträge gegebenenfalls geändert oder gelöscht werden müssen“ (SCHUFA Holding AG, 2018, S. 36f.). Bezieht man die 1000 Anfragen pro Tag auf die etwa 400.000 Anfragen pro Tag (ibid.) ergibt sich eine Problemquote von 2,5 Promille. Diese ist sehr gering, stellt aber auch mit Sicherheit die Untergrenze der tatsächlichen Datenprobleme dar, da diese den gesorten Personen vielfach nicht auffallen dürften.

Von Praktikern wurden wir bezüglich potenzieller Probleme der Entity Recognition auf die gesamte Kette der Beteiligten bei der Verwendung des SCHUFA-Scores hingewiesen: Auch wenn der Score-Anbieter den korrekt einer Person zugeordneten Score zurückliefert, kann eine fehlerhafte Zuordnung noch auf Seiten der Abfragenden passieren. So ist es möglich, dass der gelieferte Score im eigentlichen Entscheidungsprozess nicht korrekt zugeordnet wird, da bei Banken, Versicherungen, aber auch im Versandhandel und im E-Commerce eine Vielzahl von Datenquellen für die Geschäftsentscheidung eingesetzt werden. Sobald mehr als eine Datenquelle Verwendung findet, ist eine korrekte Entity Recognition nicht trivial und ohne eine eindeutige Identitäts-Kennziffer fehleranfällig. Forschungsbedarf ist hier offenkundig angezeigt, um zuerst einmal die tatsächliche Größenordnung des Problems besser abschätzen zu können.

2. Verwendung von Proxy-Variablen

Paradoxerweise verfügen die meisten Anbieter von Bonitäts-Scores, mit Ausnahme der SCHUFA, nicht über detaillierte Informationen über die individuelle Kreditlage von Verbrauchern. Stattdessen weichen diese Anbieter oft auf soziodemografische bzw. mikrogeografische Daten zurück, um beispielsweise aus der Beschaffenheit des Wohnumfelds Schlussfolgerungen über die Zahlungsmoral einzelner Verbraucher zu schließen (Kamp & Weichert, 2005). So wird beispielsweise die Zahlungsmoral einzelner Verbraucher aus der durchschnittlichen Anzahl negativer Merkmale (z. B. eingeleitete Inkassoverfahren, gerichtliche Mahnverfahren, Zwangsvollstreckungsverfahren) im selben Wohnhaus sowie der durchschnittlichen Anzahl negativer Merkmale pro Haushalt in derselben Straße abgeleitet – das sogenannte Geo-Scoring. Trotz des Verbots der abschließlichen Nutzung von Anschriftendaten (§ 31 Abs. 1

Nr. 3 BDSG, siehe näher E.I.1) wurde im Fall der Hamburger Auskunftei Bürgel im Jahr 2017 gerichtlich festgestellt, dass diese den Score eines Kunden ausschließlich aus der Wohnanschrift des Kunden ableitete.³⁶ Darüber hinaus wurden Fälle berichtet, in denen der Vorname eines Verbrauchers herangezogen wurde, um daraus Rückschlüsse auf dessen Alter zu ziehen (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014).

Ein ähnlicher Fall tritt in der Kfz-Telematik auf, bei der Nacht- und Stadtfahrten oft negativ in den Score eingehen.³⁷ Selbst wenn für die gesamte Versichertenpopulation nachgewiesen werden konnte, dass ein Kausalzusammenhang zwischen der Tageszeit bzw. dem Ort der Autofahrt und der Unfallwahrscheinlichkeit besteht und dieser Zusammenhang intuitiv erscheint (längere Reaktionszeit und möglicherweise Trunkenheit am Steuer in der Nacht bzw. höheres Verkehrsaufkommen in der Stadt und höheres Stresslevel als bei Landfahrten), lässt sich daraus nicht ein Kausalzusammenhang für die individuelle Unfallwahrscheinlichkeit ableiten. Besonders erfahrene Stadt-Autofahrer und Nachtschicht-Arbeiter werden möglicherweise zu Unrecht negativ bepunktet und unter Umständen sogar mittelbar diskriminiert (siehe auch B.II.).

Derartige Variablen, welche ein Verbrauchermerkmal nicht direkt messen, sondern indirekt über andere verfügbare Daten approximieren, werden auch als Proxy-Variablen (auch: Stellvertreter-Variable) bezeichnet. In der empirischen Sozialforschung ist allgemein anerkannt, dass auf Proxy-Variablen dann ausgewichen werden kann, wenn eine Eigenschaft nicht oder nicht mit vertretbarem Aufwand zugänglich ist. Ein Ausweichen auf Proxy-Variablen muss allerdings gut begründet sein, sie müssen mit dem fehlenden Verbrauchermerkmal hochkorreliert sein und die aus der Verwendung von Proxy-Variablen entstehenden Limitationen in der Aussagekraft des gesamten Modells müssen transparent gemacht werden (siehe für die Verwendung von Proxy-Variablen im Kontext des Bonitäts-Scoring z. B. Berg, Burg, Gombović & Puri, 2018).

³⁶ AG Hamburg, Urteil vom 16. März 2017 – 233 OWI 12/17. Siehe dazu die Meldung von Heise Online, abgerufen am 15. Mai 2018 von URL <https://www.heise.de/newsticker/meldung/Datenschutzverstoss-15-000-Euro-Bussgeld-wegen-Geoscoring-3664654.html>.

³⁷ Abgerufen am 24. Mai 2018 von URL <https://www.sparkassen-direkt.de/telematik/faq/>.

Werden Proxy-Variablen hingegen zur Prognose individuellen Verbraucherverhaltens eingesetzt – mit zum Teil erheblichen wirtschaftlichen Konsequenzen für Verbraucher –, ist die Verwendung solcher Variablen für das Scoring viel stärker begründungspflichtig. Das Risiko einer völligen Fehlbewertung angesichts der Individualität des Betroffenen ist hier am größten, der sogenannte Ökologische Fehlschluss (Kamp & Weichert, 2005). Danach wird auf Basis von Aggregatdaten auf unzulässige Weise auf Individualdaten geschlossen: Aus Informationen über die Lebensverhältnisse der Nachbarschaft einer gescorten Person werden Rückschlüsse auf die finanzielle Situation dieser Person allgemein und auf die Ausfallwahrscheinlichkeit eines Kredits konkret gezogen (Kamp & Weichert, 2005). Aus einer Korrelation, die in der gesamten Population vorzuliegen scheint, wird ein Kausalzusammenhang auf individueller Ebene abgeleitet.

Dies ist aus Sicht eines Unternehmens, das Zahlungsausfälle vermeiden will (und sich an weniger Umsatz nicht stört) hinnehmbar und sinnvoll, nicht jedoch aus Sicht eines Individuums, das fehlerhaft bewertet wird. Im Falle des Geo-Scorings werden Verbraucher durch einen Ökologischen Fehlschluss demnach in Mitverantwortung für das Fehlverhalten ihrer Nachbarn genommen und in ihrer Souveränität eingeschränkt. Daraus folgt, dass eine Person ihren eigenen Score-Wert nicht durch eine (graduelle) Änderung ihres eigenen Verhaltens verbessern kann, sondern dieses nur durch eine invasive Maßnahme wie den Umzug in eine „bessere“, d. h. im Durchschnitt solventere, Nachbarschaft möglich gemacht wird.

3. Gewichtung von Merkmalen

Welcher Score einem Verbraucher zugewiesen wird, hängt im Wesentlichen von der Ausprägung der individuellen Daten und der Gewichtung der Daten zueinander ab. Daraus kann im Extremfall folgen, dass Verbraucher, die sich mit Blick auf die meisten verwendeten Verbrauchermerkmale „Score-steigernd“ verhalten, dennoch einen schlechten Score erhalten, weil andere Daten ein unverhältnismäßig hohes Gewicht erhalten. Es lassen sich drei unterschiedliche Gewichtungsschemata unterscheiden:

Gewichtung anhand der Regressionsparameter: In einem algorithmischen Entscheidungsverfahren ergibt sich die Gewichtung eines Verbrauchermerkmals aus dem Einfluss des Verbrauchermerkmals auf das Zielmerkmal (dieses Gewicht wird oft in Form von Parametern in Regressionsgleichungen bestimmt, kann aber auch anders bestimmt werden, siehe Exkurs und auch Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Da Anbieter von Bonitäts-Scores die Gewichtung der Verbrauchermerkmale als Teil ihres Geschäftsgeheimnisses betrachten, was der BGH bestätigt hat, gibt es jedoch kaum Erkenntnisse darüber, wie diese zustande kommen und inwieweit diese objektiv sind. Konkrete gesetzliche Vorgaben zur Ableitung von Merkmalsgewichten gibt es nicht.

Heuristik aus der unternehmerischen Praxis: Im Bereich der Kfz-Telematik (im Beispiel der Tarif S-Drive von der Sparkassen Direktversicherung) wird oft eine Daumenregel („Heuristik“) verwendet, wonach die Verbrauchermerkmale Fahrweise (Beschleunigung und Bremsen), Geschwindigkeitsüberschreitungen, Nachtfahrten und Stadtfahrten mit den sehr eingängigen Gewichten 40 %, 30 %, 20 % bzw. 10 % in den Score eingehen.³⁸ Inwieweit diese Gewichte den relativen Einfluss eines der beschriebenen Verbrauchermerkmale auf das Zielmerkmal in Relation zum Einfluss aller anderen Verbrauchermerkmale widerspiegeln, lässt sich nicht sagen.

38 Abgerufen am 23. Mai 2018 von URL <https://www.sparkassen-direkt.de/telematik/faq/>.

Modellunabhängig: Im Bereich der Bonusprogramme ist bekannt, dass sich die Höhe zu erwerbender Bonuspunkte nicht zwingend nach dem gesundheitsförderlichen Effekt einer Aktivität (Verbrauchermerkmal) zur Gesundheitsverbesserung eines Verbrauchers (Zielmerkmal) richtet („modellunabhängig“), sondern z. B. nach dem Aufwand, der für den Verbraucher mit Absolvieren der Aktivität verbunden ist. Da es hierzu kaum Erkenntnisse gibt, ist eine vertiefende Diskussion in Kapitel C vonnöten. Der (statistische) Zusammenhang zwischen Verbrauchermerkmal und Zielmerkmal ist mit Blick auf den Gewichtungsfaktor also weitgehend entkoppelt.

Negative Folgen durch die Gewichtung von Merkmalen ergeben sich für Verbraucher möglicherweise dann, wenn 1) Gewichtungen der zuständigen Aufsichtsbehörde oder ggf. sogar dem Verbraucher gegenüber nicht transparent sind und der Zusammenhang zwischen eigenem Verhalten und deren Auswirkungen auf den Score nicht absehbar sind; 2) sich Gewichtungen zwischen unterschiedlichen Score-Anbietern erheblich unterscheiden: ein Verbrauchermerkmal, das bei einem Score-Anbieter ein hohes Gewicht hat, ist bei einem anderen Anbieter möglicherweise irrelevant – dies ist insbesondere dann der Fall, wenn Gewichtungen sich nicht nach objektivierbaren Kriterien wie beispielsweise der Höhe der Regressionsparameter richten; 3) Gewichtungen sich über die Zeit ändern.

VI. Konkurrenz zwischen Fairness-Kriterien

Ein weiteres Problem von Scoring tritt im Zusammenhang mit der konkreten Zusammensetzung von Datensätzen auf, die zur Bestimmung eines Scores benutzt werden. Im Bereich des maschinellen Lernens nennt man einen solchen Datensatz, der für statistische Analysen genutzt wird, „Trainings-Datensatz“ (Details zu den Grundlagen maschinellen Lernens finden sich in IV.1-2 und in Gesellschaft für Informatik, 2018: Abschnitt 4.1). Wichtig ist hierbei, dass sich Spezifika der Datengrundlage unmittelbar auf die Struktur und Vorhersage des Modells auswirken. Beispielsweise könnten in einem Datensatz über Kreditwürdigkeit drei Viertel aller als kreditwürdig erachteten Personen weiblich sein und nur ein Viertel männlich. Aus dieser Ungleichgewichtung der Datengrundlage (der Anteil beider Geschlechter beträgt in der Realität – in diesem Fall ohne Berücksichtigung „anderer“ Geschlechter – etwa 50 / 50) wird jede Statistiksoftware („Lernalgorithmus“) mit hoher Wahrscheinlichkeit „schließen“, dass andere Merkmale als das Geschlecht nur eine untergeordnete Rolle spielen und die Wahrscheinlichkeit der Kreditwürdigkeit vor allem vom Geschlecht der Antrag stellenden Personen abhängt (Gesellschaft für Informatik, 2018). Dieses Phänomen wird in der Literatur auch als Bias Amplification bezeichnet (Zhao, Wang, Yatskar, Ordonez & Chang, 2017; cf: Gesellschaft für Informatik, 2018).

Fairness bei der Berechnung von Scores ist ein grundsätzliches Problem, das unabhängig von der konkret benutzten Methode auftritt – auch Daumenregeln und „Heuristiken“ können unfair sein. Der Literaturzweig des

Fair Machine Learning bringt diese grundsätzlichen Probleme am besten auf den Punkt: Es werden quantitative Methoden diskutiert, welche die möglichst umfassende Gleichbehandlung einzelner Gruppen oder auch Individuen sicherstellen sollen (Dwork, Hardt, Pitassi, Reingold & Zemel, 2012; Gesellschaft für Informatik, 2018; Kleinberg, Mullainathan & Raghavan, 2016). Stichwortartig genannt seien hier: Gleiche Genauigkeit (Overall Accuracy Equality), statistische Parität (Statistical Parity), gleiche bedingte Gruppengenauigkeit (Conditional Procedure Accuracy Equality), gleiche bedingte Vorhersagegenauigkeit (Conditional Use Accuracy Equality) und gleiches Fehlerverständnis (Treatment Equality).

Berk et al. (2017) fassen die Vereinigung aller fünf Begriffe von „algorithmischer“ Fairness im Konzept der „Total Fairness“ zusammen (cf: Gesellschaft für Informatik, 2018). Vollkommene Fairness durch Anpassungen im Algorithmus zu erreichen, ist jedoch mit großer Wahrscheinlichkeit nicht möglich, denn die unterschiedlichen Fairness-Kriterien stehen in Konkurrenz zueinander, in dem Sinne, dass niemals alle Fairness-Kriterien gleichzeitig erfüllt sein können. Zu diesem Schluss kommen Chouldechova (2017) und Kleinberg et al. (2016) in einer Analyse von drei Fairnessmaßen, in der sie zeigen, dass aktuell keine Methode existiert, welche diese drei quantitativen Fairnessmaße zugleich erfüllt. Mit großer Wahrscheinlichkeit wird es aufgrund unterschiedlicher Risiken (Prävalenz) zwischen verschiedenen Gruppen auch niemals eine Methode geben, die alle Fairness-Kriterien simultan erreichen kann.

Konkurrenz zwischen unterschiedlichen Fairness-Maßen: Ein numerisches Beispiel³⁹

Angenommen sei, dass die tatsächlichen Risiken für Zahlungsausfälle sich stark für zwei verschiedene Gruppen unterscheiden, etwa Niedrig- und Hochverdiener: 5 % Wahrscheinlichkeit vs. 0,5 % Wahrscheinlichkeit für Zahlungsausfall. Damit ergeben sich bei gleicher Prediktionskraft eines Scores für beide Gruppen (hier seien 90 % richtige Prognosen angenommen) folgende wahren bzw. falschen Prognosen für je 10.000 Personen in jeder Gruppe:

Für die Hochrisiko-Gruppe gibt es 500 Zahlungsausfälle (5 % von 10.000), von denen 450 von dem Score erkannt und 50 übersehen werden. Von den 9.500 Fällen, die keinen Zahlungsausfall haben (95 % von 10.000), werden aufgrund der 90-prozentigen Prognosegüte 8.550 korrekt prognostiziert, aber 950 werden falsch als Zahlungsausfall markiert. Der Anteil korrekter Prognosen von Zahlungsausfällen für die Hochrisiko-Gruppe ist folglich $450 / (450 + 950) = 32 \%$.

Für die Niedrig-Risikogruppe gibt es 50 Zahlungsausfälle (0,5 % von 10.000), von denen 45 von dem Score erkannt und 5 übersehen werden. Von den 9.950 Fällen, die keinen Zahlungsausfall haben (99,5 % von 10.000), werden aufgrund der 90-prozentigen Prognosegüte 8.995 korrekt prognostiziert, aber 955 werden falsch als Zahlungsausfall markiert. Der Anteil korrekter Prognosen von Zahlungsausfällen ist folglich $45 / (45 + 955) = 4 \%$.

Der Unterschied im Anteil korrekter Prognosen ergibt sich daraus, dass es bei einem sehr kleinen Risiko sehr viele nicht-riskante Fälle gibt, die falsch klassifiziert werden (False Positive). Der Anteil korrekter Prognosen für Zahlungsausfälle wäre in beiden Gruppen nur ungefähr gleich, wenn der Algorithmus für die Niedrig-Risikogruppe im Hinblick auf den Prognosefehler 10-fach genauer als für die Hoch-Risikogruppe wäre, also die Prediktionskraft bei 99 vs. 90 % läge. Dies ist ein sehr unwahrscheinlicher Fall und im Allgemeinen nicht gegeben.

Daraus folgt für das Beispiel: Werden beide Gruppen im Hinblick auf die Spezifität = Sensitivität gleich behandelt (Conditional Procedure Accuracy Equality), unterliegt die Niedrig-Risikogruppe einem viel größeren False-Positive-Risiko (Conditional Use Accuracy Equality).

Die Implementierung von quantitativer Fairness hat zudem einen Nebeneffekt: Wählt man alle Verbrauchermerkmale aus, die einen statistisch signifikanten Effekt auf das Zielmerkmal ausüben (was moderne Verfahren des maschinellen Lernens automatisch machen), können dies auch diskriminierende und somit rechtlich geschützte Merkmale oder eng damit zusammenhängende Merkmale sein. Werden deswegen diskriminierende Merkmale aus dem statistischen Modell entfernt, wird das statistische Modell insgesamt ungenauer. Je mehr Merkmale aufgrund ihres Zusammenhangs mit der Gruppenzugehörigkeit entfernt werden, umso mehr Genauigkeit geht verloren und die Güte des statistischen Modells nimmt ab (Gesellschaft für Informatik, 2018). Hierdurch besteht also ein unlösbarer Zielkonflikt zwischen der Verwendung geschützter Merkmale, insofern diese einen signifikanten Einfluss ausüben, und der Güte des Scores: Werden statistisch signifikante Merkmale nicht benutzt, werden mehr Personen unzutreffend gescort. Es folgt wiederum daraus, dass widerstreitende Fairnessmaßstäbe vorliegen, deren simultane Erreichung im Allgemeinen nicht möglich ist. Es muss ein Optimum anhand von Prioritäten der Fairness gefunden werden. Welche Fairnessmaße bevorzugt werden bzw. nachrangig sein sollen, kann nicht durch Mathematik und maschinelles Lernen entschieden werden. Es bedarf hierzu einer gesellschaftlichen Übereinkunft über legitime Ziele und Verwendungen von Merkmalen.

³⁹ Vgl. auch Begriffserklärungen und Beispielrechnungen in Gesellschaft für Informatik (2018: Abschnitte 4.3.1 und 4.3.2).

VII. Verbraucher und Gesellschaft: Erwartungen, Wissen, Kompetenzen und Implikationen

Dieses Kapitel gibt einen Überblick über den Stand der Forschung zu den Verbrauchererwartungen bzw. der Akzeptanz von Scoring sowie zum scoring-relevanten Wissen und zur Kompetenz von Verbrauchern in Deutschland. Vorweggenommen sei, dass es in Deutschland bislang kaum unabhängige wissenschaftliche Studien gibt, die das Wissen und die Kompetenzen von Verbrauchern im Zusammenhang mit etabliertem (z. B. Bonität) und potenziell neuartigen Scoring-Verfahren (z. B. im Bereich Gesundheit, Verknüpfung von Scores aus verschiedenen Bereichen) sowie die damit assoziierten Implikationen untersuchen (vgl. für Ausnahmen Fischer & Petersen, 2018, die sich jedoch auf algorithmische Entscheidungen im Allgemeinen beziehen, sowie Müller-Peters & Wagner, 2017; PricewaterhouseCoopers, 2018). Deswegen hat der SVRV eine repräsentative Befragung in Auftrag gegeben (siehe Kapitel D), deren Ziel es u. a. war, das scoringbezogene Wissen sowie die Akzeptanz von Scoring in verschiedenen Lebensbereichen der in Deutschland lebenden Bevölkerung zu erfassen.

1. Erwartungen und Akzeptanz von Verbrauchern im Zusammenhang mit Scoring

Um eine an den berechtigten und gesellschaftlich legitimen Erwartungen von Verbrauchern orientierte Verbraucherpolitik für Scoring gestalten zu können, sollten diese Erwartungen zunächst in Erfahrung gebracht werden. In Anbetracht der Tatsache, dass es sich jenseits des traditionellen Bonitäts-Scorings beim Scoring in anderen Bereichen (z. B. Kfz-Telematik, Gesundheit) um ein vergleichsweise junges Phänomen handelt, das erst nach und nach Einzug in verschiedene Lebensbereiche von Verbrauchern erhält, sind unabhängige, aussagekräftige wissenschaftliche Studien, die Verbrauchereinstellungen bzw. -erwartungen im Zusammenhang mit Scoring beleuchten, in Deutschland bislang rar. Daher scheint es sinnvoll empirisch zu erheben, in welchen Bereichen, in welchem Umfang und in welcher Form Scoring von Verbrauchern in Deutschland als legitim angesehen bzw. was als nicht legitim erachtet wird.

Welches sind beispielsweise Merkmale, die aus Sicht der Verbraucher für die Beurteilung der Kreditwürdigkeit, der Tarifierung der Kfz- und Krankenversicherung als legitim und gerechtfertigt angesehen werden bzw. welche nicht? Darüber hinaus stellt sich die Frage, ob und falls ja, in welchem Ausmaß, Verknüpfungen von Scores und Merkmalen aus verschiedenen Lebensbereichen von Verbrauchern eher befürwortet oder abgelehnt werden. Eine ganz andere Frage ist, ob das, was die Verbraucher selbst als berechtigt und legitim erachten auch von der Gesellschaft insgesamt geteilt wird. Das herauszufinden obliegt in einer demokratischen Gesellschaft letztendlich dem parlamentarischen Gesetzgeber, der die moralischen und ethischen Vorstellungen in rechtliche Regeln übersetzen muss oder eben auch von einer Regelung Abstand nimmt.

In Bezug auf das traditionelle Bonitäts-Scoring zeigt sich, dass die Akzeptanz der übermittelten Scores vergleichsweise niedrig zu sein scheint. So erachtet in einer repräsentativen Studie über die Hälfte bis zu drei Viertel der Befragten ihren Score als nicht gerecht, wobei die Akzeptanz der Scores davon abhängt, bei welchem Unternehmen die Eigenauskunft über die persönliche Bonität eingeholt wurde (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Tendenziell ist die Akzeptanz bei höherem, d. h. günstigerem Score ausgeprägter, jedoch werden auch hohe Scores nicht zwangsläufig als gerecht empfunden. Über die Gründe für diesen Umstand kann nur spekuliert werden: Beispielsweise geben in derselben Befragung knapp die Hälfte der Befragten an, dass sie die Erläuterungen der Auskunftsteien nicht ausreichend und häufig unverständlich fänden. Dieses mag dazu beitragen, dass Bonitäts-Scores als nicht gerecht empfunden werden. Über 80 % der Befragten wünschen sich zudem mehr Transparenz und Informationen von den Auskunftsteien und befürworten eine aktive Auskunftspflicht von Auskunftsteien (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014).

Eine neuartige Methode zur Bestimmung der Bonität, bei der Daten aus sozialen Netzwerken genutzt werden (sog. *social scoring*), wird von über der Hälfte der Befragten eher als riskant eingeschätzt (56 % von n = 1.023) wie sich 2018 in einer repräsentativen Studie ge-

zeigt hat (PricewaterhouseCoopers, 2018). Dabei geben die Befragten mehrheitlich an (71 %), die Gefahr eines Fehlschlusses in Bezug auf die Bonitätsbewertung mittels der Nutzung von Daten aus sozialen Netzwerken zu sehen. Unter Erfüllung bestimmter Transparenzkriterien (z. B. Offenlegung des Scoring-Ergebnisses, Information darüber, welche Daten zum Zwecke des Scorings genutzt werden) würden jedoch über die Hälfte der jüngeren Befragten (18–25 Jahre) *social scoring* befürworten (PricewaterhouseCoopers, 2018).

Die Deutsche Akademie der Technikwissenschaften (acatech) hat Ende 2017 eine repräsentative Umfrage (n = 2.002) zum Thema Technik mit dem Schwerpunkt Digitalisierung durchführen lassen (acatech & Körber-Stiftung, 2018). Insgesamt werden digitale Technologien von der Mehrheit der Befragten eher skeptisch beurteilt. Betrachtet man beispielsweise das autonome Fahren, so zeigt sich, dass von der Mehrheit der Befragten vor allem die Besorgnis um die Datensicherheit geäußert wird. Zudem wird die Sammlung personenbezogener Daten durch das Fahrzeug von knapp zwei Dritteln der Befragten als negativ bewertet. Da durch Telematiktarife vergleichbare Daten aufgezeichnet werden, steht deren Akzeptanz auch in Frage.

Ein diffuses Unbehagen im Zusammenhang mit der Preisgabe persönlicher Daten lässt auf eine grundsätzliche Skepsis gegenüber Scoring-basierten Geschäftsmodellen schließen, die Big-Data-Analysen zur Grundlage haben. Beispielsweise befürchten zwei Drittel der Befragten in Deutschland, dass Unternehmen im Internet zu viele persönliche Daten sammeln. Die größten Sorgen, die jeweils von fast 80 % der Befragten geäußert wurden, bestehen darin, dass persönliche Informationen von Unternehmen ge- und verkauft werden, mangelnder Datenschutz in Bezug auf persönliche Daten herrscht und in der Gefahr, dass persönliche Informationen überwacht werden könnten (Center for International Governance Innovation, 2017). Nicht ohne weiteres nachvollziehbar ist jedoch, warum dennoch so viele Verbraucher in die Erfassung und Verarbeitung ihrer Daten relativ bereitwillig einwilligen. Die Praxis der

Datenschutzbeauftragten ist keineswegs von Tausenden von Beschwerden von Verbrauchern geprägt, die sich gegen die halb erzwungene Einwilligung zur Wehr setzen. Zwischen praktischem Handeln und politischen Wunschvorstellungen besteht ein erhebliches Spannungsverhältnis.

Der Akzeptanz und Erwartung von Verbrauchern im Zusammenhang mit (potenziell zukünftigem) neuartigem Scoring im Bereich telematikbasierter Kfz-Versicherung und verhaltensbasierter Krankenversicherung widmete sich eine Umfrage der Technischen Hochschule Köln.⁴⁰ Knapp die Hälfte der Befragten (46 % von n = 834) kann es sich vorstellen, Daten über das Fahrverhalten aufzeichnen zu lassen und an die Kfz-Versicherung weiterzugeben, um damit einen deutlich günstigeren Versicherungstarif zu erhalten. Zudem werden beeinflussbare Merkmale wie vorsichtiges Fahren für die Tarifierung von der Mehrheit der Befragten eher akzeptiert, wohingegen nicht durch die eigene Fahrweise beeinflussbare Faktoren wie das Fahren bei Tag vs. Nacht als Tarifierungsmerkmal von der Mehrheit eher abgelehnt werden. Bei der Krankenversicherung findet es die Hälfte bis zwei Drittel der Befragten gerecht, beeinflussbare Verhaltensweisen wie die Teilnahme an Früherkennungsuntersuchungen sowie Rauchen und Alkohol bei der Tarifierung zu berücksichtigen. Nicht beeinflussbare Merkmale, wie die genetische Vorbelastung für bestimmte Erkrankungen, als Tarifierungsmerkmal stoßen jedoch bei der überwiegenden Mehrheit auf Ablehnung. Grundsätzlich würde gut ein Drittel der Befragten an einem verhaltensbasierten Krankenversicherungstarif teilnehmen, sofern sich daraus eine Preisersparnis ergeben würde (Müller-Peters & Wagner, 2017).

Insgesamt ist die Tendenz erkennbar, dass Verbraucher prinzipiell dazu bereit sind, persönliche verhaltensbasierte Daten dem Versicherungsunternehmen zur Verfügung zu stellen, insbesondere dann, wenn eine Preisersparnis erzielt werden kann (Müller-Peters & Wagner, 2017). Welche Merkmale von Verbrauchern für die Berücksichtigung in einem Versicherungstarif als akzeptabel erachtet werden und welche nicht, scheint

40 Die Befragung wurde von einem der Versicherungsbranche nahen Institut durchgeführt; tendenziell sind die Ergebnisse jedoch mit unserer eigenen Befragung (Kapitel D) vergleichbar.

vom Einsatzbereich (Kfz vs. Gesundheit) sowie der Beeinflussbarkeit der Merkmale (z. B. genetische Vorbelastung vs. Alkoholkonsum) abzuhängen.

Bislang bleibt jedoch offen, inwiefern individuelle Faktoren wie die persönliche Betroffenheit (z. B. gesundheitlicher Zustand), der sozioökonomische Status, demographische Variablen und spezifische Einstellungen beispielsweise zum Datenschutz und zur Technik sowie Kontrollüberzeugungen die Einstellung und die Akzeptanz der Verbraucher zu Scoring moderieren. Um ein differenziertes Bild der Verbrauchererwartungen und -akzeptanz in Bezug auf Scoring zu erhalten und damit passgenaue verbraucherpolitische Maßnahmen ergreifen zu können, ist es wichtig, spezifische Verbrauchergruppen zu identifizieren.

Versicherungen, die erste Ansätze von Scoring nutzen, werben häufig ausschließlich mit positiven Konsequenzen, also einem Bonussystem, in dem durch ein bestimmtes Verhalten Punkte gesammelt werden können und bei einer bestimmten Punktezahl Sachprämien oder ein Rabatt auf den Versicherungstarif in Aussicht gestellt werden.

Der umgekehrte Fall, dass bei einem bestimmten Verhalten negative Konsequenzen drohen, also ein Malussystem, in dem beispielsweise der Versicherungstarif verhaltensabhängig teurer werden kann, wird bislang bei den Telematik-Bausteinen eines Versicherungstarifs nicht praktiziert. Die Unfallhäufigkeit ist bei den Telematik-Bausteinen bislang kein Merkmal, das für den Score herangezogen wird. Auch bei den durch die GKV angebotenen Bonusprogrammen gibt es ausschließlich Boni und eine Nichtteilnahme an Maßnahmen führt nicht zu Sanktionen im Sinne einer Verteuerung des Versicherungstarifs, was gesetzlich durch § 65a Abs. 3 SGB V geregelt ist. Es liegt jedoch nahe, dass die Einstellung und Akzeptanz der Verbraucher zu verhaltensbasierten Tarifen auch in Abhängigkeit der in Aussicht gestellten Konsequenzen variiert. Wenn bei einem Telematiktarif stets Vergünstigungen in Aussicht gestellt werden (z. B. eine Vergünstigung der Versicherungsprämie bei Einhaltung der zugelassenen Höchstgeschwindigkeit), fällt die Akzeptanz vermutlich anders aus als wenn so ein Tarif bei Nichteinhaltung bestimmter Vorgaben (z. B. nicht nur, dass es keine Vergünstigung gibt, sofern die zugelassene Höchstgeschwindigkeit überschritten wird,

sondern dass dafür auch mehr gezahlt werden muss) auch mit finanziellen Einbußen einhergeht. Daher war es ein weiteres Ziel der vom SVRV in Auftrag gegebenen repräsentativen Bevölkerungsbefragung zu erfassen, inwieweit sich ein Bonus- vs. Malussystem bei verhaltensbasierten Tarifen im Bereich Kfz und Gesundheit auf die Akzeptanz bzw. Ablehnung solcher Tarife auswirkt (siehe Kapitel D).

2. Wissen und Kompetenzen

2.1. Verbraucher und Algorithmen: Wissen und Einschätzungen

Einer aktuellen repräsentativen Umfrage (n = 1.221) der Bertelsmann-Stiftung zufolge (Fischer & Petersen, 2018) liegen bei der Mehrheit der in Deutschland lebenden Bevölkerung mangelndes Wissen, Unentschlossenheit und Vorbehalte bezüglich der Nutzung von Algorithmen vor. Da Scoring-Modellen i. d. R. Algorithmen zugrunde liegen, werden nachfolgend die Ergebnisse dieser Umfrage auszugswise dargestellt.

Obwohl drei von vier Befragten in der Studie angeben, schon einmal von dem Begriff „Algorithmus“ gehört zu haben, hat knapp die Hälfte der Befragten spontan keine Vorstellung davon, was darunter zu verstehen ist. Bei denjenigen, die zumindest schon einmal von dem Begriff Algorithmus gehört haben, wissen jedoch über die Hälfte nichts über die prinzipielle Funktionsweise von Algorithmen. Lediglich ein Zehntel dieser Personen gibt nach eigenem Verständnis an zu wissen, wie Algorithmen funktionieren. Ob die Befragten etwas mit dem Begriff Algorithmus verbinden und um dessen Funktionsweise wissen, variiert hierbei stark mit dem Alter, der Bildung und dem Geschlecht: Befragte mit Abitur können deutlich häufiger eine zumindest vage Vorstellung von Algorithmen formulieren als diejenigen mit Volks- oder Hauptschulabschluss, männliche Befragte häufiger als weibliche und Personen unter 45 Jahren eher als Personen über 60 Jahren.

Ferner zeigt sich eine bereichsspezifische Bandbreite dahingehend, wie bewusst den Befragten der Einsatz von Algorithmen ist: Während mehr als der Hälfte der Befragten bewusst war (bzw. angab, sich dessen bewusst zu sein), dass Algorithmen bei der individuellen Werbung im Internet angewendet werden und der Hälfte bzw. knapp der Hälfte bewusst war, dass Algorithmen bei der Gesichtserkennung im Rahmen von Videoüberwachung sowie bei der Bewertung der Kreditwürdigkeit Verwendung finden, ist sich etwa nur ein Drittel bewusst darüber, dass in manchen Regionen Algorithmen zur Personalbedarfsanalyse bzw. zu Einsätzen der Polizei verwendet werden, indem ermittelt wird, welche Gebiete besonders einbruchgefährdet sind (*predictive policing*). Dass Algorithmen auch von der Justiz zur Beurteilung der Rückfallwahrscheinlichkeit von Straftätern eingesetzt werden können, war hingegen nur knapp einem Fünftel der Befragten bewusst.

Gut ein Drittel der für die Bertelsmann-Studie Befragten sieht bei algorithmenbasierten Entscheidungen eher die Risiken, wohingegen nur knapp ein Fünftel der Befragten eher Chancen sieht (Fischer & Petersen, 2018). Knapp die Hälfte der Befragten sind unentschieden dahingehend, ob algorithmenbasierte Entscheidungen eher Chancen oder Risiken bergen. Demnach scheint ein großer Teil der in Deutschland lebenden Bevölkerung zu diesem Thema gegenwärtig noch keine klare Meinung dazu ausgebildet zu haben. Dieser Umstand sollte nicht überraschen, da es bislang für viele algorithmenbasierte Entscheidungen keine klare Evidenz für das tatsächliche Nutzen-Schaden-Verhältnis gibt und diese Frage selbst unter Experten umstritten ist.

Eine deutliche Meinung der Befragten ist erkennbar, wenn es darum geht, ob Entscheidungen grundsätzlich von Algorithmen oder Menschen getroffen werden sollten: Die sehr große Mehrheit (79 %) der Befragten (n = 1.221) hat bei algorithmischen Entscheidungen ein Unbehagen und zieht eine menschliche Entscheidung vor. Bereichsspezifisch zeigt sich ein etwas differenzierteres Bild: Während es insgesamt eine große Ablehnung ausschließlich algorithmenbasierter Entscheidungen gibt, würde die Mehrheit zustimmen, die Entscheidung einer effizienten Ausnutzung und Verwaltung von Lagerflächen Algorithmen zu überlassen. Knapp die Hälfte wäre zudem dafür, die individuelle Werbung im Internet und die Rechtschreibkontrolle bei der Textverarbeitung aus-

schließlich durch eine algorithmische Entscheidung vornehmen zu lassen. Hingegen befürwortet die Mehrheit der Befragten, dass die Bewertung der Kreditwürdigkeit, die Diagnose von Krankheiten und die Beurteilung des Rückfallrisikos von Straftätern als reine menschliche Entscheidung bzw. allenfalls algorithmenunterstützt durchgeführt werden sollte. Die finale Entscheidung sollte durch einen Menschen getroffen werden. Demnach ist insbesondere in sensiblen Bereichen wie der Bonität, bei strafrechtlichen und gesundheitlichen Entscheidungen die Mehrheit gegen eine (ausschließlich) algorithmenbasierte Entscheidung.

Im Einklang mit der eher negativen Einstellung gegenüber Algorithmen wird eine stärkere Kontrolle von Algorithmen über alle Bildungs- und Altersgruppen hinweg von knapp zwei Dritteln der Befragten befürwortet. Maßnahmen zur Algorithmenkontrolle wie zum Beispiel eine Kennzeichnungspflicht bei Entscheidungen durch Algorithmen, Offenlegung von Algorithmen gegenüber unabhängigen Experten und eine Einführung einer Ethikkommission findet bei der überwiegenden Mehrheit der Befragten Zustimmung (Fischer & Petersen, 2018).

Insgesamt legen die Ergebnisse der zuvor beschriebenen repräsentativen Umfrage nahe, dass in Bezug auf das Thema Algorithmen in Deutschland gegenwärtig große Wissenslücken hinsichtlich des Begriffsverständnisses und der Funktionsweise vorherrschend sind. Der Großteil der Bevölkerung hat sich bisher kaum mit dem Thema „Algorithmen“ auseinandergesetzt, was im Einklang damit steht, dass nur eine Minderheit eine eindeutige Meinung bezüglich Algorithmen im Allgemeinen hat. Gleichzeitig wird der algorithmenunterstützten bzw. der ausschließlich algorithmenbasierten Entscheidungsfindung in einer Vielzahl von Bereichen große Skepsis und Ablehnung entgegengebracht. Folglich scheint für ein besseres Verständnis, eine fundierte Beurteilung von Vor- und Nachteilen im Zusammenhang mit Algorithmen in ihren unterschiedlichen Anwendungsfeldern, ein Ausgleich des Wissensdefizits sowie der Aufbau von Kompetenzen innerhalb der in Deutschland lebenden Bevölkerung sinnvoll.

Daneben sollte ein differenzierter gesellschaftlicher Diskurs über die belegbaren Implikationen von Algorithmen angestoßen werden, um Ängsten, Ablehnung und Herausforderungen zu begegnen. Gleichmaßen bedarf es jedoch auch der Aufklärung über das empirisch begründete Potenzial und die Chancen, die neue Technologien und Algorithmen und damit auch Scoring mit sich bringen.

2.2 Scoringbezogenes Wissen von Verbrauchern

Scoringbezogenes Wissen ist bei Verbrauchern eher gering. So kennen beispielsweise viele US-amerikanische Bürger ihren persönlichen Kreditwert nicht, wobei jeder Zehnte der Befragten den eigenen Score überschätzt und damit auch die eigene Kreditwürdigkeit (Levinger, Benton & Meier, 2011). Zudem haben viele Verbraucher in den USA Schwierigkeiten damit, die genauen Details ihres Kreditwerts nachzuvollziehen. Insbesondere jüngere Verbraucher im Alter zwischen 18 und 34 Jahren wissen wenig über die Details und Implikationen ihres Kreditwerts (Consumer Federation of America & Vantage Score Solutions, 2016). Dies ist von besonderer Relevanz, da ein Feldexperiment in den USA zeigte, dass die Kenntnis über den eigenen Kreditwert mit selteneren Zahlungsverzögerungen und dann wiederum mit einem höheren Score nach einem Jahr verbunden war (Homonoff, O' Brien & Sussman, 2017).

Nicht nur in Deutschland haben Verbraucher oftmals keine Kenntnis über die Auskunftsrechte, die sie in Bezug auf ihre eigene Bonität besitzen. Dies ist insofern nicht allzu überraschend, da insgesamt die Kenntnis der Verbraucher über ihre Rechte gering ist. So weiß beispielsweise über die Hälfte der Befragten einer aus dem Jahr 2014 repräsentativen Umfrage nicht, dass sie den Anspruch haben, einmal jährlich kostenlos eine Selbstauskunft über die eigene Bonität zu erhalten (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014). Darüber hinaus wissen Verbraucher häufig auch nicht, dass die über sie gespeicherten Informationen bei Auskunfteien fehlerhaft bzw. veraltet sein können. Dies ist z. B. der Fall, wenn das Versäumnis einer Bank vorliegt, der Auskunftei zu übermitteln, dass ein Kredit vollständig abbezahlt wurde. Dieser Umstand kann sich im Einzelfall nachteilig auf den Bonitäts-Score auswirken und im

Zusammenhang damit auch darauf, ob bzw. zu welchen Konditionen ein Verbraucher einen (neuen) Kredit erhält. Zudem herrscht bei vielen Verbrauchern auch keine Kenntnis darüber, dass sie ein Anrecht zur Korrektur ihres Scores haben bzw. wo und wie sie dieses umsetzen können. Auch wissen die meisten Verbraucher nicht genau, welche Merkmale Auskunfteien zur Berechnung des Bonitäts-Scores heranziehen (siehe D).

2.3 Kompetenzen im Kontext von Scoring

Digitale Kompetenz ist eine wichtige Voraussetzung für eigenverantwortliche und informierte Entscheidungen. Dies gilt nicht nur in der digitalen Welt im Allgemeinen, sondern auch in Bezug auf Scoring im Besonderen. Obgleich die digitale Welt die Lebensrealität der Verbraucher prägt, ist digitale Kompetenz längst keine Selbstverständlichkeit, sondern muss auch durch die Politik unterstützt und gefördert werden. Darauf hat der Sachverständigenrat für Verbraucherfragen bereits in seinem Gutachten „Digitale Souveränität“ (SVRV, 2017a) aufmerksam gemacht. Digitale Kompetenz ist demnach – neben einer verbraucherfreundlichen Technologie und der rechtlichen Rahmensetzung durch Regulierung – eine wesentliche Komponente eines „Dreiecks der Digitalen Souveränität“. Angemessene Bildung, die Anwendung von Wissen und die Fähigkeit zur Selbstkontrolle sowie lebenslanges Lernen sind wesentliche Bestandteile für die Realisierung digitaler Kompetenz (SVRV, 2017a).

Für die kompetente und kritische Einordnung von Scoring und dessen Implikationen ist es indes unerlässlich, dass Verbraucher auch digitale Risikokompetenz als spezifische Form der Risikokompetenz erlangen. Darunter lassen sich Wissen, Fähigkeiten und Fertigkeiten verstehen, welche ein kritisches Verständnis und den reflektierten Umgang mit Problemen der Unsicherheit in der digitalen Welt ermöglichen. Wie die oben referierten repräsentativen Erhebungen nahelegen, sind diese in vielen Facetten nicht gegeben. Insofern sind informierte Verbraucherentscheidungen, bei denen potenzielle Vor- und Nachteile abgewogen werden sollen, in Frage gestellt. Ein Teil von Verbrauchern ist freigiebig im Umgang mit persönlichen Informationen, wenn sie dafür Vorteile wie Rabatte oder Bonuspunkte erhalten (Geslevich-Packin & Lev-Aretz, 2016), mutmaßlich ohne ein vertieftes Bewusstsein über mögliche negative Kon-

sequenzen. Es scheint nur eine vergleichsweise geringe Anzahl von Verbrauchern die Konsequenzen dieser Abwägung zu überblicken und zudem scheinen sich viele Verbraucher nicht darüber bewusst zu sein, dass Unternehmen *sensible* – von den Verbrauchern häufig bereitwillig zur Verfügung gestellte – Informationen für eigene Interessen und ggf. gegen die Interessen der Verbraucher verwenden können (siehe zusammenfassend Geslevich-Packin & Lev-Aretz, 2016).

Mit konkretem Bezug zum Scoring beruht die Kompetenz eines Verbrauchers auf folgenden Grundpfeilern: scoring-bezogenes Wissen, z. B. über Merkmale, die für das jeweilige Scoring herangezogen werden, Wissen um den Einsatz und Zweck von Scoring (wann, wo, warum) sowie über sich aus dem Scoring ergebende Konsequenzen, aber auch anwendbare Kenntnis darüber, wo und wie man über seinen eigenen Score Auskunft erhält. Konkret würde ein kompetenter Teilnehmer an einem Scoring-Verfahren dessen Qualität kritisch zu prüfen und zu bewerten wissen, das Diskriminierungspotenzial zumindest hinterfragen, und Alternativen finden und nutzen können (wenn diese verfügbar sind).

Ohne ausreichendes Wissen und die Kompetenz, scoring-basierte Entscheidungsverfahren zu verstehen und zu hinterfragen, werden Selbstbestimmung und Selbstkontrolle der Verbraucher eingeschränkt. Notwendig erscheint es, dass Verbraucher für sich persönlich potenzielle Nutzen und Schäden unter verschiedenen Gesichtspunkten (zum Beispiel durch die Herausgabe von persönlichen Daten) verstehen und abwägen können.

Relevant und wichtig ist im Umgang mit Scoring auch die so genannte *data literacy*, also die Fähigkeit planvoll mit Daten umzugehen und sie im jeweiligen Kontext einsetzen und hinterfragen zu können. Dazu gehört unter anderem die Fertigkeit, Daten erfassen, verwalten, analysieren, interpretieren und anwenden zu können (Ridsdale et al., 2015).

Ohne die beschriebenen Voraussetzungen ist nicht zuletzt ein gesellschaftlicher Diskurs über Scoring schwer möglich. Deshalb widmet sich ein Teil des Gutachtens der vorbestehenden Ausgangslage von scoringbezogenem Wissen und der Kompetenz der Bevölkerung.

3. Gesellschaftliche Implikationen

Verbraucherbezogenes Scoring wirkt sich nicht nur auf individueller Ebene aus (z. B. über verhaltensbasierte Versicherungsprämien). Vielmehr kann ein solches Scoring vielschichtige Auswirkungen auf die gesamte Gesellschaft haben. Zudem stellen sich ethisch-moralische Fragen mit Blick auf das Spannungsverhältnis zwischen Individualisierung und Solidarität. Schließlich lassen sich auch grundsätzliche Debatten führen, etwa, inwiefern Grundwerte einer demokratischen Gesellschaft, wie die persönliche Freiheit, Autonomie und Solidarität, durch Scoring herausgefordert werden. Wie bereits oben skizziert (siehe B.II), wird zudem untersucht werden müssen, ob Scoring zur sozialen Ungleichheit beiträgt oder diese vielmehr abbauen hilft und inwiefern bestimmte Verbrauchergruppen systematisch benachteiligt oder übervorteilt werden. Diese Fragen können im vorliegenden Gutachten nicht ausführlich behandelt, sondern nur skizziert werden. Zum gegenwärtigen Zeitpunkt sind verschiedene beschriebene Implikationen eines zunehmenden Scorings für die Gesellschaft in Deutschland noch spekulativ.

3.1 Gesellschaftliche Funktionen

Grundsätzlich dienen bestimmte Scoring-Formen dem Funktionieren der Wirtschaft und sind somit im legitimen Interesse der Gesellschaft. Das traditionelle Bonitäts-Scoring schützt Unternehmen vor Zahlungsausfällen und ist somit eine wesentliche Grundlage für die Kreditvergabe an Verbraucher. Da säumige Zahler eher aussortiert werden, bekommen zuverlässige Verbraucher bessere Konditionen als dies ohne Scoring möglich wäre. Scoringbasierte Entscheidungen können auch dazu beitragen, dass Verbraucher weniger Willkürentscheidungen, die ggf. auf bewussten oder unbewussten Vorurteilen oder Wahrnehmungsverzerrungen von (ausschließlich) menschlichen Entscheidern beruhen, ausgesetzt sind. Für Verbraucher und Unternehmen wird überdies die Vertragsabwicklung bei Dauerschuldverhältnissen erleichtert (z. B. der Abschluss von Handyverträgen), da auf ein im Wesentlichen verlässliches System der Kredit- bzw. Zahlungsfähigkeitsprüfung zurückgegriffen werden kann (Schröder & Taeger, 2014).

Verhaltensbasierte Tarife können hingegen aus normativ-ethischer Perspektive grundsätzlich sehr unterschiedlich bewertet werden – in der Kfz-Haftpflichtversicherung beispielsweise anders als in der Krankenversicherung (siehe E.II.2). Ein System der Beitragsbemessung, liegt der privaten Krankenversicherung (PKV) und der Kfz-Haftpflichtversicherung zugrunde. Dort wird die Versicherungsprämie individuell am Risikoprofil ausgerichtet (siehe E.II.2). In der PKV spielen hierbei beispielsweise Vorerkrankungen und das Alter eine Rolle, in der Kfz-Haftpflichtversicherung Merkmale wie die Dauer der Unfallfreiheit und der Wohnort. Jeder Versicherungsnehmer zahlt also entsprechend seines Risikoprofils. Einige Kfz-Haftpflichtversicherer haben bereits Merkmale, die u. a. das Fahrverhalten betreffen, in telematikbasierten Tarifen eingeführt (z. B. Geschwindigkeitsaufzeichnung). Nach eigenen Angaben der Versicherungen wird ein auf dem Fahrverhalten basierender Score mit einem geringeren Unfallrisiko assoziiert. Versicherungsnehmer mit einem besseren Score zahlen weniger, die Prämie wird hierbei sowohl an das Fahrverhalten direkt als auch an situative Faktoren (Stadt- vs. Landfahrt) angepasst. Hier könnte telematikbasiertes Scoring zu einem gesellschaftlichen Nutzen im Sinne einer Reduktion von unfallassozierten medizinischen Behandlungskosten führen. Sollten telematikbasierte Tarife also tatsächlich das Unfallrisiko reduzieren, würde sich damit nicht nur ein individueller, sondern auch ein gesamtgesellschaftlicher Nutzen ergeben. Bislang wurde ein solcher nicht mit angemessenen wissenschaftlichen Studien belegt.

3.2 Implikationen bezüglich der Solidarität

Scoring und die damit verbundenen Konsequenzen können auf der Ebene des einzelnen Verbrauchers als fair empfunden werden, wohingegen Scoring sich gesamtgesellschaftlich als unfair und ggf. unsolidarisch erweisen kann und umgekehrt: Ist es auf individueller Ebene nicht beispielsweise fair, wenn sich der Beitrag der Versicherungsprämie am freiwilligen Verhalten eines Versicherungsnehmers bemisst, wenn ein Score also davon abhängt wie gesundheitsbewusst sich jemand verhält bzw. wie umsichtig jemand fährt? Ist es nicht unsolidarisch zu rasen und damit potenziell andere Verkehrsteilnehmer zu gefährden oder ggf. durch unfallassozierte medizinische Behandlungen das Gesundheitssystem finanziell zu belasten? Auf individueller Ebene betrachtet, kann man sich beispielsweise auch fragen, warum man

für das Erkrankungsrisiko des rauchenden Nachbarn mitbezahlen soll, wenn man sich selbst gesundheitsbewusst verhält. Umgekehrt sprechen gute Gründe dafür, dass es ein gesellschaftlicher Wert ist, sich für die Freiheit des Nachbarn bei seiner Lebensgestaltung solidarisch zu zeigen und ggf. die durch das Rauchen anfallenden Gesundheitskosten mitzutragen. Ferner können auch Gründe der Datenschutzpräferenz dafür sprechen, sich aktiv dagegen zu entscheiden, gescoret zu werden.

Was im Kontext von Scoring als solidarisch und fair erachtet oder als ethisch-moralisch geboten eingeschätzt wird, muss in einem gesellschaftlichen Diskurs differenziert betrachtet und sorgsam abgewogen werden. Der Schutz Einzelner bzw. von spezifischen Gruppen reflektiert das gesamtgesellschaftliche Interesse einer auf Solidarität fußenden Gesellschaft. Hierbei müssen also die individuellen oder auch gruppenspezifischen Interessen mit den legitimen Interessen der Gesellschaft in Einklang gebracht werden. Das heißt auch, dass besonders schützenswerte Personen oder Gruppen durch Scoring keine (systematische) Benachteiligung erfahren sollen, etwa durch überdurchschnittlich falsch positive Scores. Im gesellschaftlichen Diskurs wäre zu bestimmen, welche Wertvorstellungen und Normen geteilt werden und wie die Interessen Einzelner oder bestimmter schützenswerter Gruppen gegen das legitime gesamtgesellschaftliche Interesse abgewogen werden können.

Die Frage, inwiefern nun Scoring die Solidarität innerhalb der Gesellschaft im Allgemeinen und das auf dem Solidaritätsprinzip fußende System der gesetzlichen Krankenversicherung (GKV, siehe E.II.3) im spezifischen unterminieren könnte, wird bereits intensiv diskutiert (z. B. Deutscher Ethikrat, 2017). Im gesellschaftlichen Sinne lässt sich Solidarität als hohes soziales Gut einer demokratischen Gesellschaft verstehen, das auf Basis geteilter Wertvorstellungen und Normen funktioniert. So besteht beispielsweise Konsens darüber, dass spezifische Gruppen, wie „verletzliche Verbraucher“ (z. B. sozial Schwache, Menschen mit Einschränkungen oder gesundheitlich Beeinträchtigte), in besonderem Maße hilfebedürftig sind, geschützt, gesellschaftlich integriert werden und ihnen soziale und ggf. auch finanzielle Unterstützung zu Teil wird (Micklitz, Oehler, Piorkowsky, Reisch & Strünck, 2010). Im Bereich der Gesetzlichen Krankenversicherung bedeutet das Solidaritätsprinzip, dass sich der Versicherungsbeitrag nicht an dem Ge-

sundheitsstatus oder am Gesundheitsverhalten orientiert, sondern jeweils am individuellen Einkommen des Versicherungsnehmers ausgerichtet. Versicherungsnehmer mit einem guten Gesundheitszustand stehen dabei solidarisch für die Versicherungsnehmer mit einem schlechteren Gesundheitszustand ein (auch wenn der schlechtere Gesundheitszustand auf möglicherweise freiwilligen Entscheidungen, wie z. B. dem Lebensstil, beruht). Allen Versicherungsnehmern stehen hierbei unabhängig von der Höhe der gezahlten Versicherungsbeiträge die gleichen Leistungen zu. Ein zentrales Prinzip der Solidargemeinschaft, das der GKV zugrunde liegt, ist der Risikoausgleich zwischen Gesunden und Kranken. Grundlage hierfür ist eine ausreichend große Anzahl an gesunden Beitragszahlern mit geringer Inanspruchnahme von Leistungen zu relativ gesehen wenigen Beitragszahlern mit Krankheiten und entsprechend höherer Leistungsanspruchnahme.

Da es sich im Kontext von Gesundheit um einen besonders sensiblen und schützenswerten Bereich handelt, ist die gesetzliche Krankenversicherung entsprechend stark reguliert. Die Einführung verhaltensbasierter Versicherungstarife in der GKV ist in Deutschland daher bislang keine Realität. Bei verhaltensbasierten Tarifen würden beispielsweise Personen systematisch benachteiligt, die nicht in der Lage sind, ein bestimmtes Verhalten auszuführen oder die nicht bereit sind, teils sensible körperbezogene Daten ihrer Versicherung zur Verfügung zu stellen. Würde beispielsweise eine bestimmte Anzahl von täglich zu absolvierenden Schritten für die Tarifierung herangezogen werden, hätten Personen mit Geheinschränkungen (z. B. bei einem Beinbruch) einen erheblichen Nachteil, denn sie hätten einen schlechteren Score und müssten unter Umständen mehr bezahlen. Insgesamt besteht im Zusammenhang mit verhaltensbasierten Tarifen also die Gefahr, dass das Prinzip der Solidargemeinschaft unterminiert wird. Im Bereich der gesetzlichen Krankenversicherungen erweisen sich die bereits existierenden Bonusprogramme, innerhalb welcher bestimmte Personen benachteiligt werden können, als diesbezüglich problematisch. Verschärft wird die Problematik überdies durch die Bonifizierung von Maßnahmen, für die das Nutzen-Schaden-Verhältnis für den Einzelnen insoweit umstritten sind (z. B. PSA-Test für Männer ab 50 Jahren, Ilic et al., 2018), dass sie nicht in den Versicherungskatalog aufgenommen werden durften.

3.3 Implikationen bezüglich der Gesellschaftsstruktur

Die potenziellen Folgen scheinen gesellschaftliche Relevanz zu besitzen. Eine zunehmende freiwillige Selbstüberwachung beispielsweise des Schlafverhaltens, des Bewegungsverhaltens, des Essverhaltens durch smarte Geräte und die gegenseitige Bewertung z. B. in sozialen Netzwerken können Druck erzeugen, sich gemäß den Idealvorstellungen einer Leistungsgesellschaft immer stärker selbst zu optimieren. Dieses könnte zu einer Aufspaltung in Optimierte und Suboptimierte führen (Selke, 2014). So ist es denkbar – wenngleich es gegenwärtig noch als fernliegend erscheint –, dass der soziale Status absehbar nicht mehr nur an der formalen Bildung, am Beruf und am Einkommen bemessen wird, sondern auch an der täglich gemessenen Anzahl von Schritten, einem allgemeinen Gesundheitscore oder dem durch das Fahrverhalten erlangten Score. Bereits heute wird in einigen Lebensbereichen der Status und Wert einer Person durch Scores beschrieben, erzeugt und festgelegt – kurz: „Zahlen machen Leute“, wie von Steffen Mau (2017) in seinem Buch „Das metrische Wir – Über die Quantifizierung des Sozialen“ treffend formuliert. Das Unternehmen Klout (Einstellung des Dienstes Ende Mai 2018) hatte beispielsweise auf Basis von Daten aus sozialen Netzwerken einen Score von Personen zwischen 1 und 100 erstellt, der den persönlichen Online-Einfluss widerspiegeln sollte.

Es ist bereits gängige Praxis, auf der Basis von Scoring eine Kategorisierung von Personen in spezifische Zielgruppen vorzunehmen, die es ermöglicht, Verbraucher unterschiedlich zu adressieren und zu behandeln. Eine solche Klassifizierung ist nicht nur deskriptiv und es geht dabei nicht nur um die bloße Beschreibung von Unterschieden. Vielmehr kann sie auch eine soziale Selektion nach „wertvollen“ und „weniger wertvollen“ Verbrauchern implizieren (Selke, 2014). Folglich besteht die Gefahr, dass es für Verbraucher aus den niedrigen Kategorien unter Umständen schwierig sein wird, beispielsweise einen Mobilfunkvertrag, eine Kfz-Versicherung oder eine Krankenversicherung zu halbwegs günstigen Konditionen abzuschließen. Hierbei stellt sich zudem die Frage, wie einfach oder schwierig es sein wird, von einer schlechten Kategorie in eine bessere wechseln zu können (Mau, 2017).

Durch die zunehmende Sammlung und Verknüpfung von Daten aus verschiedenen Lebensbereichen (z. B. Freizeitverhalten, Konsumverhalten, Zahlungsverhalten, Gesundheitsstatus, Bewegungsverhalten, Mitgliedschaften in sozialen Netzwerken, Wohngegend, Arbeitsleben, Einkommen oder Familienstand), zum Zwecke der Kategorisierung von Verbrauchern (Saetnan, Schneider & Green, 2018), erscheint es durchaus denkbar, dass Scoring die Grundlage für die Herausbildung und Etablierung einer digitalen Klassengesellschaft bilden könnte. In diesem Szenario würden die einzelnen Personen einem permanenten Wettbewerb um einen guten Score ausgesetzt sein. Die dadurch potenziell zunehmende Individualisierung der Gesellschaft bei gleichzeitiger Entsolidarisierung könnte zu einer Schwächung des gesellschaftlichen Zusammenhalts führen. Wenn, wie es sich in China bereits abzeichnen scheint, auch „Freunde“ in sozialen Netzwerken nach dem Score ausgewählt werden (Kostka, 2018), da niedrige Scores anderer auch dem eigenen Score schaden können, sind neue Formen sozialer Ausgrenzung denkbar. In diesem Zusammenhang ist der soziale Druck, sich scores zu lassen und sich „score-konform“ zu verhalten, nicht zu unterschätzen.

3.4 Implikation bezüglich der Wahlfreiheit

Im Hinblick auf Telematiktarife könnte es unter Umständen problematisch werden, dass bei hinreichender Verbreitung dieser Tarife immer größere Teile der Bevölkerung sich einem starken Anreiz bzw. Zwang unterworfen sehen könnten, selbst solche Tarife zu wählen. So ist es denkbar, dass beispielsweise bei Nicht-Teilnahme und Nicht-Preisgabe von Informationen bzw. persönlicher Daten auf ein potenziell höheres Risiko geschlossen werden könnte und es zu Stigmatisierung und (moneitären) Sanktionen kommt. Da Personen mit Voraussetzungen für einen guten Score tendenziell eher geneigt sein werden, Daten und Informationen zum Zwecke des Scorings über sich preiszugeben, ist es vorstellbar, dass sich unter Umständen auch Personen dazu gezwungen sehen, Daten und Informationen, die zur Teilnahme an einem Telematiktarif notwendig sind, preiszugeben, um etwaigen potenziellen Nachteilen, die im Kontext einer Nicht-Teilnahme auftreten könnten, zu entgehen (sog. *unravelling effect*; Peppet, 2011).

Der bereits in den 20er und 30er Jahren des letzten Jahrhunderts beschriebene Hawthorne-Effekt, der vereinfacht besagt, dass Personen ihr Verhalten aufgrund der Tatsache verändern bzw. anpassen, unter Beobachtung zu stehen (French, 1953), könnte auch beim Scoring zum Tragen kommen. Eine gesellschaftliche Herausforderung wird es daher sein, demokratische Werte wie persönliche (Wahl)freiheit, Privatsphäre, Autonomie und Solidarität als hohes Gut zu wahren. Die Wahrung dieser Werte impliziert beispielsweise auch, sich dafür zu entscheiden, vermeintlich unvernünftig zu handeln, sich nicht sozial erwünscht zu verhalten und sich aktiv dagegen zu entscheiden, gescort zu werden. Die vorgenannten gesellschaftlichen Implikationen eines zunehmenden Scorings in Deutschland sind zum gegenwärtigen Zeitpunkt in vielerlei Hinsicht noch als spekulativ zu betrachten. Sehr wahrscheinlich ist es jedoch, dass Scoring Auswirkungen auf die Gesellschaft haben wird – in welcher Form und welchem Ausmaß bleibt abzuwarten. Gerade deswegen ist es wichtig, ein Bewusstsein für mögliche gesellschaftliche Konsequenzen von Scoring zu schaffen. In diesem Zusammenhang sollte insbesondere eine Diskussion darüber geführt werden, welche ethischen und moralischen Werte für eine freie demokratische Gesellschaft unerlässlich sind und welche Implikationen im Zusammenhang mit Scoring gesellschaftlich als inakzeptabel und welche als akzeptabel gelten sollen.

VIII. Die Gefahr des Super-Scores

Verfahren, die Big Data und / oder selbstlernende Algorithmen benutzen und Verbraucher-Scoring auf Basis riesiger Datenmengen und komplexer algorithmischer Berechnungen betreiben, werden in den in diesem Gutachten betrachteten Branchen in Deutschland bislang noch nicht angewandt (siehe Kapitel C). Ein Trend zu solchen Geschäftsmodellen lässt sich außerhalb Deutschlands allerdings schon in verschiedenen Bereichen beobachten, wie in diesem Kapitel gezeigt werden soll. Diese Entwicklung ist auch vor dem Hintergrund von Möglichkeiten des Datenkaufs bei spezialisierten Datenhändlern zu sehen. Durch Datenhandel erzeugte, umfangreiche Datenmengen können potenziell die Möglichkeit eröffnen, anonymisierte oder pseudonymisierte Daten wieder personenbeziehbar zu machen (so genannte „Re-personalisierung“). Durch Verknüpfung personenbezogener Daten aus verschiedenen Lebensbereichen und Quellen ist somit auch die Bildung eines Super-Scores, ähnlich dem *Social Credit Score* in China, eine nicht auszuschließende Gefahr.

Die chinesische Regierung plant, ab 2020 das Verhalten aller Bürger und Unternehmen mit einem Sozialkreditsystem zu bewerten. „Aufrichtiges“ Verhalten im Sinne der Kommunistischen Partei soll belohnt, „unaufrichtiges“ Verhalten bestraft werden. Dies kann Auswirkungen auf Ausbildung, Karriere und den gesamten Alltag bedeuten⁴¹. Offizielles Ziel ist unter anderem, Vertrauen und Transparenz zwischen Marktteilnehmern zu erhöhen. Erste Auswirkungen sind auch auf ausländische Kooperationspartner und Kunden zu beobachten (Hoffmann, 2018). Derzeit finden in 40 chinesischen Städten Pilotprojekte statt. Zudem bieten einige kommerzielle Anbieter Sozialkreditsysteme auf freiwilliger Basis an. Die Verfahren basieren auf algorithmischen Big Data-Analysen. Eine Studie zur öffentlichen Wahrnehmung solcher Systeme zeigte, dass rund 80 % der chinesischen Internetnutzerinnen und -nutzer die staatlichen und kommerziellen Sozialkreditsysteme in ihrem Land positiv bewerten (Kostka, 2018).

Es stellt sich die Frage, ob ähnliche Scoring-Systeme auch in Deutschland denkbar sind (siehe dazu auch Al-Ani, 2018). Zwar sind solche Entwicklungen in der westlichen Welt nicht von staatlicher Seite zu erwarten, doch Unternehmen könnten mit entsprechenden Geschäftsmodellen das Ziel verfolgen, Daten aus verschiedenen Lebensbereichen zu verknüpfen und mittels Algorithmen Super-Scores zu berechnen.

In Folgenden werden exemplarisch einige Geschäftsmodelle Big Data-basierter Scoring-Verfahren außerhalb Deutschlands beschrieben und es wird auf das Thema Datenhandel eingegangen. Damit sollen aktuelle Möglichkeiten und Trends im Scoring-Bereich aufgezeigt werden, die auch für den deutschen Raum perspektivisch Relevanz entfalten könnten. Auf diese Weise kann auch die hohe Bedeutung eines in Deutschland vergleichsweise hohen Regulierungsniveaus verdeutlicht werden.

Das Szenario eines Super-Scores, der unter Umständen sogar aufgrund freiwilliger Einwilligungen von Verbrauchern, ihre Daten zur Weitergabe an Dritte zur Verfügung zu stellen, zustande kommen könnte, wird anschließend erläutert.

1. Scoring-Geschäftsmodelle im Ausland

Bonitäts-Scoring

Immer mehr Unternehmen im Finanzsektor entwickeln Scoring-Verfahren, um die Kreditwürdigkeit von Personen mittels Big Data zu prognostizieren (Jentsch, 2016). Genutzt werden auch unkonventionelle Daten, also Daten ohne direkten Bezug zur finanziellen Bonität, wie Aktivitäten in sozialen Netzwerken oder Surf- und Kaufverhalten im Internet. Zielgruppe sind meist

⁴¹ Bekannt ist beispielsweise, dass die Score-Höhe Auswirkungen darauf haben kann, ob Flugzeug- oder Bahn-Tickets erworben werden können oder ob die Kinder auf eine Privatschule dürfen. Bei einem hohen Score gibt es außerdem z. B. Vorzugsbehandlungen im Krankenhaus oder beim Carsharing muss keine Kaution hinterlegt werden (Kostka, 2018). 2014 verkündete der kommerzielle Score-Anbieter Alibaba eine Zusammenarbeit mit dem Luxemburgischen Konsulat: bei einem entsprechend hohen *Sesame-Score* können Chinesen vereinfachte Visaanträge über ein Online-Portal stellen (Alibaba Group, 2014).

die sogenannten „underbanked“, also Menschen, die z. B. mangels einer traditionellen Finanzhistorie oder aufgrund fehlender Bonität bei klassischen Kreditinstituten kein Darlehen bekommen würden. Märkte vor allem in Entwicklungs- oder Schwellenländern sollen so erschlossen werden.

Die Firma Kreditech⁴² wurde 2012 als Start-up gegründet und hat ihren Hauptsitz in Hamburg (Gründerszene.de, 2018). Über Tochterfirmen verarbeitet Kreditech derzeit in Russland, Polen, Spanien, Tschechien und Mexiko Kreditanträge⁴³. Anders als klassische Auskunfteien sucht Kreditech nicht primär nach Versäumnissen in der Vergangenheit, um die Kreditwürdigkeit zu ermitteln – die Firma benutzt offenbar Proxy-Variablen und wertet nach eigenen Angaben Big Data mittels künstlicher Intelligenz und „selbst lernender“ Algorithmen aus (die mit immer wieder zusätzlichen Daten gefüttert werden und deswegen etwas irreführend in der Öffentlichkeit als „selbst lernend“ bezeichnet werden) (Seibel, 2015).

Nach Angaben des Unternehmens identifiziere seine Software Kunden online innerhalb weniger Sekunden und errechne vollautomatisch ihren Kreditscore. Wird ein Antrag genehmigt, würde das Geld nach durchschnittlich 15 Minuten auf das Konto des Kunden überwiesen.

Bis zu 20.000 verschiedene Einzelinformationen verarbeitet der Kreditech-Algorithmus laut Webseite pro Anfrage, darunter neben klassischen Bonitäts-Informationen, die von der SCHUFA bezogen werden, auch ortsbezogene Daten (GPS), Social Graphs (Likes, Freunde, Locations, Posts), Online-Shoppingverhalten oder Gerätedaten (Schulz, Müller & Rosenbach, 2013).

Neben den Informationen, die beim Ausfüllen des Antrags angegeben werden, wird auch der Ausfüllprozess des Fragebogens selbst bewertet: Wird der Kreditantrag von einem (teuren) iPad aus gestellt oder von einem anderen Gerät? Wie lange dauert das Ausfüllen des Fragebogens? Ebenso werden die Fehlerhäufigkeit und der Einsatz der Löschtaste registriert. Kreditech-Gründer Sebastian Diemer im Interview mit der Tageszeitung *Die Welt*: „Wir haben festgestellt, dass Leute, die ihren Kredit nicht zurückzahlen, eine ganz bestimmte Schriftart auf ihrem Computer hatten.“⁴⁴ (Seibel, 2015). Zusätzlich werden Informationen ausgewertet, die über die Person im Internet zugänglich sind. Mit der Zustimmung des Antragstellers werden soziale Medien wie Facebook und Twitter ebenso analysiert wie Amazon oder Ebay. Wie viele Facebook-Freunde haben studiert? Wie kreditwürdig sind diese? Anhand des Facebook-Profiles wird auch kontrolliert, ob Foto und Ort mit anderen Informationen etwa bei Xing oder LinkedIn übereinstimmen.

Kreditech ist nur ein Beispiel für eine Reihe von Firmen, die ähnliche Geschäftsmodelle verwenden, um ihren Kunden Kredite zu gewähren oder zu vermitteln. Als weitere Beispiele genannt seien ZestFinance⁴⁵ und Kabbage⁴⁶ mit Sitz in den USA, Wonga⁴⁷ und Big Data Scoring⁴⁸ mit Sitz in Großbritannien oder Lenddo⁴⁹ mit Sitz in Singapur. 2015 wurde Facebook ein Patent auf ein Verfahren gewährt, dass unter anderem Geldverleihern ermöglichen soll, die Kreditwürdigkeit von Facebook-Mitgliedern anhand deren Freundeskreis zu beurteilen (Lunt, 2014).

42 Abgerufen am 18. Juni 2018 von URL <https://www.kreditech.com/>.

43 In Deutschland hat Kreditech seinen Service nach drei Wochen wieder eingestellt, nachdem die Finanzaufsicht BaFin eine Prüfung des Geschäftsmodells angekündigt hatte (Schulz, Müller & Rosenbach, 2013).

44 Es handelt sich hierbei um eine Schriftart, welche von Casino- und Poker-Programmen verwendet wird. Bei der genannten Kundengruppe handelt es sich also laut Kreditech mit erhöhter Wahrscheinlichkeit um Spieler im Internet (Seibel, 2015).

45 Abgerufen am 18. Juni 2018 von URL <https://www.zestfinance.com/>.

46 Abgerufen am 18. Juni 2018 von URL <https://www.kabbage.com/>.

47 Abgerufen am 18. Juni 2018 von URL <https://www.wonga.com/>.

48 Abgerufen am 18. Juni 2018 von URL <https://bigdatascoring.com/>.

49 Abgerufen am 18. Juni 2018 von URL <https://www.lenddo.com/>.

Gekennzeichnet sind die Geschäftsmodelle alle durch den Einsatz von komplexen Big Data-Analysen und algorithmischen Entscheidungsprozessen und zeigen, was technisch inzwischen möglich ist. Die verwendeten Daten gehen wie oben beschrieben weit über traditionelle Variablen wie Zahlungshistorie oder Einträge in Schuldnerverzeichnissen hinaus.

Scoring im Gesundheitsbereich

Auch im Gesundheitsbereich entwickeln sich zunehmend Geschäftsmodelle, die zwar nicht immer auf Big Data basieren, zumindest aber algorithmische Scoring-Verfahren nutzen, um den Gesundheitszustand oder das Gesundheitsverhalten von Personen zu quantifizieren bzw. zukünftige Gesundheit vorherzusagen (Kolany-Raiser, 2016; Wiegard & Breitner, 2017; Budzinski & Schneider, 2017). Vor allem für Krankenversicherungsunternehmen sind solche Produkte interessant, falls so das Verhalten von Versicherten gesteuert und kontrolliert sowie das Kostenrisiko beeinflusst und kalkuliert werden können. Grundlage für die Datenbasis bilden unter anderem Fitnessarmbänder und andere Wearables, die Gesundheitsparameter wie z. B. Blutdruck und Herzfrequenz messen sowie sportliche Aktivitäten und Bewegung aufzeichnen. Auch Schlaf und Ernährung können über Wearables und Apps protokolliert werden. Unsere Recherchen haben freilich ergeben, dass bei Scoring-Projekten in der Krankenversicherung zumindest im stark regulierten deutschen Markt bislang vor allem die Kundengewinnung und -bindung im Vordergrund steht (siehe C.III).

Ein privater Dienstleister im Gesundheitsbereich ist das Start-up Dacadoo⁵⁰, ein 2010 gegründetes, schweizerisches Unternehmen mit Hauptsitz in Zürich. Dacadoo bewertet den aktuellen Gesundheitszustand und das Wohlbefinden mithilfe des persönlichen Dacadoo Gesundheitsindex – eines Werts zwischen 1 (maximal schlecht) und 1000 (maximal gut). Der Index ändert sich nach oben bzw. unten in Echtzeit, abhängig von Körperwerten, emotionalen Wohlbefinden und Lebensstil (Bewegung und Sport, Ernährung, Stress und Schlaf). Aktivitäten werden mit dem Aktivitäten-Tracker

in der Dacadoo App aufgenommen. Zusätzlich werden verschiedene Fitness-Tracking-Geräte und -Apps unterstützt. Außerdem kann Dacadoo sich mit digitalen Waagen, Blutdruckmessgeräten und Herzfrequenzmessbändern verbinden. Schließlich können Daten auch manuell über die App oder über den Web-Zugang eingegeben werden. Dacadoo nutzt auch Angaben zu früheren Gesundheitsproblemen (darunter Einzelheiten zur Familiengeschichte) oder Krankenhausaufenthalten sowie zu Lebensgewohnheiten.

Soweit die Aussagen des Unternehmens Dacadoo. Bislang gibt es allerdings keine klinische Studie, die zeigt, dass die Nutzung der Dacadoo-Plattform auch tatsächlich in Bezug auf Verhaltensänderung und Gesundheit wirksam ist.

Ein anschauliches Beispiel für ein umfangreiches Gesundheits-Scoring im Rahmen einer Krankenversicherung ist etwa das Versicherungsunternehmen Aviva in Großbritannien. Wer bei Aviva eine private Krankenversicherung abschließt, kann kostenlos an dem Programm „MyHealthCounts“⁵¹ teilnehmen. Mit dem sogenannten „Q Score“ wird der aktuelle Gesundheitszustand beurteilt. Mit Hilfe von Trackern können Gewicht, Blutdruck, Fitness, Rauchverhalten, Cholesterin und mehr überwacht werden. Nach der Registrierung füllt der Versicherte außerdem einen Fragebogen aus. Der Fragebogen umfasst zum Beispiel Fragen zu Ernährungs- und Bewegungsverhalten sowie Krankengeschichte (eigene und Familienangehörige). Auf dieser Basis wird der personenbezogene Score berechnet. Die Gesundheit einer Person wird dabei im Vergleich zu 100 Personen desselben Alters, Geschlechtes und ethnischen Hintergrunds bewertet. Das Programm ermöglicht Teilnehmern die Chance, bis zu 15 % Rabatt auf Versicherungsbeiträge zu erhalten.

50 Abgerufen am 18. Juni 2018 von URL <https://www.dacadoo.com>.

51 Abgerufen am 18. Juni 2018 von URL <https://www.myhealthcounts.co.uk/>.

Avivas Partner für das MyHealthCounts-Programm ist die britische Firma Roadtohealth, welche auch die Gesundheitsscoring-App Quealth⁵² entwickelte. Die App errechnet das individuelle Risiko, in der Zukunft verschiedene Krankheiten zu entwickeln. Dazu werden vom Nutzer der App Fragebögen zu Krankheitsgeschichte, Biometrie und Lifestyle beantwortet. Algorithmen berechnen den Quealth Score – ein Wert zwischen 1 und 100. Wie auch bei MyHealthCounts und Dacadoo ist der genaue Algorithmus nicht bekannt. Das Erkrankungsrisiko wird derzeit für fünf nicht übertragbare Krankheiten geschätzt: Diabetes, Herz-Kreislauf-Erkrankung, Krebs, Demenz und chronisch obstruktive Lungenerkrankung. Das Unternehmen gibt auf seiner Webseite an, die Quealth Algorithmen seien evidenzbasiert und hätten eine nachgewiesene starke prädiktive Kraft. Eine unabhängige klinische Studie, die diese Behauptungen belegt hat, ist bislang jedoch nicht bekannt.

Neben der Einschätzung des Risikos werden nach Angaben des Unternehmens die erfassten Informationen genutzt, um Gesundheitsinformationen, Ratschläge und Coaching individuell angepasst zu vermitteln.

Ein weiteres Beispiel dafür, wie lernende Algorithmen, die immer wieder mit neuen Daten trainiert werden, im Gesundheitsbereich eingesetzt werden können, stammt aus den USA: SelfieQuote.com ist ein Test-Projekt des US-Lebensversicherungsunternehmens Legal & General⁵³. Kunden können über die Webseite ein Angebot für eine Lebensversicherung erhalten, indem sie ein Foto ihres Gesichtes hochladen. Die Technologie wertet das Foto anhand von hunderten Punkten auf dem Gesicht aus und berechnet damit Proxy-Informationen zu Body Mass Index, Geschlecht und biologischem Alter. Ein weiteres Feature der Technologie ist die Möglichkeit, abzuschätzen, ob die Person Raucher ist. Die so gewonnenen Daten werden nach Angaben von Legal & General zur Verifizierung der durch die Kunden gemachten Angaben genutzt und ein personalisiertes Versicherungspaket wird erstellt. Der genaue Vorgang ist außerhalb des Unternehmens nicht bekannt. Wahr-

scheinlich dient Selfiequote derzeit vor allem als Marketing-Instrument um junge, technikaffine Kunden zu gewinnen (Moorcraft, 2018).

Das Projekt von Legal & General entstand in Partnerschaft mit dem Technologieunternehmen Lapetus Solutions⁵⁴. Es gibt vor, mit einer Kombination von Gesichtsanalysen, biodemografischer Informationen und Befragungen Gesundheitsstatus und Lebenserwartung bzw. Sterberisiko von Personen berechnen zu können. Derzeit untersucht wird die Möglichkeit, mit der Technik erste Anzeichen von Krankheiten wie Diabetes, Herzfehler oder Demenz erkennen zu können.

Auch in den USA ansässig ist das Startup Aspire Health, welches von der Google-Holding Alphabet mitfinanziert wird⁵⁵. Hauptgeschäftsfeld sind Angebote der Palliativ-Medizin. Die Firma hat einen Algorithmus entwickelt, der die verbleibende Lebensdauer schwerkranker Patienten berechnen soll. Gesundheitsdaten über die Patienten werden gesammelt und deren Krankheitsbilder mit Mustern häufiger Behandlungen abgeglichen (Welchering, 2017). Auf diese Weise soll bewertet werden, welche weitere Behandlung sinnvoll erscheint. Der Algorithmus gibt eine Empfehlung zur weiteren Behandlung ab, der Arzt trifft darüber die Entscheidung. Dem Verfahren liegt die Annahme zugrunde, dass eine teure ärztliche Behandlung dann nicht sinnvoll sei, wenn die Überlebenschance niedrig ist. Die Profitabilität des Gesundheitswesens könne so gesteigert werden.

Eine kritische Betrachtung, insbesondere in Bezug auf die Prädiktionskraft und Wirksamkeit der einzelnen Scoring-Modelle, ist nicht nur bei den beschriebenen Beispielen zwingend. Denn in der Regel fehlt es an unabhängiger wissenschaftlicher Evidenz dafür, dass die Scores auch wirklich halten, was sie versprechen. Darüber hinaus ist auch eine Berücksichtigung der Profitabilität als Entscheidungskriterium im Gesundheitswesen kritisch zu sehen. Die Beispiele verdeutlichen jedoch, dass die Quantifizierung der Gesundheit mit Hilfe von Algorithmen vielseitig einsetzbar ist und auch bereits

52 Abgerufen am 18. Juni 2018 von URL <https://www.quealth.co/>.

53 Abgerufen am 18. Juni 2018 von URL <https://term.lgamerica.com/selfie-quote/#/>.

54 Abgerufen am 18. Juni 2018 von URL <https://www.lapetusolutions.com/>.

55 Abgerufen am 18. Juni 2018 von URL <http://aspirehealthcare.com/>.

eingesetzt wird. Es bleibt abzuwarten, wie sich der Trend entwickelt und inwiefern sich auch in Deutschland ähnliche Geschäftsmodelle etablieren (können).

2. Datenakkumulation und Datenhandel

Grundlage eines jeden Verbraucher-Scorings sind Verbraucherdaten, teilweise in erheblichen Mengen (Dixon & Gellmann, 2014). Es gibt kaum noch Lebensbereiche, in denen Verbraucher keine Datenspuren hinterlassen. Die Digitalisierung hat dazu geführt, dass das Sammeln von personenbezogenen Daten so einfach ist wie nie zuvor. Ebenso ist das wirtschaftliche Interesse gewachsen, diese Daten gewinnbringend zu nutzen, z. B. in den Bereichen Werbung, Onlinehandel, Marktforschung und Politik. Ziel ist meistens, Betrug zu vermeiden, Personen zu identifizieren oder auch Marketingkampagnen und Kundenkommunikation effizienter zu gestalten, indem konkrete Zielgruppen präzise adressiert werden (OECD, 2013).

Daten, die nicht selbst erhoben werden können, werden häufig von Datenhändlern dazugekauft. Das Geschäftsmodell solcher Unternehmen besteht im Sammeln, Aufbereiten und Verkaufen von Daten. In Deutschland lag 2014 das Marktvolumen im Bereich des Handels mit Adress- und anderen personenbezogenen Daten bei rund 610 Mio. Euro (Goldmedia, 2017).

Die persönlichen Daten, die über Verbraucher gesammelt werden, lassen sich unterschiedlichen Kategorien zuordnen (OECD, 2013): demografische Daten (z. B. Geburtsdatum, Geschlecht, Familienstand, Bildungsgrad, Einkommen), nutzer-generierte Inhalte (z. B. Blogs, Kommentare, Rezensionen, Fotos, Videos), Verhaltensdaten im Internet (z. B. Eingaben in Suchmaschinen, Onlineshopping), Daten des sozialen Umfelds (z. B. Kontakte und Freunde in sozialen Netzwerken), Ortungsda-

ten (z. B. Adresse, GPS-Daten, IP-Adresse) und offizielle, personenbezogene Identifikationsdaten (z. B. Ausweisnummern, Kontonummern, Polizeiakten).

Datenhändler erhalten Daten in der Regel nicht direkt vom Verbraucher, sondern durch Dritte, und häufig ohne dass Verbraucher sich darüber im Klaren sind. Datenhändler nutzen eine Vielzahl an Quellen, um ihre Datenbanken aufzubauen, z. B. aus dem Onlinehandel und von Versandhäusern, durch Kunden- oder Cashback-Karten, über Preisausschreiben, aus öffentlich zugänglichen Statistiken und Verzeichnissen oder von Datensammlern wie Facebook, Amazon, Twitter oder anderen Webseitenbetreibern (Goldmedia, 2017).

Die Datensammler wiederum können persönliche Daten im Internet auf verschiedenste Weise erlangen (Palmethofer, Semsrott & Alberts, 2016): Zum einen gibt es einen unmittelbaren Zugang, wenn Daten freiwillig übermittelt werden, z. B. wenn sich Nutzer bei einem Onlineservice anmelden oder ein Profil in einem sozialen Netzwerk anlegen. Weiterhin ergeben sich Daten, wenn eine Person beobachtet wird und so Informationen mittelbar preisgibt, z. B. durch Browser-Verläufe oder GPS-Positionen. Schließlich werden Daten wie Online-Profile, Likes oder Bewertungen aber auch nicht-personenbezogen erscheinende Daten analysiert und ausgewertet und so persönliche Daten, wie Alter oder Geschlecht, abgeleitet.

Den Weg seiner Daten zu verfolgen ist für den Verbraucher quasi unmöglich. Die meisten Datenhändler erhalten Daten auch von anderen Datenhändlern und verkaufen sie selbst wiederum an weitere Datenhändler. Eine Studie der Federal Trade Commission (2014) stellte fest, dass sieben von neun untersuchten Datenhändlern in den USA sich gegenseitig Daten zu Verfügung stellen.⁵⁶ Das bedeutet, dass Verbraucher wenige Möglichkeiten haben, der Nutzung der Daten im Nachhinein zu widersprechen, falsche Daten zu korrigieren und Einordnungen in bestimmte Kategorien anzufechten, da sie nicht wissen, wer überhaupt ihre Daten hat und welche Daten bereits im weltweiten Datenhandel verfügbar sind.

⁵⁶ Untersucht wurden die neun Unternehmen Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, Peek You, Rapleaf, Recorded Future (Federal Trade Commission, 2014).

Ein zentrales Thema ist in diesem Zusammenhang die informierte Einwilligung, welche die rechtliche Grundlage für das Sammeln vieler der genannten Daten bildet. Der Verbraucher erlaubt durch eine freiwillige Einwilligung zum Beispiel eine Aufzeichnung des Nutzungsverhaltens, welche ansonsten verboten wäre. Die Einwilligung erfolgt normalerweise durch die Zustimmung von Verbrauchern zu Datenschutzerklärungen oder Allgemeinen Geschäftsbedingungen (AGBs). Allerdings sind mit der informierten Einwilligung einige Probleme verbunden (Hofmann & Bergemann, 2017)⁵⁷: Zum einen zeigen Studien, dass die Datenschutzbestimmungen oder AGBs der Anbieter von Verbrauchern mehrheitlich nicht gelesen werden. Das Setzen eines entsprechenden Häkchens vor dem Herunterladen einer App oder dem Beitreten zu einem sozialen Netzwerk ist zu einer habituellen Alltagshandlung geworden. Ein Faktor ist dabei, dass der Zugang zu einer entsprechenden Dienstleistung ohne eine Zustimmung meist überhaupt nicht möglich ist. Der Nutzer sieht also keine ernsthafte Alternative, ohne den Alltag erheblich einzuschränken oder auf gesellschaftliche Teilhabe zu verzichten. Eine weitere Kritik an der informierten Einwilligung ist, dass die umfassende Informiertheit des Verbrauchers zum Zeitpunkt der Einwilligung angezweifelt werden kann. Datenschutzerklärungen und AGBs sind oft seitenlang und für Laien unverständlich formuliert. Außerdem können Nutzer die Folgen, die die Einwilligung in die Speicherung und Verarbeitung ihrer Daten haben kann, meist überhaupt nicht abschätzen. Der SVRV setzte sich deshalb bereits in seinem Gutachten „Digitale Souveränität“ (2017a) dafür ein, dass Unternehmen den Verbraucher mit einem verständlichen One-Pager (500 Wörter) über AGB und Datenschutzerklärungen zu informieren haben. Zur Stärkung der Informiertheit wird von einigen Verbraucherschützern auch die Einführung vorformulierter Datenschutzbestimmungen vorgeschlagen (Pollmann & Kipker, 2016).

Auch wenn einzelne Quellen möglicherweise nur wenige Datenpunkte über einen Verbraucher zur Verfügung stellen, können Datenhändler viele verschiedene Datenelemente zusammenführen. Auf diese Weise kann es gelingen, ein detailliertes Gesamtbild über individuelle

Lebensführung und Verhalten zu erlangen. Datenhändler aggregieren außerdem nicht nur Rohdaten, wie z. B. den Namen, Alter, Zahlungsvorgänge oder die Suchhistorie einer Person, sondern auch daraus abgeleitete Daten, also Schlussfolgerungen wie z. B. Produktvorlieben, Kaufkraft oder Fitnesslevel, aber auch höchst sensitive Einordnungen in Gruppen nach ethnischen Hintergrund, Einkommenslevel oder Gesundheitsstatus (zu den hiermit verbundenen Gefahren mittelbarer Diskriminierung siehe B.II.2). Eine Studie in den USA hat beispielsweise belegt, dass eine Analyse von Facebook-Likes mit gewisser Zuverlässigkeit ermöglicht, auf Eigenschaften wie sexuelle Orientierung, ethnische Zugehörigkeit, Religion, politische Einstellung, Intelligenz, Suchtmittelkonsum, Alter, Beziehungsstatus, Geschlecht und weiteres zu schließen (Kosinski, Stillwell & Graepel, 2013). Auf verschiedene Persönlichkeitsmerkmale lässt sich z. B. auch mittels Auswertung von Mobiltelefonaten (Chittaranjan, Blom & Gatica-Perez, 2011; De Montjoye, Quoidbach, Robic & Pentland, 2013) oder Suchmaschinenutzung und Browserverhalten schließen (Kosinski, Stillwell, Kohli, Bachrach & Graepel, 2012).

Die zur Verfügung stehenden Informationen werden weiterhin dazu genutzt, Verbraucher in Gruppen und Kategorien einzuordnen. Die Gruppenzugehörigkeit beruht meist nicht aufgrund eines einzigen Merkmals, sondern aufgrund einer Kombination verschiedener Informationen. Beim Datenhändler Acxiom (s.u.) gibt es beispielsweise Kategorien wie „bürgerliche Mitte“, „urban Aktive“ oder „statusorientierte Arbeiter“ (Junge, 2012).

Wie zu Anfang dieses Kapitels anhand einiger Beispiele gezeigt wurde, nutzen neuartige Scoring-Geschäftsmodelle zunehmend Big Data und algorithmenbasierte Verfahren. Da zur Bildung eines Scores keine kausale Wirkung zwischen zwei verschiedenen Merkmalen, sondern nur eine statistische Korrelation bestehen muss, werden statistische Methoden eingesetzt, um große Mengen an Verbraucherdaten zu analysieren und darin Muster und Zusammenhänge zu erkennen, mit dem Ziel Prognosen oder Einschätzungen über deren zukünftiges Verhalten zu treffen (Christl, 2014). Das Interesse, möglichst große Datenmengen zu akkumulieren, ist

⁵⁷ Die genannten Herausforderungen machen deutlich, dass die informierte Einwilligung des Verbrauchers als Instrument des Datenschutzes nicht allein eine zentrale Stellung einnehmen kann, sondern durch weitere Mechanismen zu ergänzen ist. Zentral sind insoweit die Gewährleistungen des Zweckbindungsgrundsatzes (Art. 5 Abs. 1 Buchst. b DSGVO) und des Kopplungsverbots (Art. 7 Abs. 4 DSGVO, siehe hierzu bereits SVRV, 2016).

daher groß. Jeanette Hofmann sieht darin eine Gefahr zur datenbasierten Monopolbildung: „Es ist tatsächlich so, dass die Generierung und Analyse von Daten in der Hand von sehr wenigen Organisationen liegt. Und natürlich ist es auch so, dass je mehr Daten man hat, desto bessere Algorithmen kann man auch bauen und damit ist sozusagen in diesem Markt ein inhärenter Trend zum Monopol angelegt.“ (Bilger, Löwel & Tomaszewski, 2018; siehe dazu z. B. auch Rubinfield & Gal, 2017).

Im Folgenden werden einige Akteure im Datenhandel weltweit und insbesondere in Deutschland exemplarisch dargestellt, um ihre Geschäftspraxis zu veranschaulichen.

Beispiele Datenhändler – international und in Deutschland

Die Acxiom Corporation⁵⁸ ist ein Unternehmen aus den USA, welches Unternehmenskunden unter anderem mit Verbraucherdaten beim Marketing unterstützt. In den USA verfüge Acxiom über Daten zu 250 Millionen Bürgern. Weltweit agiert Acxiom laut Webseite in über 60 Ländern und verspricht seinen Kunden, 2,5 Milliarden Verbraucher adressieren zu können⁵⁹. Jeder Person, über die Daten vorliegen, wird eine 13-stellige Nummer zugeordnet, an die alle gespeicherten Informationen geknüpft sind (McLaughlin, 2013): demografische Daten, Haushaltscharakteristika, finanzielle Situation, Lebensereignisse, Interessen, Kaufverhalten, Sozialverhalten – die Liste der Merkmale ist lang⁶⁰. Das Gruppieren von Verbrauchern in bestimmte Zielgruppen gehört genauso zu den Services von Acxiom wie das Paaren von Daten mit konkreten Emailadressen, Anschriften oder Telefonnummern. Besondere Spezialität des Unternehmens sind Offlinedaten, z. B. von Regierungsbehörden, mit denen die Online-Daten angereichert werden. Außerdem bietet Acxiom seinen Kunden an, komplette Kundenkarteien zu lagern und zu verwalten. Zu den

Kunden gehören z. B. Kreditkartenanbieter, Automobilhersteller, Versicherungen, Einzelhandelsketten u.v.m.

Acxiom ist seit 2004 auch in Deutschland tätig und bietet Kunden unter anderem Adressdaten mit zahlreichen Zusatzinformationen zur jeweiligen Adresse und den dahinter stehenden Verbrauchern. Die Datenbestände können auf konkrete Konsumabsichten sowie Verhaltensweisen und Einstellungen der Verbraucher zugeschnitten werden. Daten zu 44 Millionen Deutschen habe Acxiom Deutschland bereits gesammelt.⁶¹

Neben Acxiom Deutschland sind die ABIS GmbH (gehört zur Deutsche-Post-Adress-Gruppe), die AZ Direct GmbH (Teil der Bertelsmann Printing Group), die Deutsche Post Direkt GmbH, die EOS Holding GmbH (eine Tochter der Otto Group) und die Schober Information Group Deutschland GmbH (Christl, 2014; Goldmedia, 2017) die dominierenden Unternehmen auf dem deutschen Markt des Datenhandels. Sie alle sind vor allem im Bereich Adresshandel tätig, d. h. sie sammeln Postanschriften von Verbrauchern, verifizieren und bereiten diese auf und reichern sie mit weiteren Zusatzinformationen an. An die Adressen gelangen sie durch öffentlich zugängliche Verzeichnisse wie Telefonbücher, Vereins- und Firmenregister. Außerdem kaufen die Datenhändler Daten von anderen Unternehmen, z. B. von Versandhändlern und Verlagen.

Die deutsche AZ Direct GmbH verwaltet beispielsweise 37 Millionen Haushaltsanschriften und bietet diese ihren Kunden für Direktmarketing-Aktivitäten an⁶². Durch ihr *Audience-Targeting-System AZ DIAS* stünden zu 40 Millionen Haushalten, 70 Millionen Personen und 20 Millionen Gebäuden Profilinginformationen zu Soziodemografie, Psychografie, Konsumeigenschaften, Lebensphasen, Lage/Geodaten uvm. zur Verfügung. Namhafte Unternehmen wie die Weltbildgruppe, Gruner + Jahr oder der Klingel-Konzern überlassen ihre Adressbestände der AZ Direct GmbH. Auf der Webseite⁶³ lassen sich Beschrei-

58 Abgerufen am 18. Juni 2018 von URL <http://www.acxiom.de/ueber-acxiom/>.

59 Abgerufen am 18. Juni 2018 von URL <https://www.acxiom.com/>.

60 Eine Auswahl ist zu finden unter: <https://www.acxiom.com/what-we-do/infobase/>, abgerufen am 18. Juni 2018.

61 Abgerufen am 18. Juni 2018 von URL <https://www.linkedin.com/company/acxiom-deutschland-gmbh>.

62 Abgerufen am 18. Juni 2018 von URL <http://www.az-direct.com/site/multichannel-marketing-produkte/direct-mail/>.

63 Abgerufen am 18. Juni 2018 von URL <http://www.az-direct.com/site/fileadmin/ikat/listinfos/index.html>.

bungen von über 2.500 Listen einsehen, die Kunden von Kooperationspartnern in Segmente wie „Ökologiebewusste Personen“, „Spendenaffine Akademiker im Ruhestand“ oder „Mehrfachpostkäufer mit sozialer Gesinnung“ einteilen.

3. De-Anonymisierung von Daten

Anonymisierte Daten, also solche, die nicht auf den jeweiligen Verbraucher rückführbar sind, fallen nicht unter das Datenschutzgesetz und können daher in Deutschland einschränkungslos erhoben und verwendet sowie von Datenhändlern ge- und verkauft werden. Der konkrete Personenbezug (typischerweise Name, Geburtsdatum, Anschrift) wird also aus den Datensätzen entfernt. Möglich ist auch eine Pseudonymisierung, das heißt Attribute wie Name oder Anschrift werden durch Pseudonyme oder Kennzeichen ersetzt⁶⁴.

Nicht zu unterschätzen ist aber die Gefahr der „De-Anonymisierung“ oder „Re-Personalisierung“ solcher anonymisierten oder pseudonymisierten Daten. Schon mit dem Vorliegen weniger verknüpfter Daten sind Personen normalerweise wieder identifizierbar. So zeigte Sweeney (2000), dass Postleitzahl, Geschlecht und Geburtsdatum ausreichen, um 87 % der US-amerikanischen Bevölkerung eindeutig zu identifizieren. Zur Identifizierung von immerhin 18 % der Bürger reichen Landkreis, Geschlecht und Geburtsdatum.

Die Möglichkeiten der De-Anonymisierung von Datensätzen sind vielfach belegt: Wissenschaftlern an der Universität Texas gelang es beispielsweise mittels eigens entwickelter Algorithmen einen anonymisierten Netflix-Datensatz mit den Filmbewertungen von 500.000 Nutzern teilweise zu de-anonymisieren. Indem die Daten mit nicht-anonymisierten Filmbewertungen auf der Internetseite „Internet Movie Database“ abgegli-

chen wurden, konnte die Identität einzelner Nutzer herausgefunden werden. Auch war es möglich, deren politische Präferenzen und andere sensible Informationen aufzuzeigen (Narayanan & Shmatikov, 2008). 2009 zeigten dieselben Autoren, wie Nutzer, die sowohl bei Twitter als auch bei der Fotoplattform Flickr aktiv waren, mit einer Fehlerquote von nur 12 % in einem anonymisierten Twitter-Datensatz re-identifiziert werden konnten (Narayanan & Shmatikov, 2009). Weitere Studien konnten zeigen, dass sich anonymisierte Mobilitätsdaten, die mittels GPS-Sensoren in Smartphones, Computern und Fahrzeugen gesammelt wurden, de-anonymisieren lassen, indem die Datensätze mit zusätzlichen Informationen, z. B. aus sozialen Netzwerken angereichert werden (z. B. De Montjoye, Hidalgo, Verleysen & Blondel, 2013; Ji, Li, Srivatsa, He & Beyah, 2016; Srivatsa & Hicks, 2012). In Deutschland gelang es 2016 Reportern vom Norddeutschen Rundfunk einen von einem Datenhändler erlangten Datensatz zu de-anonymisieren, der rund zehn Milliarden Internetadressen, aufgerufen von rund drei Millionen deutschen Internetnutzern, umfasste (Norddeutscher Rundfunk, 2016). Detaillierte Aussagen zum Surfverhalten konnten konkreten Personen zugeordnet werden. Der Datensatz war allerdings anscheinend auf rechtlich nicht zulässigem Weg entstanden: Eine Browsererweiterung des Unternehmens „Web of Trust“ hatte laut Aussage des NDR ohne Einwilligung der Nutzer deren besuchte Webseiten protokolliert und die Daten auf Servern im Ausland gespeichert.

Man kann also argumentieren, dass in Zeiten von Big Data jeder Datenpunkt aufgrund unzähliger Vernetzungsmöglichkeiten mit anderen Persönlichkeitsdaten potenziell personenbezogen ist (Boehme-Neßler, 2016). Das Zusammenführen und Abgleichen von mehreren Datensätzen mit personenspezifischen Merkmalen wie Kaufverhalten, Browserverläufen, Suchhistorien etc. aus verschiedenen Quellen ermöglicht den Rückschluss auf konkrete Verbraucher. Big Data und leistungsfähige Algorithmen lassen dies ohne großen Aufwand zu. Weiterhin ergibt sich daraus die Möglichkeit durch Datenhändler erworbene Daten auch für Verbraucher-Scoring zu verwenden, da diese in der Regel personenbezogen gebildet werden.

⁶⁴ Die Frage, inwieweit pseudonyme Daten nicht personenbezogen sind, ist umstritten. Im 26. Erwägungsgrund der DSGVO heißt es: „Einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, sollten als Informationen über eine identifizierbare natürliche Person betrachtet werden.“

Zusammenfassend lässt sich feststellen, dass Datenakkumulation und -handel in Zeiten von Big Data eine wichtige Rolle spielen und dass mit Algorithmen ein Identifizieren konkreter Verbraucher auch in großen, anonymisierten Datensätzen möglich ist. Doch was bedeutet dies im Hinblick auf das Thema Verbraucher-Scoring?

4. Verknüpfung von Daten zum Super-Score

Mit der steigenden Bedeutung des Datenhandels wächst das Potenzial, Daten aus verschiedensten Lebensbereichen in einer Datenbank und in einem Unternehmen zu vereinen. Potenziell können also viele Daten aus verschiedenen und vermeintlich unabhängigen Lebensbereichen konkreten Verbrauchern zugeordnet werden und dann auch als Grundlage für Scoring verwendet werden. Ein solcher Super-Score würde bedeuten, dass einzelne Verhaltensweisen von Verbrauchern lebensbereichsübergreifend weitreichende Konsequenzen nach sich ziehen können (siehe B.II.4).

Hinweise auf eine „Zweckentfremdung“ von Daten und Scores finden sich zum Beispiel in den USA. Bonitätsauskünfte (z. B. der FICO Score) werden von Autoversicherungsanbietern und Hauseigentümersicherungen zur Beitragskalkulation genutzt (Consumer Reports, 2015; Dixon & Gellmann, 2014). Von Interesse ist hierbei nicht, ob der Kunde mit einer bestimmten Wahrscheinlichkeit einen Kredit zurückzahlen wird, sondern die Informationen fließen in die Berechnung der Wahrscheinlichkeit ein, dass der Kunde bereit ist, höhere Versicherungsbeiträge zu zahlen, oder dass er einen Schaden melden wird (O’Neil, 2016).

Eine Umfrage der Society of Human Resource Management zeigte, dass fast die Hälfte (47%) der 430 befragten Arbeitgeber einen Bonitätscheck durchführt, bevor sie einen Bewerber einstellt (Society for Human Resource Management, 2012). Hierbei geht es ebenfalls nicht darum, konkret die Kreditwürdigkeit einer Person zu erfahren, sondern darum, aus Bonitätsinformationen Eigenschaften wie Vertrauenswürdigkeit oder Zuverlässigkeit abzuleiten (O’Neil, 2016).

Die Zweckentfremdung von Daten wird durch Datenhandel befördert. Insbesondere Bonitäts-Scoring ist bereits eng mit dem Geschäftsbereich Datenhandel – oft in einem undurchsichtigen Gewirr von Unternehmensgruppen und Tochterunternehmen – verknüpft. Teile der Bertelsmann-Gruppe sind beispielsweise sowohl der Datenhändler AZ Direct GmbH sowie die Auskunft Arvato Infoscore. Creditreform Boniversum betreibt außer seinem Geschäft als Auskunft mit dem Tochterunternehmen Microm⁶⁵ auch eine Konsumentendatenbank mit soziodemografischen, sozioökonomischen und psychographischen Merkmalen. Das Potenzial unkonventionellere Merkmale, z. B. das Verhalten im Internet, für das Bonitäts-Scoring sowie anderes Verbraucher-Scoring zu nutzen, ist also zweifellos gegeben.

Zu erwähnen ist an dieser Stelle außerdem, dass ein solches Potenzial auch unabhängig von Datenhändlern vorhanden ist – beispielweise bei großen Versicherungsunternehmen, die verschiedene Versicherungssparten betreiben. Datenbestände könnten hier mit den Einverständnis der Versicherten zusammengelegt und im Zusammenhang ausgewertet werden, auch verhaltensbasierte Daten. Die Generali Versicherung AG bietet beispielsweise aktuell Smart Insurance- und Telematik-Optionen in vier Sparten an: in der Risikolebensversicherung sowie in der Berufsunfähigkeitsversicherung („Generali Vitality“), in der Kfz-Versicherung („Generali Mobility“) und in der Hausratversicherung („Generali Domocity“). In der privaten Krankenversicherung soll das Vitality Programm in naher Zukunft eingeführt werden. Auch wenn betont werden muss, dass Generali keinerlei Pläne zu einer Zusammenlegung der Datenbestände geäußert hat, ergeben sich im Hinblick auf die Gefahr von Super-Scores durchaus interessante Frage-

65 Abgerufen am 18. Juni 2018 von URL <https://www.microm.de/>.

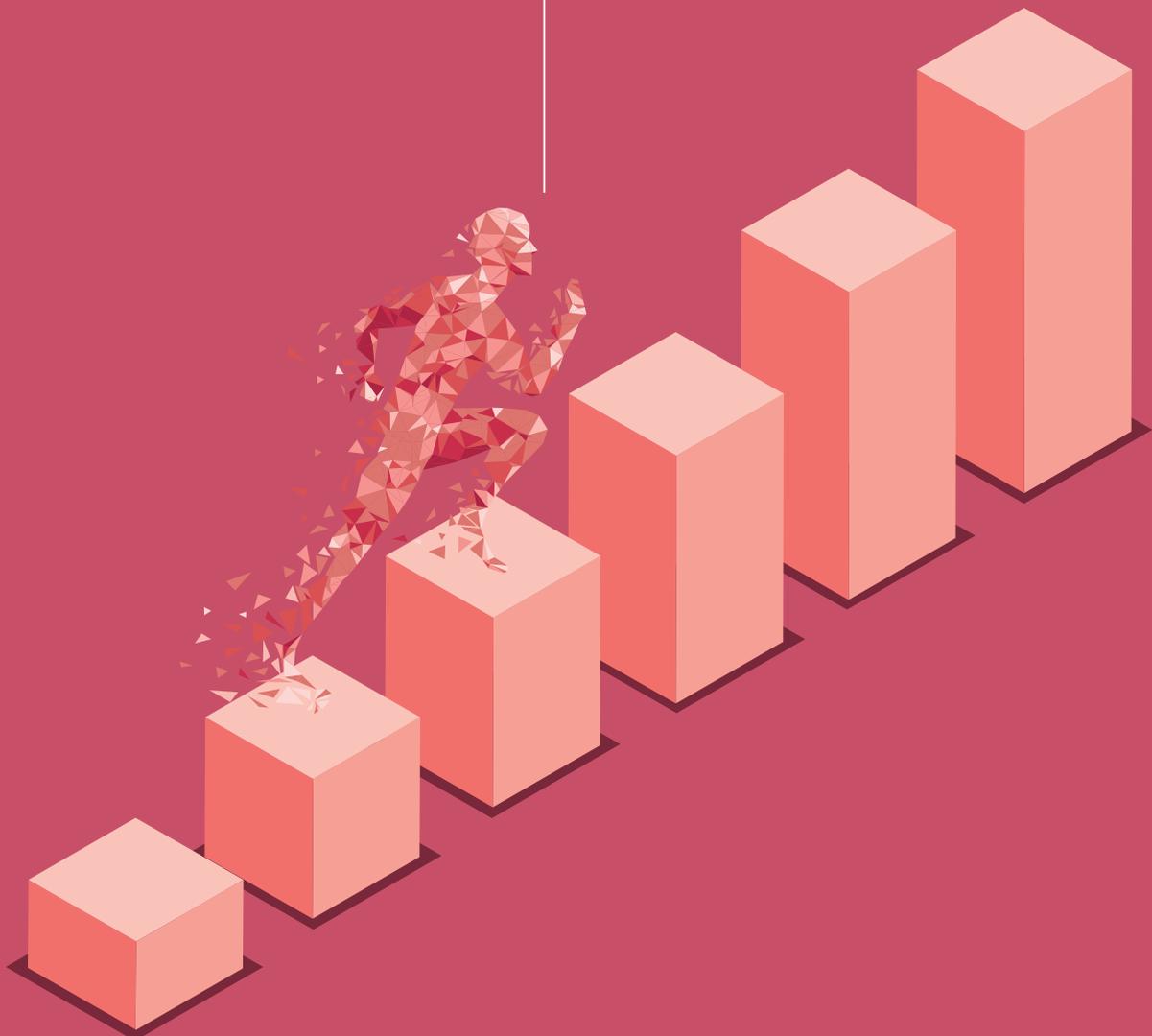
stellungen (vgl. auch Marktstudie in Kapitel C): Könnten nicht Fahrdaten relevant für die Beitragsberechnung in der Krankenversicherung sein? Schließlich können Unfälle mit Personenschaden die Gesundheitskosten steigern. Und umgekehrt? Erhöht nicht ein hohes Herzinfarktisiko die Unfallwahrscheinlichkeit?

Das Szenario eines Super-Scores, in dem Verhalten in verschiedenen Lebensbereichen verfolgt, analysiert und für die Score-Berechnung verwendet wird, erscheint im beschriebenen Kontext von innovativen Geschäftsmodellen, Big Data, Datenhandel und De-Anonymisierung durchaus plausibel. Während Entwicklungen wie das Sozialkreditsystem in China von staatlicher Seite in Deutschland nicht zu erwarten sind, stellt sich die Frage, ob Verbraucher besser vor ähnlichen Entwicklungen von Unternehmensseite her geschützt werden müssen. Wenn große Datenmengen über Verbraucherverhalten leicht zu erhalten und Analysewerkzeuge profitabel

einsetzbar sind, wird sich Datenhandel und Verbraucher-Scoring weiter verbreiten. Dass Unternehmen daran großes Interesse haben, zeigt z. B. Douglas Merrill, der Gründer der Auskunftei Zest Finance, dessen Gegenwartsanalyse auch als Warnung verstanden werden kann: „We feel like all data is credit data, we just don't know how to use it yet“ (Hardy, 2012).



Marktbefragung:
Auskunfteien, Kfz-Telematik
und Krankenversicherungen



I. Einleitung und zentrale Fragestellungen

Um das derzeitige Angebot von Verbraucher-Scoring in den im Zentrum dieses Gutachtens stehenden drei Bereichen (Bonitäts-Scoring, Telematik-Tarife, Gesundheits-Scoring) auf dem deutschen Markt zu analysieren, führte der SVRV im Frühjahr 2018 eine Marktstudie durch und befragte Unternehmen und Versicherer in den verschiedenen Marktsegmenten.

Die Methode der standardisierten, schriftlichen Befragung wurde gewählt, um Aussagen direkt von den Unternehmen und Versicherern zu erhalten und so möglicherweise mehr zu erfahren, als eine Recherche auf Unternehmenswebseiten o.ä. ergeben würde. Ein Anliegen war außerdem, neben der Perspektive der Verbraucher (siehe Kapitel D) auch die Perspektive und Interessen der Scoring-Nutzer bzw. -Akteure darzustellen.

Ziel war zum einen, die derzeitige Verbreitung von Scoring-Verfahren und verhaltensbasierten Geschäftsmodellen in den untersuchten drei Bereichen festzustellen. Die zentralen Fragen dabei waren:

- Welche Angebote gibt es und welche sind im Entstehen?
- Warum wird gescort, welche Ziele werden verfolgt?
- Welche Daten werden verwendet?
- Welche Gütekriterien erfüllen die Scoring-Verfahren?
- Wie transparent gehen die Unternehmen mit den Scoring-Verfahren um?

Zum Zweiten sollte den Befragten die Gelegenheit gegeben werden, ihre Erfahrungen, Meinungen und Pläne zum Thema Scoring bzw. verhaltensbasierte Tarife mitzuteilen. So interessierte beispielsweise, welche positiven und negativen Aspekte sie in Scoring sehen und wie sie die Zukunft solcher Verfahren in ihrer Branche einschätzen. Um möglichst offene Antworten zu erhalten, wurde den Unternehmen und Versicherern Anonymität bei der Auswertung zugesichert.

In der Analyse der Befragungsergebnisse sollen – auch unter der Berücksichtigung der identifizierten Problemfelder in Teil B – Verbraucherprobleme in Verbindung mit Scoring-Geschäftsmodellen herausgearbeitet werden, um eine empirische Grundlage für die Handlungsempfehlungen dieses Gutachtens zu schaffen.

II. Erhebungsdesign

Die Marktuntersuchung, die von der Geschäftsstelle des SVRV durchgeführt wurde, umfasste erstens Wirtschaftsauskunfteien, zweitens Kfz-Versicherungen und drittens sowohl gesetzliche als auch private Krankenversicherungen. Zu den drei Bereichen wurde eine Marktrecherche durchgeführt, um für die Befragung relevante Unternehmen zu identifizieren, und eine Marktübersicht erstellt, die als Grundlage der Erhebung diente.

Während sich in den einzelnen Marktbereichen sicherlich spezifische Verbraucherprobleme ergeben würden, sollte auch eine bereichsübergreifende Diskussion möglich sein. Deshalb wurde bei der Durchführung der Befragung darauf geachtet, den Grad der Standardisierung so zu halten, dass ein Großteil der Fragen allen Befragten in den drei Bereichen in einheitlicher oder ähnlicher Formulierung gestellt wurde (siehe Anhang II).

Gleichzeitig musste aber auch auf bereichsspezifische Besonderheiten Rücksicht genommen werden. Zum Beispiel wurde auf mögliche Zukunftstrends und Einflussfaktoren eingegangen, die für die drei Marktbereiche jeweils spezifisch sind. So wurde den Auskunfteien eine Frage zum möglichen Einsatz von Social Media-Daten gestellt, die Kfz-Versicherer um eine Einschätzung zum Thema europäische eCall-Verordnung gebeten und die Krankenversicherer auf das Thema elektronische Patientenakte angesprochen.

Das Berücksichtigen spezifischer Rahmenbedingungen bei der Befragung betraf vor allem den Bereich der Krankenversicherung. Sowohl in der gesetzlichen als auch in der privaten Krankenversicherung werden in Deutschland derzeit keine verhaltensbasierten Tarife angeboten. Jedoch können Versicherte im Rahmen der Bonusprogramme bei gesundheitsförderlichen Aktivitäten und Durchführung von Präventivmaßnahmen, die mit Hilfe von Bonuspunkten quasi „gescort“ werden⁶⁶, Geld- oder Sachleistungen erhalten. Bei den Auskunfteien und den Telematik-Tarifen in der Kfz-Versicherung ist der Grad der Automatisierung des Scorings jedoch wesentlich fortgeschrittener. Bei der Befragung der Krankenversicherer stand daher vielmehr im Fokus, herauszuarbeiten, wo Ansätze von Gesundheits-Scoring

bereits erkennbar sind und welche Wünsche, Pläne und Ansichten zu dem Thema von den Befragten geäußert werden. Dabei wurden die Krankenversicherer im Gegensatz zu den Auskunfteien und Kfz-Versicherern nicht konkret mit dem Begriff des „Scorings“ konfrontiert, da das Konzept der „verhaltensbasierten Tarife“ hier üblicher und damit zielführender schien. Auf diesem Grund entfiel bei den Krankenversicherern auch ein Großteil der Fragen zu Score-Berechnung, Modellbildung, statistischen Gütemaßen etc.

Insgesamt ist also eine vergleichende Analyse der Befragungsergebnisse möglich, während zugleich auf spezifische Rahmenbedingungen und mögliche Branchen-Trends eingegangen werden kann. In der Auswertung können alle drei Marktbereiche betreffende Verbraucherprobleme in Verbindung mit Scoring hervorgehoben sowie mögliche Best Practices herauskristallisiert werden.

Für die Befragung gab es keinen Pre-Test, allerdings wurden die Fragebögen auf Basis von Hintergrundgesprächen mit Unternehmen, Versicherern und Scoring-Experten entwickelt.

Einschränkend ist zu erwähnen, dass es sich bei den durch diese Marktstudie erlangten Ergebnissen ausschließlich um Aussagen der Unternehmen und Versicherer handelt, die nicht immer validiert werden können (beispielsweise durch öffentlich verfügbare Quellen). Ziel der Untersuchung war jedoch vorrangig zu erfahren, welche Informationen und in welchem Umfang die Unternehmen und Versicherer auf Nachfrage zu ihren Scoring-Verfahren zur Verfügung stellen und darüber hinaus, welche Einschätzungen, Pläne und Wünsche im Kontext Scoring sie bei zugesicherter Anonymität äußern.

Aus diesem Grund wurde auch der Großteil der Fragen offen formuliert, d. h. nur in wenigen Fällen konnten die Befragten vorgegebene Antwortmöglichkeiten ankreuzen. Für die Auswertung der offenen Fragen wurde eine quantitative Inhaltsanalyse durchgeführt und ein Kategorisierungssystem mit zusammenfassenden Oberbe-

⁶⁶ Im Rahmen einiger Bonusprogramme gibt es sogar die Möglichkeit, Daten aus Fitness-Apps zu nutzen, um Punkte zu sammeln, sowie den Gesamtpunktestand digital einzusehen und zu verwalten.

griffen und zugehörigen Kodes entwickelt. Sämtliche Antworten wurden von drei Ratern unabhängig voneinander kodiert.⁶⁷

In den Bereichen Bonitäts-Scoring und Telematik-Tarife in der Kfz-Versicherung wurden jeweils Rücklaufquoten von über 50 % (Auskunfteien: 60 %, Kfz-Versicherer: 63 %) und bei den Krankenversicherern eine Rücklaufquote von 41 % erreicht. Die gewonnenen Ergebnisse können nicht jeweils die gesamte Branche repräsentieren und die dargestellten Aussagen sind nur eingeschränkt verallgemeinerbar.

1. Anbieterübersicht

Auskunfteien

Um den deutschen Markt für Bonitäts-Scores zu analysieren, wurden verschiedene Wirtschaftsauskunfteien befragt. Dabei handelt es sich um privatwirtschaftlich geführte Unternehmen, die wirtschaftsrelevante Daten und Kreditwürdigkeitseinstufungen über Personen an Geschäftspartner übermitteln. Es wurden alle fünf Auskunfteien, die Mitglied im Verband „Die Wirtschaftsauskunfteien e.V.“ (vormals: „Verband der Handelsauskunfteien e.V.“) sind und Bonitätsauskünfte über Privatpersonen erteilen, angeschrieben. Eine Übersicht über die teilnehmenden Unternehmen findet sich im Anhang I.1.

Kfz-Versicherer

Ein weiterer Fragenbogen wurde an Anbieter von Telematik-Tarifen in der Kfz-Versicherung versandt. Es gibt keine Anbieter von Kfz-Versicherungen, die ausschließlich Telematik-Tarife im Portfolio haben, vielmehr handelt es sich um zusätzliche Bausteine, die zu bestehenden Kfz-Versicherungen angeboten werden.

Identifiziert wurden 15 Telematik-Tarife bzw. Telematik-Bausteine in der Kfz-Versicherung. Von den kontaktierten Versicherungsunternehmen nahmen fünf an der Befragung teil, die die Fragen teilweise auch für Tochterunternehmen beantworteten. Die Sparkassen Direktversicherung hatte zwischen 2013 und 2015 im Rahmen eines Pilotprojekts einen Telematiktarif getestet und nahm ebenfalls an der Befragung teil. Zusammengekommen liegen damit Aussagen zu zehn Telematik-Tarifen vor, eine Übersicht findet sich im Anhang I.2.

Gesetzliche und private Krankenversicherer

Weiterhin erfolgte im Rahmen dieser Marktstudie eine Befragung sämtlicher in Deutschland tätigen Krankenversicherer. Obgleich Gesundheits-Scoring hier nicht konkret praktiziert wird, sollten die einzelnen Bonusprogramme näher beleuchtet werden, um zu erarbeiten, in welcher Form und in welchem Ausmaß das Gesundheitsverhalten von Versicherten bereits erfasst und bonifiziert wird. Darüber hinaus wurden die Versicherer zu Einstellungen zu verhaltensbasierten Tarifen oder derartigen Plänen in der näheren Zukunft befragt.

Derzeit gibt es auf dem deutschen Markt 110 gesetzliche⁶⁸ und 43 private⁶⁹ Krankenversicherer, von denen insgesamt 62 (47 GKV und 15 PKV) an der Befragung teilnahmen⁷⁰ (siehe Anhang I.3).

67 Bei nicht-übereinstimmender Kodierung wurde die Kategorie mit zwei Übereinstimmungen gewählt.

68 Angabe des GKV-Spitzenverbandes, Abgerufen am 10. Juli 2018 von URL <https://www.gkv-spitzenverband.de/>.

69 Angabe des Verbandes der Privaten Krankenversicherung, Abgerufen am 10. Juli 2018 von URL <https://www.pkv.de/>.

70 Die teilnehmenden AOKen beantworteten einige der gestellten Fragen wortgleich (weiterführende Einschätzungen zum Thema verhaltensbasierte Tarife und Blick in die Zukunft). Bei der Auswertung wurde jedoch jede Antwort für jede an der Befragung teilnehmende AOK einmal (also insgesamt elf Mal) gewertet.

2. Fragebogen

Für jedes Marktsegment wurde ein spezifischer Fragebogen angefertigt (siehe Anhang II.). Die Fragebögen wurden Ende März 2018 an die oben genannten Anbieter per Post und per Email versendet. Eine Auswertung der Antworten ohne direkte Zuordnung zum Namen des Unternehmens wurde zugesichert.

Insbesondere wurden Fragestellungen zu folgenden Themenbereichen formuliert:

1. Geschäftsmodelle
2. Datenerhebung und Score-Berechnung
3. Aufsicht und Auskunftsrechte
4. Einschätzungen zum Thema Scoring und Blick in die Zukunft

Den Versicherern und Auskunfteien wurden zunächst einige Fragen zum Geschäftsmodell gestellt. Bei den Auskunfteien war etwa interessant zu erfahren, unter welchen Voraussetzungen Bonitäts-Scores weitergegeben werden. Bei den Telematik-Tarifen und Bonusprogrammen der Krankenversicherungen sollte außerdem u. a. herausgefunden werden, ob eine bestimmte Zielgruppe fokussiert wird und welche Merkmale die an solchen Programmen teilnehmenden Kunden auszeichnen. Gestellt wurden weiterhin Fragen zu den Vorteilen für die Teilnehmer und zu Auswirkungen auf deren Fahr- bzw. Gesundheitsverhalten.

Weiterhin gab es einen Fragenblock zur Datenerhebung und zur Berechnung der Scores für die Auskunfteien und die Kfz-Versicherungen. Welche Quellen bzw. Technik die Unternehmen zur Datenerhebung verwenden (z. B. Telematik-Stecker, Apps auf dem Smartphone), war ebenso von Interesse wie die Frage, wer die Berechnung der Score-Werte vornimmt. Es folgten Fragen zu den in die Scores einfließenden Merkmalen sowie den verwendeten Schätzmodellen und deren Güte.

Fragen zur Aufsicht und Qualitätssicherung bezogen sich auf Maßnahmen zu Gewährleistung von Aktualität und Richtigkeit der verwendeten Daten und Prüfung der Verfahren durch Aufsichtsbehörden.

Nach Fragen zu den Auskunftsrechten der Kunden und deren Möglichkeit zur Fehlerkorrektur wurden die Versicherungsunternehmen und Auskunfteien gebeten, ihre Einschätzungen zum derzeitigen Markt ihrer Branche, zu Scoring-Verfahren und zu möglichen Entwicklungen in der Zukunft zu geben. Welche Daten würden die Unternehmen gerne zusätzlich erheben und verwerten, gibt es aus ihrer Sicht Bedarf, Gesetze und Regulierungen anzupassen? Die Vor- und Nachteile von Scoring-Verfahren sollten ebenfalls erörtert werden. Vor allem bei den Krankenversicherungen stellte sich die Frage, ob verhaltensbasierte Tarife bereits konkret geplant und / oder als vorteilhaft bewertet werden, da diese ja derzeit noch nicht angeboten werden.

III. Diskussion der Ergebnisse und der sich daraus ergebenden Verbraucherprobleme

Eine ausführliche Deskription sämtlicher Ergebnisse der Marktstudie befindet sich im Anhang III. Im Anhang IV. befindet sich außerdem eine tabellarische Darstellung der Ergebnisse. Im Folgenden werden die zentralen Erkenntnisse und die sich ergebenden Verbraucherprobleme in Hinblick auf die in Kapitel B bearbeiteten Handlungsfelder diskutiert.

1. Verbreitung von Scoring in den betrachteten Marktberreichen

Während Bonitäts-Scoring von Wirtschaftsauskunfteien ein etabliertes Geschäftsfeld ist, zeigt die Marktstudie, dass die Verbreitung von Scoring-Verfahren in den Bereichen Kfz- und Krankenversicherung noch deutlich weniger fortgeschritten ist als man dies angesichts der öffentlichen Debatte vermuten könnte.

Von den rund 90 Versicherungsanbietern, die derzeit in Deutschland Kfz-Versicherungen vertreiben⁷¹, bieten – so ergab die Recherche – erst 16 Unternehmen Telematik-Tarife, die auf einem Scoring der Fahrweise eines Versicherten beruhen, an. Darunter befinden sich auch große Versicherungskonzerne mit hohen Versichertenzahlen, doch die Marktbefragung zeigt, dass der Anteil der Versicherten, die einen Telematik-Tarif nutzen, noch sehr gering ist. Bei einigen Unternehmen liegt die Quote bei unter einem Prozent der Gesamtversichertenzahl. Sechs der zehn Anbieter in dieser Umfrage sprechen mit ihren Telematik-Tarifen auch gezielt nur junge Fahrer und Fahranfänger an, da diese Versichertengruppe traditionell eher hohe Prämien zahlen muss und eine mögliche Reduzierung der Kosten für sie einen besonderen finanziellen Anreiz darstellt.



„Sachlicher Grund für diesen Produktzuschnitt ist, dass junge Erwachsene bis 25 Jahre im Straßenverkehr häufiger und schwerer verunglücken als Fahrer anderer Altersgruppen.“

Muss Telematik in der Kfz-Versicherung als ein Nischenprodukt gesehen werden oder wird es sich weiter verbreiten? Beide Einschätzungen sind in den Aussagen der Unternehmen zu finden. Einige Anbieter sind der Meinung, durch Telematik würden sich die Geschäftsmodelle in der Branche grundlegend verändern und der Trend zu verhaltensbasierten Tarifen sei nicht aufzuhalten.



„Wir glauben, dass der Kunde heute von seinem Versicherer mehr erwartet, als die Regulierung im Schadensfall.“

Ein Aspekt, der dagegen häufig als Negativargument für Telematik aus Unternehmensperspektive angeführt wird, sind die hohen Kosten, die sich für Versicherer aus dem Telematik-Tarif ergeben. Besonders für die Geschäftsmodelle, die eingebaute Telematik-Boxen nutzen, scheinen sich Aufwand und Kosten für Einbau der Technik und Hardware in Kombination mit zu leistenden Rabatten bei gutem Fahrverhalten nicht zu rentieren, wie die Aussagen in der Befragung zeigen. Dies ist womöglich auch der Grund dafür, dass Anbieter mitunter ganz auf einzubauende Hardware verzichten und Daten ausschließlich über eine App auf dem Smartphone erfassen. Eine Lösung der Kostenproblematik sehen

⁷¹ Angabe des GDV, Abgerufen am 10. Juli 2018 von URL <https://www.gdv.de/de/zahlen-und-fakten/versicherungsbereiche/ueberblick-4660>.

einige Anbieter in der Nutzung der eCall-Technologie⁷², die seit diesem Jahr in jedem neuen Fahrzeug verbaut wird. Die von den Befragten häufig geäußerte Befürchtung eines Datenmonopols seitens der Hersteller muss in diesem Kontext gesehen werden.



„Das Produkt Kfz-Versicherung ist mit den heute schon vorhandenen Tarifierungsmerkmalen sehr risikogerecht kalkuliert. Es bleibt daher abzuwarten, ob Telematik-Tarife tatsächlich bereits vorhandene Tarifierungsmerkmale ablösen oder zumindest ergänzen können.“

Verhaltensbasierte Tarife in der Krankenversicherung, bei denen sich die zu zahlenden Beiträge am fortlaufend verfolgten, d. h. gescorten Gesundheitsverhalten der Versicherten bemessen, gibt es in Deutschland nicht. Allerdings kann man bei fast allen gesetzlichen Krankenkassen bei Nachweis über die Absolvierung gesundheitsförderlicher Aktivitäten und Maßnahmen Boni erhalten. In der privaten Krankenversicherung scheinen sich vergleichbare Bonusprogramme nicht etabliert zu haben.⁷³ Die Bonusprogramme werden außerdem nur von wenigen Versicherten in Anspruch genommen: Wie die in der Befragung gemachten Angaben zeigen, sind in den meisten Fällen weniger als 10% der Versicherten auch Programmteilnehmer.

Der Marktstudie nach zeigen sich aber in den Bonusprogrammen – vor allem solchen, die mit Punktesystemen arbeiten – die ersten Ansätze von Gesundheits-

Scoring. Das Verhalten von Verbrauchern wird hier mit Punkten (also Zahlenwerten) belegt. Im Gegensatz zu Bonitäts-Scoring und Telematik bemessen Bonusprogramme das Gesundheitsverhalten der Versicherten bislang in der Regel aber auf analogem Wege, z. B. mit Bonusheften und Teilnahmebestätigungen und dem Aufaddieren der erzielten Punkte. Ein „Tracking“ von Verhaltensweisen mit Wearables und Apps wird bislang kaum genutzt, auch wenn ein Interesse daran an verschiedenen Stellen der Befragung deutlich wird. Komplexe statistische Verfahren oder gar Algorithmen, die einen statistischen Zusammenhang zwischen ergriffenen Verhaltensweisen und deren Auswirkungen auf die Gesundheit berechnen, kommen jedoch nicht zum Einsatz, weswegen grundsätzliche Fragen nach der Fairness der Bonusprogramme drängend sind.



„Die Digitalisierung verbunden mit dem Trend der Selbst-Optimierung der Menschen wird auch Auswirkungen in der Gesundheitswirtschaft haben. Die Skepsis gegenüber der Datenerfassung und -weitergabe bei digitalen Anwendungen und Therapien wird abnehmen. Wenn es zum Alltag gehört, seine Gesundheitsdaten zu teilen, werden Kunden auch erwarten, individuelle Präventionsangebote und Gesundheitservices zu erhalten. Künftig muss es deshalb mehr Spielraum geben, um entsprechende Angebote und Tarife zu gestalten.“

72 Hersteller müssen das von der Europäischen Union vorgeschriebene automatische Notrufsystem für Kraftfahrzeuge namens eCall (Kurzform für emergency call) seit dem 31. März 2018 in alle neuen Modelle von Pkw und leichten Nutzfahrzeugen einbauen. Im Fahrzeug montierte Geräte sollen einen Verkehrsunfall automatisch an die einheitliche europäische Notrufnummer 112 melden und durch die rascher initiierten Rettungsmaßnahmen die Zahl der Verkehrstoten senken. Die Einführung von eCall bedingt unter anderem die Ausstattung von Fahrzeugen mit einem Airbag-Sensor, der die Information bereit stellt, ob ein Unfall mit Verletzungsgefahr passiert ist, einem Satelliten-Empfänger, damit das Auto via GPS und/oder Galileo seinen Standort zum Zeitpunkt des Unfalles ermitteln kann (wird vielfach zusammen mit dem Navigationssystem genutzt), eine Mobilfunkantenne zum Senden des Notrufes an die Notrufzentrale und eine eCall-Steuerung (teilweise im Infotainment integriert), die die für den eCall benötigten Daten sammelt, nach einem Unfall übermittelt und eine Sprachverbindung zur Notrufzentrale öffnet.

73 Allerdings plant der Anbieter Generali sein App-basiertes Präventionsprogramm „Vitality“ ab Ende 2018 / Anfang 2019 auch in der Krankenversicherung anzubieten (Abgerufen am 1. Oktober 2018 von URL <https://www.generali.de/ueber-generali/presse-medien/pressemitteilungen/generali-vitality-wird-in-deutschland-weiterausgebaut-25562/>).

Obgleich Gesundheits-Scoring im engeren Sinne auf dem deutschen Versicherungsmarkt bislang nicht praktiziert wird, weisen die Aussagen der Versicherer auf eine zunehmende Relevanz des Themas hin. So zeigt sich ein Teil der antwortenden Versicherer offen für verhaltensbasierte Tarife. 8% der Antwortenden sagten aus, dass sie keinerlei Nachteile durch die Einführung verhaltensbasierter Tarife sehen (5 von 62 Versicherern).

Viele Versicherer wünschen sich außerdem beim Einsatz von Fitness-Trackern und der Belohnung von gemeinnützigen Verhaltensweisen (ohne individuellem gesundheitlichen Nutzen, s.u.) mehr Gestaltungsspielraum.

Über 40% der antwortenden Krankenversicherer können sich vorstellen, Daten aus der elektronischen Patientenakte ihrer Versicherten für Bonusprogramme oder verhaltensbasierte Tarife zu nutzen. Betrachtet man die Bonusprogramme zusätzlich im Kontext von kommerziellen Anbietern wie Dacadoo und privaten Versicherern wie Generali lässt sich das Entwicklungspotenzial von Gesundheits-Scoring und damit dessen steigende Verbraucherrelevanz erkennen.



„Auch der deutsche Gesundheitsmarkt wird sich dem nicht verschließen können, nur dauert in Deutschland alles länger.“

2. Transparenz

Verbrauchergerechtes Scoring setzt die Transparenz über die verwendeten Merkmale voraus, aus denen der Score gebildet wird. Die Befragung zeigt, dass es zwischen den betrachteten drei Marktsegmenten wesentliche Unterschiede in Bezug auf die Transparenz der Merkmale gibt.

Beim Bonitäts-Scoring wird deutlich, dass es für den Verbraucher schwer nachzuvollziehen ist, welche Daten zum Zwecke der Bonitätsauskunft gesammelt werden und welche Daten genau in die Berechnung seines Scores einfließen (vgl. zur Verbrauchersicht Kapitel D in diesem Gutachten). Es lässt sich festhalten, dass die verschiedenen Auskunftsteile teils unterschiedliche Merkmale für ihre Modelle nutzen. In der Marktbefragung zeigt sich der unterschiedliche Dateneinsatz für Scoring zum Beispiel in den Aussagen der Auskunftsteile zur Anzahl der verwendeten Merkmale: Eine Auskunftsteil nennt als Minimum der verwendeten Merkmale die Zahl Acht, eine andere Auskunftsteil die Zahl 25. Die Anzahl der verwendeten Merkmale variiert außerdem, je nachdem, um was für einen Branchenscore es sich handelt.

Die Unterschiede zwischen den Auskunftsteilen in Bezug auf die von ihnen für das Bonitäts-Scoring verwendeten Daten ergeben sich zum Teil aus den unterschiedlich zur Verfügung stehenden Datenquellen. Während alle Auskunftsteile Zugriff auf öffentlich zugängliche Register und Verzeichnisse sowie amtliche Bekanntmachungen haben, gilt dies nicht für Informationen von individuellen Vertrags- und Kooperationspartnern wie z. B. Finanzinstitute, (Versand-) Händler, Telekommunikations- oder Energieversorgungsunternehmen. Eine erschöpfende Auflistung der verwendeten Merkmale für die Score-Berechnung ist bei keiner Auskunftsteil zu finden.

Die Kunden von Telematik-Tarifen werden nach den Ergebnissen der Marktbefragung über die Kriterien, die in die Score-Bildung einfließen, in der Regel gut informiert. Üblicherweise bilden drei bis neun Merkmale wie Brems-, Beschleunigungs- und Geschwindigkeitsverhalten die Grundlage für die Scoring-Modelle. Über die Zusammensetzung des Scores wird nach Auskunft der Versicherer in der Befragung in den Vertragsunterlagen (Allgemeine Bedingungen für die Kfz-Versiche-

rung, Nutzungsvereinbarungen, Datenschutzerklärung etc.) und auch auf den einzelnen Webseiten und in den Telematik-Apps der Versicherungen aufgeklärt.

Die meisten Anbieter stellen laut Befragung ihren Telematik-Kunden in Aussicht, dass sie bei der Verbesserung ihrer Fahrweise unterstützt werden und dass sich insgesamt die Sicherheit im Straßenverkehr erhöhen wird. Tarifbezeichnungen wie „BetterDrive“ oder „Secure Drive“ suggerieren ebenfalls, dass durch eine Teilnahme am Telematik-Tarif die eigene Fahrweise verbessert wird und dementsprechend auch der Score die eigene Fahrsicherheit zum Ausdruck bringt. Eine solche Selbstdarstellung der Telematik-Anbieter ist dann kritisch zu sehen, wenn auch Merkmale wie Zeitpunkt einer Fahrt (z. B. Tag- oder Nachtfahrten) oder Ort einer Fahrt (z. B. Stadt- oder Landfahrten) in die Score-Berechnung einfließen, denn diese kann der Autofahrer selbst oft nicht beeinflussen. Von den zehn Anbietern in der Befragung berücksichtigen neun Anbieter Tag oder Uhrzeit, sieben Anbieter den Straßentyp, fünf Anbieter den Ort einer Fahrt, zwei Anbieter die Bevölkerungsdichte und ein Anbieter die Länge bzw. Dauer einer Fahrt. Bei all diesen Merkmalen handelt es sich um nur schwer beeinflussbare Parameter – es kann also nicht davon gesprochen werden, dass allein der Fahrstil (wie z. B. Bremsen, Beschleunigen, Kurvenverhalten) für die Höhe des Scores ausschlaggebend ist. Differenzen zwischen Werbeversprechen („besseres Fahren“) und eigentlichem Zielmerkmal des Scorings („Schadenseintrittswahrscheinlichkeit“) sind aus Verbrauchersicht problematisch einzuschätzen (siehe bereits B.I.5).

Ähnlich wie bei den Telematik-Tarifen der Kfz-Versicherungen erfahren die Teilnehmer an Bonusprogrammen der gesetzlichen Krankenversicherungen, mit welchen Aktivitäten sie Punkte sammeln können und ab wann sie Boni ausgezahlt bekommen. Eine verhaltenssteuernde Wirkung wäre andernfalls kaum zu erzielen. Die Befragung zeigt, dass vielen Bonusprogrammen ein langer Katalog von Maßnahmen und Aktivitäten zugrunde liegt. Nach einem Bericht der Verbraucherzentrale Nordrhein-Westfalen (2015) sind die Bonusprogramme aus Sicht der Versicherten sehr unübersichtlich. Eine Orientierung dahingehend, „wie viele und welche Maßnahmen wahrscheinlich in Betracht kommen und wie hoch der Gewinn unter den konkreten Umständen ausfällt“ sei erschwert. Ähnlich wie bei den Kfz-Versiche-

rungen gibt es bei den Bonusprogrammen teilweise das Problem, dass den Teilnehmern Gesundheitseffekte in Aussicht gestellt werden, aber auch Aktivitäten wie gemeinnützige Maßnahmen (Blutspende, Erste-Hilfe-Kurse etc.) bonifiziert werden. Eine eindeutige Darstellung dessen, was durch das Bonusprogramm erzielt werden soll, wäre erstrebenswert.



„Das Programm soll verantwortungsvolles und sicheres Fahren junger Autofahrer unterstützen.“

Im Kontext von Transparenz spielt auch eine Rolle, ob, wie und wie oft dem Verbraucher sein aktueller Score mitgeteilt wird. Hier zeigen sich zwischen den drei betrachteten Markt Bereichen ebenfalls erhebliche Unterschiede.

Beim Bonitäts-Scoring können Verbraucher eine kostenlose Selbstauskunft beantragen, um zu erfahren, welche Daten eine Auskunft über sie gespeichert hat und wie der aktuelle Score lautet. Doch hier gilt: Die Initiative muss vom Verbraucher ausgehen. Der Verbraucher muss sich aktiv darum bemühen, herauszufinden, bei wem welche Daten vorliegen. Dies ist im Hinblick auf den Grad der Bekanntheit der einzelnen Unternehmen problematisch zu sehen: Eine Verbraucherbefragung des Unabhängigen Landesentrums für Datenschutz Schleswig-Holstein und der GP-Forschungsgruppe (2014) ergab, dass die SCHUFA mit hohem Abstand die höchste Bekanntheit hat (98% der Befragten), gefolgt von Creditreform Boniversum (33% der Befragten). Die restlichen Auskunfteien sind weit unter 20% der Befragten bekannt. Vielen Verbrauchern ist die Rolle von Auskunfteien im Wirtschaftsleben also nicht bewusst und sie wissen nicht, wann welche Unternehmen sie scoren, z. B. wenn automatische Bonitätsabfragen beim Warenkauf im Onlinehandel erfolgen.

Bei den Telematik-Tarifen der Kfz-Versicherung, können Verbraucher in der Regel über die App der Versicherung jederzeit einsehen, wie hoch ihr aktueller Score ist. Nach jeder Fahrt bzw. wenn genügend Fahrdaten gesammelt wurden, wird der Score aktualisiert. Der Kunde kann

theoretisch nachvollziehen, wie sich einzelne Fahrten auf den Score auswirken.

Die Teilnehmer an Bonusprogrammen in der Krankenversicherung erkennen ihren Status häufig in Form eines Punktestandes. Wie häufig dieser mitgeteilt wird, variiert allerdings. Bei vielen Versicherungen lässt sich der Versicherte die Absolvierung von Aktivitäten und Maßnahmen analog in einem Bonusheft, Bonuspass etc. per Stempel oder Eintrag bestätigen oder reicht eine Teilnahmebescheinigung beim Versicherer ein. Ebenso häufig wird der aktuelle Stand im Bonusprogramm erst mitgeteilt, wenn bereits eine Auszahlung der Prämie erfolgt. Sieben Versicherer bieten an, den Punktestand im Bonusprogramm über eine App oder ein Onlineportal einzusehen.

Verbraucherfreundlicher erscheint es, wenn der Verbraucher, sofern er es möchte, jederzeit in Erfahrung bringen kann, welchen Score (oder Punktestand) er aktuell erreicht hat. Besonders bei einigen Krankenversicherungen gibt es daher erkennbares Entwicklungspotenzial.

3. Score-Berechnung und statistische Güte

Scoring-Verfahren sollen ein bestimmtes Verhalten oder bestimmte Verhaltenskonsequenzen möglichst genau prognostizieren.

Bei den Auskunfteien ist dieses Zielmerkmal (abhängige Variable) laut Aussagen der drei Unternehmen in dieser Befragung das Auftreten eines Negativmerkmals innerhalb eines bestimmten Zeitraums.

Bei den Kfz-Versicherungen sollen die Schadenseintrittswahrscheinlichkeit oder die Häufigkeit zukünftiger Schadensfälle vorausgesagt werden.

Das vorgegebene Ziel von Gesundheits-Scoring ist neben der Steuerung von Verbrauchern in Richtung gesundheitsförderliches Verhalten eine Vorhersage des zukünftigen Gesundheitsstatus des Gescorten. Auch in

den Bonusprogrammen lassen sich diese Dimensionen zu einem gewissen Grad ausmachen. Dem gesetzlichen Auftrag folgend, sollen sie gesundheitsbewusstes Verhalten fördern. Laut über 70% der Krankenkassen mit Bonusprogrammen in der Befragung ist dies auch der den Teilnehmern in Aussicht gestellte Effekt. Verbunden mit dem Ziel der Förderung gesundheitsbewussten Verhaltens ist die Vorstellung, dass damit eine zukünftige Senkung von Krankheitskosten einhergeht. Personen, die sich jetzt gesundheitsbewusst verhalten, würden in Zukunft seltener oder weniger schwerwiegend krank werden und der Versicherung somit weniger Ausgaben bescheren. Daher gilt in der Regel: Je mehr Aktivitäten im Bonusprogramm absolviert werden, desto höher der Bonus. Impliziert wird damit: je höher der Punktestand eines Teilnehmers, desto niedriger werden dessen Gesundheitskosten in der Zukunft sein.

Zur Berechnung der Scores können algorithmische Verfahren Anwendung finden. Die übliche Methode für die Berechnung von Bonitäts-Scores ist eine multiple logistische Regression. Eine Auskunftei wählte bei der Frage nach der verwendeten Schätzmethode zusätzlich noch die vorgegebenen Antwortmöglichkeiten „Entscheidungsbäume“, „Ensemblemethoden“ und „Neuronale Netze / Deep Learning“ aus.

Auch bei den Telematik-Scores scheint die logistische Regression häufig Anwendung zu finden.

Bei den Bonusprogrammen der Krankenversicherungen werden keine komplexen algorithmischen Entscheidungsverfahren angewendet – gibt es Punktesysteme, werden die Bonuspunkte einfach zu einem Gesamtpunktestand addiert. Gesundheits-Scores, deren Verfahren wesentlich ausdifferenzierter sind und denen von Auskunfteien und Telematik-Anbietern ähneln, finden sich bislang nur bei privaten Anbietern wie Dacadoo und Generali Vitality.

Alle drei untersuchten Branchen machen wenig aussagekräftige Angaben zur statistischen Qualität der von ihnen verwendeten Scores. Für die statistischen Modelle zur Score-Berechnung stehen verschiedene Gütemaße zur Verfügung, die beschreiben, wie sehr sich die Modelle eignen, Verbraucher hinsichtlich ihres Risikos (Kreditfallrisiko, Unfallwahrscheinlichkeit, Krankheitswahrscheinlichkeit usw.) zu klassifizieren.

Einige Auskunfteien machten in der Befragung Angaben zu den von ihnen verwendeten Gütemaßen, alle drei verwenden den Gini-Koeffizienten. Bei einer Auskunftei erreiche dieser einen Wert von 0.60 bis 0.75, bei einer anderen einen Wert zwischen 0.60 und 0.80⁷⁴. Diese Kennziffern machen freilich keinerlei Aussagen zu gruppenspezifischen Unterschieden der Prognosegüte, d. h. dem Ausmaß der in verschiedenen Dimensionen erreichten Fairness (siehe B.VI).

Bei den Kfz-Versicherungen blieben die Fragen nach objektivierbaren Gütekriterien wie der Höhe des Gini-Koeffizienten zur korrekten Identifikation „guter“ und „schlechter“ Autofahrer unbeantwortet.

Die Güte des Score-Modells hängt von den Merkmalen ab, die zur Score-Berechnung verwendet werden sowie der Gewichtung dieser Merkmale zueinander. Es sollten nur solche Merkmale genutzt werden, die einen statistisch signifikanten Einfluss auf das Zielmerkmal haben. Faktoren deren kausaler Einfluss stärker ist als andere, sollten stärker gewichtet werden und umgekehrt. Die Auskunfteien und Kfz-Versicherungsunternehmen wurden deshalb gefragt, anhand welcher Kriterien sie die Merkmale auswählen und die Gewichtungen festlegen. Auch bei den Krankenversicherungen interessierte, wie Aktivitäten und Maßnahmen ins Bonusprogramm aufgenommen werden und wie entschieden wird, welches Verhalten wie viele Punkte bringt.

Bei den Auskunfteien gestaltet es sich schwierig, Informationen über die Gewichtung der verwendeten Merkmale zu erhalten. Die Unternehmen berufen sich hier auf das Geschäftsgeheimnis (siehe B.I.1 und B.I.3). Die statistische Signifikanz der Merkmale stehe aber laut Aussage der Unternehmen sowohl bei Merkmalsauswahl als auch bei der Gewichtung im Mittelpunkt. Wenn ein Merkmal die Vorhersagegenauigkeit des Scores erhöht, wird es in das Modell mit aufgenommen. Die Gewichtung ergibt sich aus der Größe des statistischen Zusammenhangs zwischen Merkmal und Zielmerkmal. Interessant ist in diesem Zusammenhang, dass eines der befragten Unternehmen weitere Kriterien für die Merkmalsauswahl anführt, nämlich: daten-

schutzrechtliche Beurteilung, Verfügbarkeit, Kosten sowie Stabilität und Zuverlässigkeit des Merkmals. Gibt es zu einem Merkmal beispielsweise datenschutzrechtliche Bedenken, wird es nicht für die Score-Berechnung verwendet, selbst wenn es eine gewisse Vorhersagekraft innehat. Zur Datenqualität wurden keine spezifischen Aussagen gemacht; dass sie genannt wird deutet darauf hin, dass sie in der wissenschaftlichen Literatur, welche die formalen Eigenschaften von Scores stark betont, bislang unterschätzt wird (vgl. dazu Hand, 2005; Hand, 2006; Verbeke, Dejaeger, Martens, Hur & Baesens, 2012; Britz, 2008).



„Es gibt Merkmale, die signifikant sind, die sich aber dem menschlichen Verständnis erstmal intuitiv nicht so einfach erschließen. Bei solchen Merkmalen wird man entscheiden müssen, ob man diese in ein Modell aufnimmt oder nicht.“

Bei den Telematik-Anbietern ist es branchenüblich, bestehende Statistiken zu Fahrweisen und Unfallursachen, z. B. vom Statistischen Bundesamt oder der Royal Society for the Prevention of Accidents, auszuwerten, um die Merkmale zu identifizieren, die sich für die Prognose einer Schadenseintrittswahrscheinlichkeit eignen. Laut Auskunft der Unternehmen wird dann die statistische Signifikanz der einzelnen Merkmale bei der Modellbildung berücksichtigt.

Aussagen einiger Telematikanbieter, dass auch die menschliche Intuition und Heuristik aus unternehmerischen Erfahrungen bei der Auswahl der Variablen eine Rolle spielen, werden in Kapitel B.V reflektiert.

Bei den Bonusprogrammen der Krankenversicherung wird den Versicherten durch die Teilnahme ein präventiver Gesundheitseffekt in Aussicht gestellt. Den disku-

⁷⁴ Für ein perfektes Vorhersageinstrument hätte der Gini-Koeffizient den Wert 1.00. Ein Wert von 0.00 bedeutet, dass nur rein zufällig vorhersagt werden kann (siehe B.IV).

tierten Gütekriterien folgend sollte jede für einen Bonus zu absolvierende Aktivität auch tatsächlich gesundheitsförderlich sein und sich das Maß der Gesundheitsförderung auch in der relativen Bepunktung gegenüber anderen Aktivitäten widerspiegeln. Eine Maßnahme, die gesundheitsförderlicher ist als eine andere, sollte auch mehr Punkte bringen.

Bei der Auswahl der zu bonifizierenden Maßnahmen richten sich viele Krankenkassen laut der Marktbefragung danach, ob die Maßnahme gesundheitsförderlich ist. Sie folgen dabei den gesetzlichen Vorgaben gemäß § 65a SGB V sowie den Vorgaben des Bundesversicherungsamts (BVA) und des GKV-Spitzenverbandes⁷⁵. Einige wenige Krankenkassen orientieren sich jedoch auch an anderen Kriterien wie z. B. „Markttrends“ und Attraktivität der Aktivitäten für Versicherungen.

Das BVA kommt in seinem Sonderbericht zum Wettbewerb in der gesetzlichen Krankenversicherung zu dem Schluss, dass der gesundheitliche Nutzen einiger Bonusprogramme nicht qualitätsgesichert sei, „weil nicht nur zertifizierte, dem Leitfaden Prävention des GKV-Spitzenverbandes gemäß § 20 Abs. 2 SGB V entsprechende Präventionsangebote an einen Bonus gekoppelt werden, sondern auch ‚vergleichbare‘ qualitätsgesicherte Angebote gemäß § 65a Abs. 1 S. 1 Nr. 3 SGB V, wie das Erlangen eines Sportabzeichens oder die Mitgliedschaft in einem Fitnessstudio, hierfür genügen“ (Bundesversicherungsamt, 2018).

Besonders interessant ist es zu sehen, dass manche Krankenkassen ihren Versicherten ermöglichen, mit gemeinnützigen Aktivitäten Bonuspunkte zu sammeln, die keinen individuellen Gesundheitseffekt versprechen und somit keinen präventiven Charakter haben. Gleichwohl werden solche Programme gesundheitsförderlich dargestellt. Daneben stellt sich die Frage, ob die Bonifizierung von altruistischem Verhalten von § 65a SGB V gedeckt ist. Fünf Versicherer in der Befragung bonifizieren Blutspenden, Erste-Hilfe-Kurse, Knochenmarktypisierung, Organspendeausweise oder Patientenverfügungen. Begründet wurde das Vorgehen damit, dass die Versicherer auf diesem Wege altruistisches Verhalten

und die Gesundheitsförderung von Mitmenschen belohnen wollen.



„... [W]er nicht nur an sich selbst denkt, sondern sich auch aktiv für die Gesundheit anderer Menschen einsetzt, hat Bonuspunkte verdient.“

Das Bundesversicherungsamt, das die Rechtsaufsicht über die bundesunmittelbaren Träger der gesetzlichen Krankenversicherungen führt, genehmigt gemeinnützige Aktivitäten aufgrund der fehlenden Gesundheitseffekte nicht (Bundesversicherungsamt, 2018), weshalb nur wenige, landesunmittelbare Bonusprogramme so vorgehen können. In der Befragung wurde auch deutlich, dass einige Versicherer den Wunsch hegen, hier vom BVA mehr Gestaltungsspielraum eingeräumt zu bekommen. Dies gilt auch für den Einsatz von Fitness-Trackern. Bisher ermöglichen nur drei der befragten Krankenkassen das Sammeln von Bonuspunkten mit Daten, die über Fitness-Tracker gesammelt wurden. Vor allem Schrittzahl, aber auch Herzfrequenz und Kalorienverbrauch werden erfasst. Aus Verbrauchersicht stellt sich die Frage, wie qualitätsgesichert solche Ansätze sind. Haben die Krankenkassen Belege dafür, dass die mittels Fitness-Trackern erfassten Aktivitäten tatsächlich präventiv wirken? Oder werden vielmehr Marketing- und Kundenbindungseffekte angestrebt?

Auch die Gewichtung der Aktivitäten richtet sich bei den Bonusprogrammen augenscheinlich nicht unbedingt nach dem gesundheitsförderlichen Effekt. Bei vielen Versicherungen gibt es gar keine Gewichtung – alle Maßnahmen bringen gleich viele Punkte. Als häufigste Antwort auf die Frage zur Gewichtung wurde von den Versicherern der Aspekt des Aufwands und der Unkompliziertheit genannt. Den Versicherten soll es so leicht wie möglich gemacht werden, im Bonusprogramm Punkte zu sammeln.

⁷⁵ Im Leitfaden Prävention des GKV-Spitzenverbandes werden „die inhaltlichen Handlungsfelder und qualitativen Kriterien für die Leistungen der Krankenkassen und der Primärprävention und betrieblichen Gesundheitsförderung“ festgelegt. Abgerufen am 10. Juli 2018 von URL https://www.gkv-spitzenverband.de/krankenversicherung/praevention_selbsthilfe_beratung/praevention_und_bgf/leitfaden_praevention/leitfaden_praevention.jsp

Befinden sich Fehler in der Datengrundlage, kann dies unmittelbar Auswirkungen auf die Aussagekraft des Scores haben und zu weitreichenden Konsequenzen für Verbraucher führen. Deshalb wollten wir von den Unternehmen im Bereich Bonitäts-Scoring und Telematik wissen, welche Möglichkeiten Verbraucher haben, Fehler in ihren Daten korrigieren zu lassen.

Alle Auskunftsteien bieten Verbrauchern Korrekturmöglichkeiten, wenn diese feststellen, dass falsche oder veraltete Informationen über sie vorliegen. Dieser Weg kann allerdings nur beschritten werden, wenn auch die Kenntnis über falsche Daten vorhanden ist (siehe B.I.3), weshalb noch einmal auf die Transparenzproblematik beim Bonitäts-Scoring hingewiesen sei (siehe oben).

Best Practice stellen hier weiterhin die Unternehmen in der Befragung dar, die laut eigenen Angaben kontinuierliche Datenüberwachung und Einhaltung von Löschfristen praktizieren, um die Qualität ihres Datenbestandes zu wahren. Eine Auskunftstei gibt an, unter anderem Schulungen bei ihren Kunden, regelmäßige Datenanalysen, automatisierte Plausibilitäts-Checks sowie regelmäßige Bestandsabgleiche und Dateninventuren durchzuführen. Weiterhin unterlägen die Vertragspartner einer Aktualisierungspflicht, d. h. sie müssen Änderungen im Datenbestand melden.



„Wenn Verbraucher auf fehlerhafte Daten hinweisen, werden diese umgehend korrigiert.“

Zusätzlich zu den Angaben der Unternehmen in der Befragung ist zu erwähnen, dass eine Auskunftstei, die SCHUFA, Verbrauchern außerdem ermöglicht, sich bei Problemen an einen Ombudsmann, also einen externen, neutralen Schlichter wenden können. Voraussetzung ist, dass der Verbraucher sein Anliegen zuvor nicht mit dem Kundenservice des Unternehmens klären konnte. Laut aktuellem Tätigkeitsbericht gab es 2017 984 eingereichte Verbraucheranliegen (2016: 1.107). Davon waren 366 zulässig und in 42 Fällen wurde zugunsten des Verbrauchers entschieden. 618 Fälle

waren nicht zulässig, weil die für das Schlichtungsverfahren notwendige vorherige Kontaktaufnahme mit dem Privatkunden ServiceCenter der SCHUFA noch nicht stattgefunden hatte. In der Mehrzahl der Fälle ging es um das vorzeitige Löschen eines oder mehrerer negativer Einträge (SCHUFA Holding AG, 2018b). Im Tätigkeitsbericht des Hessischen Datenschutzbeauftragten des Jahres 2017 wird beschrieben, dass es sich bei den eingereichten Beschwerden bezüglich der SCHUFA um Fälle handelte, in denen negative Bonitätsinformationen falschen Personen zugeordnet worden waren. Dies sei durch fehlerhafte manuelle Bearbeitung verursacht worden (Der Hessische Datenschutzbeauftragte, 2017).



„Wichtig ist daher zum einen, das Möglichste zu tun, um solche Messfehler zu erkennen und auszuschließen. Zum anderen aber auch dafür zu sorgen, dass es ‚Toleranzen‘ gibt, so dass nicht der eine fehlerhafte Datenpunkt den Score eines Fahrers gleich dauerhaft zerstört. Und zu guter Letzt klar und deutlich zu sagen, dass jeder Fahrer ein gewisses Maß an fehlerhaften Daten in seinen Aufzeichnungen haben wird [und dass] somit aber auch jeder Fahrer davon betroffen ist und dies in der Kalibrierung der Algorithmen also schon mit enthalten ist.“

Wenn Telematik-Kunden Fehler bei der Aufzeichnung der Fahrdaten auffallen, sollten sie diese ebenfalls vom Anbieter prüfen und gegebenenfalls korrigieren lassen können. Nicht alle Anbieter ermöglichen das ihren Kunden, wie die Befragung zeigt. Problematisch ist dies vor allem angesichts der Tatsache, dass eine gewisse Fehleranfälligkeit bei den Unternehmen bekannt ist. So wird im Rahmen der Befragung von Fällen berichtet, in denen falsche Fahrdaten zustande kommen, weil das Kartenmaterial veraltet ist, oder weil es Störungen beim GPS-Signal oder beim Übertragen der Daten gab.

4. Verhaltenseffekte

Bei Scoring-Verfahren mit verhaltenssteuernden Komponenten stellt sich die Frage, ob die gewünschten Effekte auf diesem Wege tatsächlich erreicht werden können. Verbessert sich die Schadenhäufigkeit von Versicherten durch die Teilnahme an Telematik-Tarifen? Ändert sich der Gesundheitszustand der Teilnehmer an Bonusprogrammen ihrer Krankenversicherung? Gefragt wurden die Versicherer danach, ob sie selbst Erkenntnisse über solche Effekte haben und wie sie zu diesen gelangten.



„Erste Analysen bestätigen die Hypothese, dass Kunden, die den Telematik-Baustein gewählt haben, im Vergleich zum durchschnittlichen Kunden ohne Telematik-Baustein vorausschauender fahren.“

Bei den Kfz-Versicherungen scheint das übliche Vorgehen zu sein, Telematik-Kunden und Kunden ohne Telematik-Tarif im Hinblick auf deren Schadenhäufigkeit oder verursachten Kosten zu vergleichen – eine „echte“ Kontrollgruppe, die eine Vergleichbarkeit beider Gruppen hinsichtlich aller relevanten Merkmale herstellt, scheint von den Anbietern nicht gebildet zu werden.

Ein Kfz-Versicherer räumt explizit ein, dass nicht ausgeschlossen werden könne, dass ohnehin umsichtige und sichere Fahrer eher bereit sind, Telematik-Angebote wahrzunehmen. Finden Vergleiche zwischen Teilnehmern und Nicht-Teilnehmern in Bezug auf Schadenhäufigkeit statt, sind die Ergebnisse aus diesem Grund verzerrt. Die Schadenhäufigkeit der betrachteten Personen war ohnehin niedrig und sank nicht nur durch die Teilnahme.



„Kunden, die wissen, dass sie umsichtig und verantwortungsvoll Auto fahren, sind auch eher bereit, an einem Telematik-Tarif teilzunehmen.“

Ein ähnliches Vorgehen zeigt sich auch bei den Krankenkassen. Die meisten Versicherer weisen die Gesundheitseffekte indirekt durch Nachweis der Senkung der Leistungsausgaben bzw. der Gesundheitskosten von Bonusprogrammteilnehmern gegenüber Nicht-Teilnehmern nach. Nach § 65a SGB V müssen die Aufwendungen für Maßnahmen der Bonusprogramme mittelfristig aus Einsparungen und Effizienzsteigerungen, die durch diese Maßnahmen erzielt werden, finanziert werden. Kriterien für die Evaluationsberichte, mit denen die Krankenkassen ihrer Rechenschaftspflicht nachkommen, wurden vom Bundesversicherungsamt 2005 den bundesunmittelbaren Krankenkassen in einem Rundschreiben⁷⁶ mitgeteilt⁷⁷. Folgende Mindestanforderungen gibt das BVA darin unter anderem vor:

- Eine „nicht-randomisierte kontrollierte Studie“ ist erforderlich. Es muss sich um eine Voll- und nicht nur um eine Stichprobenerhebung handeln.
- „Kontrollgruppenverfahren (Matching-Verfahren)“ müssen angewendet werden. Die Kontrollgruppen müssen ausreichend groß sein, und den wesentlichen soziodemografischen Merkmalen der Teilnehmer entsprechen. Konkret werden für das Matching die Kriterien Alter, Geschlecht, Versichertenstatus, Region und Kosten vorgegeben.
- Ein „Vorher- / Nachhervergleich (Prä-Post-Vergleich)“ hat zu erfolgen.
- Alle Kosten (für Errichtung, Durchführung, Dokumentation und Evaluation der Bonusprogramme) müssen berücksichtigt werden.

⁷⁶ Abgerufen am 1. Oktober 2018 von URL https://www.bundesversicherungsamt.de/fileadmin/redaktion/Risikostrukturausgleich/Rundschreiben/Rundschreiben_Bonus.pdf

⁷⁷ Das BVA räumt ein, dass bei der Prüfung der Wirtschaftlichkeit auf die Datenbasis der Krankenkassen vertraut werden müsse und, dass nur geprüft werden könne, ob die Evaluationsberichte vollständig und plausibel sind sowie die methodischen Vorgaben des Rundschreibens eingehalten wurden (Bundesversicherungsamt, 2018). Im Tätigkeitsbericht 2016 des BVA heißt es, lediglich acht von 66 bundesunmittelbaren Krankenkassen hätten die Wirtschaftlichkeit ihrer Bonusprogramme für die Vergangenheit nicht nachweisen können (Bundesversicherungsamt, 2016).

Diese Vorgaben sind methodisch sinnvoll, insbesondere da die Kosten (vor Wahl eines Bonus-Tarifs) berücksichtigt werden sollen. D. h. die Bonus-Gruppe soll mit Gruppen verglichen werden, die zuvor dieselben durchschnittlichen Kosten verursacht haben. Ob diese Vorgaben von den Krankenkassen eingehalten werden, insbesondere was die korrekte Durchführung von Kontrollgruppenvergleichen angeht, ist schwer nachzuverfolgen. In der Befragung machten die Krankenkassen keine näheren Angaben zu ihren Evaluationsverfahren. Das BVA räumt ein, dass bei der Prüfung der Wirtschaftlichkeit auf die Datenbasis der Krankenkassen vertraut werden müsse und, dass nur geprüft werden könne, ob die Evaluationsberichte vollständig und plausibel sind sowie die methodischen Vorgaben des Rundschreibens eingehalten wurden (Bundesversicherungsamt, 2018). Im Tätigkeitsbericht 2016 des BVA heißt es, lediglich acht von 66 bundesunmittelbaren Krankenkassen hätten die Wirtschaftlichkeit ihrer Bonusprogramme für die Vergangenheit nicht nachweisen können (Bundesversicherungsamt, 2016).

Ähnlich wie bei den Telematik-Tarifen der Kfz-Versicherung ist es bei den Bonusprogrammen wahrscheinlich, dass vor allem gesundheitsbewusste Versicherte partizipieren, deren Gesundheitskosten ohnehin gering sind. Diesen Eindruck bestätigt zum Beispiel auch eine Studie des Robert-Koch-Institutes, die zeigte: „Hohes Gesundheitsbewusstsein und gesundheitsbewusste Gesundheitsverhalten wie Nichtrauchen und sportliche Aktivität sind wichtige Einflussfaktoren für eine Teilnahme am Bonusprogramm. Die Ergebnisse sind Hinweise darauf, dass insbesondere Versicherte an [Bonusprogrammen] teilnehmen, die sich im Allgemeinen bereits sehr gesundheitsbewusst verhalten und im Sinne eines sogenannten Mitnahmeeffektes die präventiven Maßnahmen ohnehin wahrgenommen hätten“ (Jordan, von der Lippe, Starker, Hoebel & Franke, 2015).

Aus Verbrauchersicht wäre es zu begrüßen, wenn der Verhaltenseffekt, der durch Scoring-Angebote angestrebt wird, auch wissenschaftlich fundiert ist. Dies gilt sowohl für die einzelnen Verhaltensweisen bzw. Merkmale, die in der Score einfließen, als auch das Programm in seiner Gesamtheit. Und es sollte breit(er) diskutiert werden, ob Bonus-Programme, die überwiegend von ohnehin Gesunden wahrgenommen werden, gesellschaftlich erwünscht sind (selbst wenn diese Program-

me sich auch bei genauer Betrachtung als wirtschaftlich erweisen, d. h. durch das Bonusprogramm die Gesundheit von ohnehin Gesunden nochmals ansteigt).

5. Diskriminierung

Während es Sinn des Scorings ist, Personen voneinander zu unterscheiden und individuell zu bewerten, sollen sie dabei aber nicht ohne Grund benachteiligt oder diskriminiert werden (siehe B.II.2). Ein Anknüpfungspunkt ist hierbei die Frage, welche Merkmale für das Scoring verwendet werden. Einige Merkmale, wie ethnische Herkunft oder Religion, sind durch das AGG geschützt und ihre Verwendung daher von vornherein nur unter bestimmten Umständen erlaubt. Doch auch bei anderen Merkmalen ist zu diskutieren, ob sich durch ihre Einbeziehung ins Scoring Benachteiligungen für bestimmte Verbraucher(-gruppen) ergeben, die gesellschaftlich nicht akzeptabel sind. Diskussionswürdig erscheinen insbesondere solche Merkmale, die vom individuellen Verbraucher nicht oder nur schwer beeinflussbar sind, weshalb eine Bewertung anhand dieser Kriterien unfair erscheint (Vergleiche auch Ergebnisse der Bevölkerungsbefragung, Kapitel D).

Ein viel diskutierter Aspekt im Zusammenhang mit Bonitäts-Scoring ist das Geoscoring (siehe auch B.III.3). Geoscoring ist aus Verbrauchersicht problematisch, weil der Gescorte in „Sippenhaft“ für die Zahlungsmoral seiner Nachbarschaft genommen wird und es nicht mehr um das individuelle Zahlungsverhalten geht. Damit besteht in Bezug auf dieses Merkmal keine Möglichkeit, auf die Höhe des eigenen Scores Einfluss zu nehmen. Verschiedene Daten- und Verbraucherschützer haben dieses Vorgehen daher bereits als diskriminierend bezeichnet (z. B. Schaar, 2008; siehe auch Britz, 2008).

Während eine Auskunft in der Befragung bis auf Ausnahmefälle Geoscoring komplett ausschließt, gibt eine andere an, regelmäßig „Zahlungserfahrungen im unmittelbaren Wohnumfeld“ zu berücksichtigen.

Auch Scoring anhand von Kriterien wie Alter oder Geschlecht kann aufgrund der Unbeeinflussbarkeit dieser

Merkmale als diskriminierend aufgefasst werden. Zwei Auskunfteien gingen bei der Frage zu potenziellen Diskriminierungsproblemen auf die Variable Geschlecht ein. Eine Auskunftei schließt dessen Verwendung von vorne herein aus, die andere gibt an, diese nur zu verwenden, wenn sie sich im Modell als „trennscharf“ erweise.

Zwei der drei Auskunfteien sagen also, dass sie bereit sind nach dem Geschlecht zu unterscheiden, wenn sich dies aus den Daten ergibt. Aus dieser Praxis können diskriminierende Effekte resultieren (siehe E.III.5). Die dritte Auskunftei, die das ausschließt, ignoriert freilich indirekte Diskriminierung durch Korrelation von Merkmalen wie z. B. Geschlecht oder Nationalität mit anderen – zulässigen – Merkmalen. Nun mag es sein, dass faktisch das Geschlecht und die Nationalität empirisch für die errechneten Scores keine Rolle spielen, aber es überrascht, dass das Problem der Diskriminierung – sei sie direkt oder indirekt – von manchen Unternehmen grundsätzlich abgestritten wird. Die Aussage einer Auskunftei, dass mathematisch-statistische Verfahren nicht diskriminieren würden, verkürzt die Problematik in unzulässiger Art und Weise (siehe B.II und E.III.5).



„Mathematische, statistische Verfahren diskriminieren nicht.“

Diskriminierungs- oder Benachteiligungspotenziale der Bonusprogramme der Krankenversicherungen müssen auch im Kontext ihrer Rahmenbedingungen betrachtet werden: Die gesetzliche Krankenversicherung fußt auf dem Solidaritätsprinzip. Zusammengefasst bedeutet dies, dass die Erkrankungsrisiken von allen Versicherten gemeinsam getragen werden. Die Beitragsbemessung orientiert sich an dem Einkommen der Versicherten und nicht an Merkmalen wie Alter, Geschlecht oder Gesundheitsverhalten. Der Leistungsanspruch richtet sich nach der individuellen Bedürftigkeit, alle Versicherten sind in gleichem Maße abgesichert.

Die Teilnahme an einem Bonusprogramm hat auf die Beitragsbemessung eines Versicherten also keinen Einfluss. Allerdings lässt sich argumentieren, dass durch Boni in Form von Geld- und Sachleistungen die individuellen Kosten für die Krankenversicherung sinken. Potenziell sind bei einigen Krankenkassen Boni von mehreren hundert Euro pro Jahr möglich⁷⁸. Es stellt sich also die Frage, ob durch die Bonusprogramme bestimmte Verbrauchergruppen benachteiligt werden, weil ihnen die Teilhabe an solchen Vergünstigungen und damit niedrigeren Versicherungskosten erschwert oder verwehrt wird.

Die Bonusprogramme stehen, wie auch die Befragung zeigt, grundsätzlich allen Versicherten offen. Doch nicht alle Maßnahmen scheinen auch für alle Verbraucher gleichermaßen absolvierbar. Beispielsweise setzen sportliche Aktivitäten, Teilnahmen an Sportveranstaltungen oder Sportkursen eine gewisse Mobilität und körperliche Verfassung voraus. Kranken, körperlich beeinträchtigten oder alten Menschen wäre es erschwert oder verwehrt, auf diesem Wege Bonuspunkte zu sammeln. Ähnliche Probleme finden sich, wenn der Nachweis von Körperdaten und Laborwerten (BMI, Blutzuckerspiegel etc.) bonifiziert wird (anstatt z. B. positive Veränderungen dieser Werte zu bonifizieren). Auch strukturell bedingte Benachteiligungseffekte sind denkbar: Bei einer Reihe von Maßnahmen muss der Versicherte in finanzielle Vorleistung gehen (z. B. Fitnessstudio). Versicherte, die sich dies nicht leisten können, werden damit ebenfalls potenziell benachteiligt.



„Verhaltensbasierte Tarife können dazu führen, dass einzelne Versichertengruppen dies auf Kosten von Menschen, deren Krankheiten nicht verhaltensbedingt sind, ausnutzen. Insofern sehen wir hier sehr kritisch drauf.“

⁷⁸ Eine Übersicht über die möglichen Boni in den Bonusprogrammen der einzelnen Krankenkassen findet sich unter: <https://www.gesetzlichekrankenkassen.de/bonuszahlung/bonuszahlung.html> (Abgerufen am 1. Oktober 2018).

Weiterhin kann diskutiert werden, ob mit den Bonusprogrammen eine indirekte Risikoselektion vorgenommen wird, wenn sowieso schon Gesunde und Gesundheitsbewusste verstärkt profitieren. Die Befragung zeigt, dass einige Krankenkassen gezielt gesundheitsbewusste Versicherte ansprechen wollen. Dies kann als ein Versuch gewertet werden, möglichst junge und gesunde Versicherte dauerhaft an sich zu binden.



„Durch die möglicherweise eintretende Antiselektion kann eine gute Absicherung der Gesundheit für Personen, die sich nicht für verhaltensbasierte Tarife entscheiden, gegebenenfalls zu deutlich erhöhten Prämien führen. Hintergrund ist, dass sich gute Risiken eher für die günstigeren verhaltensbasierten Tarife entscheiden und die Beiträge für die Personen in den nicht verhaltensbasierten Tarifen damit ansteigen werden, da sich die Risikomischung verschlechtert.“

In verhaltensbasierten Tarifen wäre dieser Effekt sicherlich verstärkt vorzufinden. Die fortlaufende Orientierung der individuellen Beiträge am jeweiligen Gesundheitsstatus und Gesundheitsverhalten würde automatisch zu einer Benachteiligung gesundheitlich beeinträchtigter Verbraucher führen.

Die Marktbefragung zeigte jedoch auch, dass die meisten (gesetzlichen) Versicherer dem Solidaritätsprinzip einen hohen Wert zuschreiben und dieses schützen wollen. Von knapp 50 % der Befragten wird die Gefährdung des Solidaritätsprinzips oder die Gefahr von Benachteiligung und Diskriminierung als Nachteil von verhaltensbasierten Tarifen genannt. 30 % der Befragten führten das Solidaritätsprinzip als Hauptgrund gegen eine Einführung von Malus-Systemen, bei denen gesundheits-schädliches Verhalten bestraft wird, an.

Auch bei den Telematik-Tarifen in der Kfz-Versicherung ist es sinnvoll, etwaige Diskriminierungseffekte durch die verwendeten Merkmale zu prüfen. Hinter der Ein-

beziehung von Scoring-Variablen wie Zeit einer Fahrt, Ort einer Fahrt oder genutzter Straßentyp liegt die Annahme, dass mit bestimmten Rahmenbedingungen einer Fahrt ein erhöhtes Unfallrisiko einhergeht. Es stellt sich aber die Frage, ob die Verwertung solcher vom Fahrer nur schwer beeinflussbaren Merkmale bestimmte Verbrauchergruppen benachteiligt. Ein Telematik-Nutzer, der in der Stadt wohnt, und daher auch viel in der Stadt fährt, wird regelmäßig einen schlechteren Score erhalten, als ein Fahrer, der aufgrund seiner Lebensverhältnisse vor allem Überlandfahrten tätigt. Ein Nachtschichtarbeiter, der vor allem nachts den Weg zu seiner Arbeitsstätte zurücklegt, wird weniger gut bewertet werden, als jemand, der vor allem tagsüber fährt. Ein solches Vorgehen erscheint problematisch, da dadurch eine strukturelle Benachteiligung einzelner Verbrauchergruppen nicht auszuschließen ist. Einzelne Verbrauchergruppen könnten aufgrund von Faktoren benachteiligt werden, die sie selbst kaum beeinflussen können. Außerdem stellt sich, wie bereits diskutiert, die Frage, ob über die Verwendung speziell dieser Merkmale so ausführlich aufgeklärt wird, dass der Verbraucher über die Sinnhaftigkeit einer Telematik-Option in seiner individuellen Lebenssituation selbst entscheiden kann.

6. Verknüpfung von Daten, Einbeziehung neuer Verbrauchermerkmale

Die Unternehmen in der Marktbefragung wurden gefragt, ob sie sich vorstellen können, auch Daten aus anderen Lebensbereichen ihrer Kunden für ihr Scoring zu verwenden. Diese wurde von allen Umfrageteilnehmern verneint.

Aus Verbrauchersicht wäre es beispielsweise kritisch, wenn die Datenbasis der Scoring-Verfahren von Auskunfteien um Merkmale aus intuitiv bonitätsirrelevanten Bereichen, wie z. B. Fahr- oder Gesundheitsverhalten, erweitert würde. Die würde einer Verknüpfung verschiedener Lebensbereiche entsprechen und einem bereits in Kapitel B diskutierten Super-Score nahekommen. Insofern ist es interessant, dass alle Auskunfteien

in der Befragung angeben, den sich bereits im Ausland abzeichnenden Trend der Einbeziehung von Internet- und Social-Media-Daten in das Bonitäts-Scoring (siehe B.VIII) zu beobachten. Eine Auskunftei spricht in diesem Kontext sogar von einem erheblichen Entwicklungspotenzial. Erwähnenswert ist in diesem Zusammenhang, dass die SCHUFA bereits 2012 eine Zusammenarbeit mit dem Hasso-Plattner-Institut (HPI) eingegangen war, um Grundlagenforschung rund um die technische Verarbeitung öffentlicher Web-Daten zu betreiben. Der Vertrag wurde allerdings vom Forschungsinstitut aufgrund des öffentlichen Drucks wieder gekündigt.⁷⁹

Bei einigen Krankenversicherungen zeichnet sich ein Interesse ab, die Datenbasis um weitere Gesundheitsdaten zu erweitern. Rund 37% der Befragten können sich vorstellen, Daten aus der elektronischen Patientenakte für den Bereich Bonusprogramme oder verhaltensbasierte Tarife zu nutzen.

Nur sehr wenige (nämlich fünf gesetzliche) Krankenversicherer in der Befragung befürworten die Möglichkeit, bei der Ausgestaltung von Bonusprogrammen oder aber auch bei der Prämiengestaltung von Krankenversicherungstarifen neben gesundheitsförderlichen auch gesundheitsschädliche Verhaltensweisen von Versicherten zu berücksichtigen (zur Diskussion der Problematik von Bonus- vs. Malus-Systemen siehe auch Kapitel D).



„Bei der Ausgestaltung von Bonusprogrammen ist es aus unserer Sicht durchaus vorstellbar auch gesundheits-schädliche Verhaltensweisen bei der Ermittlung des Bonus entsprechend zu berücksichtigen. Durch ein Bonus-Malus-System bei einem freiwilligen Bonusprogramm kann sicherlich ein besserer Effekt als bei einem reinen Bonus-System erzielt werden, da der Ansatz dann ganzheitlicher ist.“

Bei den Kfz-Versicherungsunternehmen ist eine Bestrebung, die Datenbasis um neue Verbrauchermerkmale zu erweitern, aus der Befragung nicht erkennbar. Grundsätzlich gilt die Meinung, nur Merkmale mit Bezug zum Fahrverhalten der Kunden seien für das Scoring interessant. Die Hälfte der Anbieter schätzt die Akzeptanz eines Malus-Systems als sehr niedrig bzw. überhaupt nicht vorhanden ein (eine Einschätzung, die sich mit den Erkenntnissen aus der Bevölkerungsbefragung deckt, siehe Kapitel D).

7. Aufsicht

Da sich die betreffenden Unternehmen bei ihren Scoring-Verfahren auf das Geschäftsgeheimnis berufen können und diese Verfahren somit nicht öffentlich machen müssen (siehe B.I.3), ist eine umfassende Aufsichtspraxis im Sinne der Verbraucher. Verbraucher, welche die Scoring-Verfahren nicht im Detail nachvollziehen können, müssen sie sich darauf verlassen können, dass die zuständigen Aufsichtsbehörden die Scoring-Algorithmen kritisch und gründlich prüfen können und tatsächlich prüfen (siehe B.III).

Die Auskunfteien werden regelmäßig durch die Aufsichtsbehörden für den Datenschutz überwacht. Zuständig sind jeweils die Datenschutzbeauftragten in dem Bundesland, in welchem sich der Hauptsitz des Unternehmens befindet.

In der Befragung geben die drei antwortenden Unternehmen an, den zuständigen Landesdatenschutzbehörden ihre Verfahren offengelegt zu haben. Es stellt sich die Frage, wie sich die Aufsicht der Datenschutzbehörden in der Praxis darstellt: Während eine Auskunftei angibt, ihre Scoring-Verfahren sämtlichen Datenschutzbeauftragten des Bundes und der Länder offengelegt zu haben, berichtet eine andere Auskunftei ausschließlich von regelmäßigen Treffen mit dem zuständigen Datenschutzbeauftragten.

⁷⁹ Siehe Pressemitteilung des HPI, abgerufen am 1. Oktober 2018 von URL <https://hpi.de/pressemitteilungen/2012/schufa-forschungsprojekt-gekuendigt.html>.

Der Hessische Datenschutzbeauftragte schrieb in seinem Tätigkeitsbericht für das Jahr 2014: „In keinem anderen Bereich als der Tätigkeit von Handelsauskunfteien gehen bei mir mehr Beschwerden ein“ (Der Hessische Datenschutzbeauftragte, 2014).

Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit erklärt in seinem Tätigkeitsbericht 2014/2015 in Bezug auf Auskunfteien: „Für Datenschutzaufsichtsbehörden ist es nahezu nicht möglich, selbst zu prüfen, ob die zur Berechnung eines Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich sind“ (Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit, 2016). Im Mai 2018 zeigte eine Recherche von Reportern des Bayrischen Rundfunks, dass die Überprüfung der Scoring-Verfahren durch die Datenschutzbehörden zu einem wesentlichen Teil auf der Grundlage von Gutachten basiert, die die Auskunfteien selbst bei Universitäten und Wissenschaftlern in Auftrag gegeben haben (Kerler, Köppen, Schnuck & Zierer, 2018).

Einige Datenschützer, wie der ehemalige Datenschutzbeauftragte des Landes Schleswig-Holstein, Thilo Weichert, der frühere Bundesdatenschutzbeauftragte Peter Schaar oder die Datenschutzbehörde Nordrhein-Westfalen, äußern sich in dem Report kritisch gegenüber dieser Praxis. Sie sehen Interessenskonflikte, wenn die Gutachten von den Auskunfteien bezahlt werden und fordern unabhängige Prüfungen sowie die dazu notwendigen finanziellen Mittel. Die Behörde des Hessischen Datenschutzbeauftragten, welche über mehrere Auskunfteien Aufsicht führt, hält das Vorgehen dagegen für unproblematisch. Es sei immer auch die Möglichkeit gegeben, zusätzliche eigene Gutachten zu beauftragen, was allerdings bislang nie erforderlich gewesen sei.

Bezüglich der Aufsicht bei Telematik-Angeboten der Kfz-Versicherung ergibt sich in der Befragung ein heterogenes Bild. Fünf der befragten zehn Anbieter geben an, es gäbe überhaupt keine staatliche Aufsicht in Bezug auf Telematik in der Kfz-Versicherung. Vier Anbieter nennen dagegen die BaFin als zuständige Aufsichtsbehörde. Diese Divergenz der Antworten dürfte durch den Umstand erklärbar sein, dass zwar die BaFin die Aufsicht über

Versicherungsunternehmen führt (siehe im Einzelnen § 4 Abs. 1 FinDAG, §§ 320 ff. VAG), dabei aber kein „telematikspezifisches“ Aufsichtsprogramm verwirklicht (siehe E.IV). Nähere Hinweise auf eine Überprüfung der Scoring-Verfahren durch eine Aufsichtsbehörde werden in der Unternehmensbefragung nicht gebracht.

Die Krankenkassen müssen nach § 65a Abs. 3 SGB V der zuständigen Aufsichtsbehörde mindestens alle drei Jahre Rechenschaft darüber ablegen, dass sie ihre Bonusprogramme durch Einsparungen und Effizienzsteigerungen finanzieren. Laut BVA „muss bei der Prüfung der Wirtschaftlichkeit auf die Datenbasis der Krankenkassen vertraut werden, die dem Evaluator zur Verfügung gestellt wird, da keine eigenen Daten durch das Bundesversicherungsamt erhoben werden.“ (Bundesversicherungsamt, 2018). Es könne daher nur geprüft werden, ob die Evaluationsberichte vollständig und plausibel sind.

Für den Verbraucher, denen ein Gesundheitseffekt versprochen wird, ist problematisch, dass aus den Evaluationsberichten für das BVA nicht ableitbar ist, ob die Teilnahme an einem Bonusprogramm einen objektiven gesundheitlichen Nutzen bedeutet. Das BVA kann lediglich prüfen, ob die Bonusprogramme das Kriterium der Wirtschaftlichkeit erfüllen, nicht aber, ob sich gesundheitsbewusstes Verhalten bei den Teilnehmern erhöht hat.

Aus Verbrauchersicht weiterhin kritisch zu bewerten ist außerdem die unterschiedliche Zulassungspraxis zwischen dem für bundesunmittelbare Träger der gesetzlichen Krankenversicherung zuständigen BVA und regionalen Aufsichtsbehörden. Dies betrifft zum Beispiel die Zulassung gemeinnütziger Aktivitäten zur Erlangung von Boni sowie den Einsatz von Fitness-Trackern. Rund 20 % der Versicherer in der Befragung sprechen diesen Umstand bei der Frage nach einem möglichen Bedarf, den Gesetzes- und Regulierungsrahmen anzupassen, an.

D

Wissen über und
Wertungen von Scoring
in der Bevölkerung



I. Vorstudie 2017

Empirische Informationen über das Wissen der Bevölkerung in Deutschland zum Thema Scoring und algorithmischen Entscheidungen und deren Bewertung sind nur spärlich vorhanden (siehe B.VII). Deswegen hat der SVRV eine repräsentative Bevölkerungsbefragung in Zusammenarbeit mit dem Erhebungsinstitut infas durchgeführt. Diese Befragung wurde durch eine nicht-repräsentative Vorstudie vorbereitet. Die Ergebnisse beider Studien werden in diesem Kapitel berichtet.

Im November 2017 wurde eine Vorstudie in Berlin durchgeführt, um Teile des für die repräsentative Befragung entwickelten Fragebogens zu testen. Die Studie wurde in einem Berliner Kinosaal mit 91 Teilnehmern (52 % Frauen), im Alter zwischen 17 und 75 Jahren durchgeführt (vgl. Rebitschek, Gross, Brümmer, Gigerenzer & Wagner, 2018).

Eingesetzt wurden Paper-and-Pencil-Fragebögen (Rebitschek, Gross, Brümmer, Gigerenzer & Wagner 2018) mit Fragen u. a. zu Wissen (Merkmale, die in etablierte Bonitäts-Scorings eingehen) und der Akzeptanz von Scoring-Szenarien zu kombinierter Kfz-Krankenversicherung (kombiniertes Versicherungs-Scoring), Telematik-Meldung an die Polizei (Strafzettel-Scoring) sowie einem umfassenden Bürger-Scoring in verschiedenen Lebensbereichen. Als Intervention, die das Erleben eines fiktiven *Social Credit Systems* darstellt, wurde die Episode „Nosedive“ der Netflix-TV-Serie *Black Mirror* gezeigt. In diesem Film werden von den Menschen ständig jede Dienstleistung sowie nahezu jede soziale Interaktion bewertet, jeder hat einen persönlichen Score und kann zugleich den jedes anderen in Echtzeit erfahren.

Die Akzeptanz von neuartigen Scoring-Verfahren, welche Merkmale aus verschiedenen Scoringbereichen kombinieren, war vor und nach der Intervention durch die Filmvorführung für alle Szenarien auf eine Minderheit beschränkt (Tabelle 1, Anhang V), wobei das Szenario „Bürgerscoring“ von 92 % (vorher) bzw. 96 % (nachher) der Teilnehmer abgelehnt wurde, das „Strafzettel-Scoring“ von 31 % (vorher) bzw. 25 % (nachher) befürwortet wurde. Knapp 28 % der Teilnehmer befürworteten mindestens eines der drei Szenarien. Dieser Anteil lag vor der Filmvorführung noch etwas höher; dies war zu erwarten, da die meisten Teilnehmer vor der Filmvorführung über ein solches Szenario vermutlich nicht vertieft nachgedacht haben werden.

Die Intervention, die den Teilnehmern zusätzliches Wissen brachte, hat die erwarteten Ergebnisse gezeigt, nämlich mehr Skepsis bei mehr Wissen. Freilich sind die Effekte so klein, dass man davon ausgehen kann, dass die Ergebnisse einer Repräsentativ-Befragung ohne Intervention aussagekräftiger sind. Eine detailliertere Darstellung dieser Studie kann Anhang V entnommen werden.

II. Repräsentativbefragung 2018

Um repräsentative Ergebnisse zu erzielen, wurde im Auftrag des SVRV vom Erhebungsinstitut infas eine Telefonbefragung (so genanntes *computer-assisted telephone interview*, CATI) im Zeitraum von Anfang Februar bis Ende April 2018 mit einer durchschnittlichen Dauer von 22,5 Minuten durchgeführt. An der repräsentativen Bevölkerungsbefragung nahmen 2.215 Personen (1.123 Männer, 1.092 Frauen) im Alter zwischen 16 und 94 Jahren teil. Da die Teilnahmebereitschaft in verschiedenen Bevölkerungsgruppen unterschiedlich hoch war, wurden diese Unterschiede bei der Auswertung durch eine Gewichtung aus den Ergebnissen herausgerechnet (infas, 2018). Alle im Folgenden angegebenen Prozentzahlen beziehen sich auf die gewichteten Ergebnisse (die absoluten Fallzahlen sind nicht gewichtet). Aufgrund von Rundungen addieren sich die Ergebnisse in einigen Fällen nicht immer exakt zu 100% auf.

Erfragt wurden Wissen und Bewertungen zu den drei Themenkomplexen des vorliegenden Gutachtens: Bonität, Kfz-Haftpflichtversicherung und Gesundheit. Im Weiteren wurde nach der Akzeptanz bezüglich der Verknüpfung von Scores aus verschiedenen Bereichen gefragt. Hierbei sollten die Befragten angeben, ob sie einen Kombinationstarif nutzen würden, der zu einer Kostenersparnis bei der Kfz-Haftpflichtversicherung führt, sofern sie dem Unternehmen auch Daten über das Gesundheitsverhalten zur Verfügung stellten und umgekehrt. Zudem wurde die Akzeptanz eines umfassenden sozialen Bürger-Scorings analog zum chinesischen Sozialkreditsystem erfragt. Um die Bedeutung des „Framings“, mit dem ein Scoring beispielsweise im Zusammenhang mit einem verhaltensbasierten Versicherungstarif beworben wird, empirisch einschätzen zu können, wurden jeweils der Hälfte der Befragten zur Kfz-Haftpflichtversicherung und Krankenversicherung die Folgen des Scorings als Bonus (= geringere Versicherungsbeiträge bei „guten“ Scores) bzw. als Malus (= höhere Versicherungsbeiträge bei „schlechten“ Scores) präsentiert.

Der Fragenkomplex zum Thema Kfz endete mit der Präsentation eines fiktiven Szenarios, bei dem eine allgemeine Aufzeichnungspflicht der Fahrgeschwindigkeit in allen Autos einschließlich einer obligatorischen Übermittlung von Geschwindigkeitsübertretungen direkt an die Polizei postuliert wurde. Im Fragenkomplex zum Thema Gesundheit wurde abschließend ein fiktives Szenario, in dem die Aufzeichnungspflicht der täglichen Schrittzahl und ggf. die Beteiligung an Arztkosten bei zu wenigen Schritten pro Tag postuliert wurde, präsentiert.

Die Struktur und der Inhalt des Fragebogens, anhand dessen die Telefonbefragung durchgeführt wurde, sind in Anhang V ausführlich beschrieben. Der Fragebogen inklusive des genauen Wortlauts der Fragen anhand dessen die Telefonbefragung durchgeführt wurde, kann infas (2018) entnommen werden. Zudem sind die Ergebnisse im Anhang V im Detail deskriptiv dargestellt. In den folgenden Abschnitten erfolgt eine stärker analytische Präsentation der Ergebnisse.

1. Analyse der Ergebnisse

1.1 Scoring: Was wissen die Menschen?

Von den meisten etablierten Auskunfteien werden u. a. die Merkmale Zahlungsverhalten in der Vergangenheit, Alter und Kreditvertrag (laufend) für die Beurteilung der Bonität herangezogen. Hingegen werden die Merkmale Vermögen, Verhalten in sozialen Netzwerken, Beruf und Nationalität nicht dafür verwendet.

In Tabelle D.1 dargestellt sind die Antworten der Befragten auf die Frage, ob etablierte Auskunfteien das genannte Merkmal zur Berechnung des Bonität-Scores heranziehen oder nicht.

Es wird deutlich, dass das Wissen über die Verwendung von Merkmalen für das Bonitäts-Scoring mäßig ausgeprägt ist. Interessanterweise vermutet die Mehrheit der Verbraucher fälschlicherweise, dass auch das Vermögen, der Beruf sowie die Nationalität für das Bonitäts-Scoring berücksichtigt werden. Grundsätzlich ist festzustellen, dass Verbraucher die Verwendung von Merkmalen überschätzen, denn es wird mehrheitlich vermutet, dass sechs der sieben präsentierten Merkmale anstatt der drei tatsächlichen Verwendung finden.

i

Zwischenfazit

Über alle Befragten hinweg ist das Wissen über die Verwendung von Merkmalen zur Beurteilung der Bonität durch etablierte Auskunfteien mäßig ausgeprägt.

Tabelle D.1

Wissen über die Verwendung von Bonitäts-Merkmalen. Verteilung der Antworten (in %) auf die Frage, ob das Merkmal für die Beurteilung der Bonität von Auskunfteien verwendet wird („ja“) oder nicht („nein“). Grün bedeutet, dass die Verwendung bzw. Nicht-Verwendung eines Merkmals mehrheitlich korrekt identifiziert wurde; rot bedeutet, dass die Verwendung bzw. Nicht-Verwendung eines Merkmals mehrheitlich nicht korrekt identifiziert wurde. n = 2.215. (Quelle: infas, 2018).			
	JA	NEIN	VERWEIGERT / WEISS NICHT
Zahlungsverhalten	78	16	6
Alter	71	23	7
Kreditvertrag	82	12	6
Vermögen	64	27	8
Soziale Netzwerke	30	57	13
Beruf	66	26	8
Nationalität	51	40	9

Tabelle D.2

Frequenz der Score-Übermittlung. (SCHUFA: tatsächlicher Abruf in %; Wunsch nach Rückmeldung des Scores in %).				
	JA		NEIN	VERWEIGERT / WEISS NICHT
Abruf SCHUFA-Score in letzten 5 Jahren	21		79	0
Akzeptanz bei Kfz-Score	44		55*	1
	täglich / wöchentlich / monatlich	nur bei starken Änderungen		
	27	17		
Akzeptanz Gesundheits-Score	82		12	6
	täglich / wöchentlich / monatlich	nur bei starken Änderungen		
	28	18		

*Die Antwortkategorie hier lautete „gar nicht“. (Quelle: infas, 2018).

Für eine tiefergehende Analyse stellt sich die Frage, ob das Wissen darüber, welche Merkmale von etablierten Auskunfteien für das Bonitäts-Scoring herangezogen werden und welche nicht, mit bestimmten Faktoren wie beispielsweise dem sozioökonomischen Status variiert. So ist zum Beispiel denkbar, dass Befragte mit einer überdurchschnittlichen formalen Bildung auch überdurchschnittlich viel über verwendete Verbrauchermerkmale wissen. Zudem ist es möglich, dass Verbraucher einer bestimmten Alterskohorte über mehr Scoring-Wissen verfügen, da sie in bestimmten Lebenssituationen stärker mit den Auswirkungen einer Kreditwürdigkeitsprüfung konfrontiert sind, wie z. B. bei der Aufnahme eines Immobilienkredits, welche in der Regel im Alter zwischen 30 und 50 Jahren⁸⁰ erfolgt. Unter diesen und weiteren potenziellen Einflussfaktoren (Monatseinkommen, Geschlecht, Größe des Wohnorts, Erfahrungswissen durch Abruf des SCHUFA-Scores in den vergangenen 5 Jahren) sollen jene Faktoren iden-

tifiziert werden, die im Zusammenhang mit dem Ausmaß des scoringbezogenen Wissens stehen. Untersucht werden soll demnach mittels Regressionsanalyse (siehe D.II.2) über alle in Tabelle D.1 angegebenen Merkmale hinweg, von welchen Faktoren die Wahrscheinlichkeit abhängt, dass die Befragten korrekterweise angeben, ob ein Merkmal tatsächlich bzw. tatsächlich nicht von Auskunfteien zur Berechnung des Bonitäts-Scores verwendet wird.

80 Abgerufen am 7. August 2018 von URL <https://www.sparkasse.de/themen/eigenheim-finanzieren/das-richtige-alter-beim-immobilienkauf.html>

Die gängige Praxis der Übermittlung des aktuellen Scorewerts an Verbraucher reicht von gar nicht bzw. nur auf Anfrage bis hin zu der Möglichkeit, den eigenen Score jederzeit z. B. in einer App einzusehen, wie dies bei Anbietern von Kfz-Telematiktarifen der Fall ist. Vor dem Hintergrund dieser breiten Streuung der Häufigkeit der Rückmeldung durch die Übermittlung von Scorewerten stellt sich die Frage, wie oft Verbraucher üblicherweise ihren Score abrufen (im Falle des SCHUFA-Scores) bzw. welche Häufigkeit der Rückmeldung Verbraucher bei neuartigen Scoring-Verfahren wie Telematik- bzw. Gesundheits-Scores begrüßen würden.

Tabelle D.2 kann entnommen werden, dass nach eigenen Angaben nur etwa ein Fünftel der Befragten in den letzten fünf Jahren den eigenen SCHUFA-Score im Rahmen einer Selbstauskunft erfahren hat. Befragt nach der Akzeptanz der Häufigkeit der Rückmeldung des Scores im Bereich Kfz- und Gesundheits-Telematik gibt etwa die Hälfte der Befragten an, ihren Score gar nicht erfahren zu wollen. Von denjenigen Befragten, die grundsätzlich eine Übermittlung des Scorewerts befürworten, spricht sich etwa ein Drittel dafür aus, nur über starke Veränderungen ihres Scores informiert werden zu wollen; zwei Drittel der Befürworter einer Score-Übermittlung sprechen sich dafür aus, mindestens monatlich (zum Teil wöchentlich oder sogar täglich) über ihren Score informiert zu werden – unabhängig davon, ob sich der Score stark verändert hat oder nicht.

Fazit

Die Befragten sind mehrheitlich dagegen, eine Rückmeldung zu ihrem Score zu erhalten. Knapp die Hälfte der Befragten möchte jedoch grundsätzlich über den Score mit unterschiedlicher Frequenz informiert werden. Denkbar wäre die gängige Praxis (über eine opt-in Option) dahingehend anzupassen, die automatische Übermittlung eines Scores immer dann zu veranlassen, wenn sich dieser stark verändert hat und für den Verbraucher mit (potenziellen) Konsequenzen einhergeht, beispielsweise bei einer qualitativen Änderung des Scores, die mit dem (drohenden) Abstieg von einer Kategorie in die nächst darunterliegende einhergeht.

1.2 Schutz persönlicher Merkmale

Für die Berechnung von Scores wird in der Regel auf persönliche Verbraucherdaten zurückgegriffen wie z. B. Zahlungsverhalten, Vitaldaten und Bewegungsdaten. Die Befragungsteilnehmer wurden sowohl für die Bereiche Kfz- als auch Gesundheits-Telematik nach ihrer

Akzeptanz der Erhebung und Auswertung persönlicher Merkmale befragt. Die daraus resultierenden Akzeptanz- bzw. Ablehnungsraten sind in Tabelle D.3 zusammenfassend dargestellt, wobei nach mehrheitlicher Ablehnung (< 50 % der Befragten stimmen zu) bzw. Akzeptanz (> 50 % der Befragten stimmen zu) aufsteigend sortiert ist.

Tabelle D.3

Akzeptanz bzw. Ablehnung der Erfassung von Merkmalen für einen Kfz- bzw. Gesundheits-Telematiktarif (in %; Rangfolge: niedrig bis hoch).*				
	MERKMAL	AKZEPTANZ	ABLEHNUNG	VERWEIGERT / WEISS NICHT
		(„voll und ganz“ bzw. „eher gerechtfertigt“)	(„eher nicht“ bzw. „überhaupt nicht gerechtfertigt“)	
mehrheitliche Ablehnung (< 50%)	Schlafen ⁺	14	85	1
	Tag-/Nachtfahrt [#]	17	79	3
	Gehen ⁺	26	73	2
	Land-/Stadtfahrt [#]	30	69	1
	Gewicht ⁺	35	65	0
	Alkohol ⁺	43	56	1
	Beschleunigung [#]	45	53	3
mehrheitliche Akzeptanz (> 50%)	Rauchen ⁺	58	41	0
	Geschwindigkeit [#]	62	37	1
	Krebsfrüherkennung ⁺	63	37	0
	Handynutzung zum Lesen oder Schreiben von Nachrichten [#]	66	33	1

*Angegeben sind die Befragungsergebnisse für ein Bonus-Framing, in dem das Einhalten von Vorgaben belohnt wird. #: Merkmale für den Kfz-Telematiktarif; +: Merkmale für den Gesundheits-Telematiktarif (Quelle: infas, 2018; eigene Berechnungen).

Die Erhebung und Verwendung potenziell sensibler Merkmale zur Bildung von Scores wird für die Mehrheit der abgefragten Merkmale im Bereich Gesundheit und Kfz unter dem Bonus-Framing mehrheitlich abgelehnt. Lediglich die Merkmale Rauchen und Krebsfrüherkennung (jeweils bei Gesundheits-Telematik) sowie Geschwindigkeit und Handynutzung (jeweils bei Kfz-Telematik) finden eine Mehrheit, die der Verwendung dieser Merkmale für die Tarifierung zustimmen würde.



Zwischenfazit

Insgesamt ist die Akzeptanz hinsichtlich der Merkmale, die für das präsentierte Kfz- bzw. Gesundheits-Scoring herangezogen werden sollen, niedrig. Dieses deutet darauf hin, dass die Mehrheit der Befragten bei den meisten der präsentierten verhaltensbasierten Merkmalen dagegen ist, dass diese bei der Ausgestaltung von Versicherungstarifen berücksichtigt werden.

Mittels Regressionsanalyse wird geprüft werden, welche Faktoren die Akzeptanz für die Nutzung persönlicher Merkmale zum Zwecke des Scorings beeinflussen (siehe D.II.2). Die Leithypothese ist hier, dass das Ausmaß der Akzeptanz für die Verwendung eines bestimmten Merkmals zur Tarifierung auch von der persönlichen Betroffenheit mit Bezug auf dieses Merkmal abhängt. Wird beispielsweise die Fahrt auf dem Land bonifiziert, während der Score bei Stadtfahren sinkt (weil mit Landfahrten ein geringeres Unfallrisiko assoziiert wird als mit Fahrten in der Stadt), liegt es nahe, dass Befragte, die in der Stadt wohnen, eine geringere Akzeptanz mit Blick auf dieses Merkmal zeigen. Zudem erscheint es plausibel, dass Faktoren wie der Gesundheitszustand (z. B. könnten Personen mit einem schlechteren allgemeinen Gesundheitszustand verhaltensbasierte Tarife tendenziell eher ablehnen), die Art der Fortbewegung und der Wohnort im Zusammenhang mit der Akzeptanz bzw. Ablehnung von Merkmalen, die für die Tarifierung herangezogen werden, aufweisen.

1.3 Akzeptanz neuer und potenziell möglicher Entwicklungen auf dem Scoring-Markt

1.3.1 Würden Verbraucher telematikbasierte Kfz- und Gesundheits-Scores nutzen?

Die Befragten lehnen neuartige und in Zukunft potenziell mögliche Scoring-Verfahren im Bereich Kfz und Gesundheit mehrheitlich ab (etwa zu zwei Dritteln). Interessant ist hierbei, dass dies für die beiden Bereiche Kfz und Gesundheit gleichermaßen gilt und die Effekte des Bonus-Malus-Framings wie erwartet evident, aber vor allem im Gesundheits-Bereich betragsmäßig sehr klein sind (siehe Tabelle D.4). Im Umkehrschluss bedeutet dies, dass sich etwa ein Drittel der Befragten die Nutzung eines verhaltensbasierten Tarifs verbunden mit Scoring sowohl im Bereich Kfz als auch im sensiblen Bereich der Gesundheit vorstellen kann. Letzteres Ergebnis ist insofern interessant, als die tatsächliche Nutzerzahl

verhaltensbasierter Tarife in der Kfz-Haftpflichtversicherung deutlich unter einem Drittel der Versicherten liegt (siehe Marktstudie, Kapitel C) und die Befragten sich deutlich skeptisch gegenüber der Erhebung potenziell sensibler Merkmale zeigen (siehe D.II.1.2), ohne welche die Implementierung eines derartigen Tarifsystems allerdings gar nicht möglich wäre. Dieses Paradoxon deutet darauf hin, dass die Intention etwas tun zu wollen, nicht zwangsläufig dazu führt, dieses auch in eine Handlung zu übersetzen. Zum anderen könnte die Diskrepanz, dass die Befragten einerseits eine hohe Skepsis bzgl. der Erhebung und Verwendung von Merkmalen für einen Telematiktarif aufweisen, andererseits jedoch ca. ein Drittel und mehr sich vorstellen können, einen Telematiktarif zu nutzen (um ggf. potenziell zu sparen), auch mit dem Datenschutz-Paradoxon erklärt werden.

Tabelle D.4

Akzeptanz neuartiger Scores (in %). Die Befragten sollten angeben, ob sie einen telematikbasierten Tarif nutzen würden oder nicht. Die Antworten sind für die Gesamtstichprobe im Bereich Kfz bzw. Gesundheit und unterteilt nach Bonus bzw. Malus-Framing dargestellt.							
	JA			NEIN			VERWEIGERT / WEISS NICHT*
	gesamt	Bonus	Malus	gesamt	Bonus	Malus	
Kfz	36	43	30	63	57	69	1
Gesundheit	34	36	32	65	64	67	1

Kfz: gesamt: n = 1.104, Bonus: n = 536, Malus n = 568; Gesundheit: gesamt: n = 1.111, Bonus: n = 570, Malus: n = 541 (Quelle: infas, 2018).

Beispiele für Argumente gegen die Nutzung eines telematikbasierten Kfz-Haftpflichtversicherungstarifs



„Weil sehr viele persönliche Daten erhoben werden würden, die ich nicht preisgeben will.“

(männlicher Befragter, 40 J.)

„Ich empfinde das als Kontrolle, die zu stark in meinen privaten Bereich eingreift.“

(weibliche Befragte, 60 J.)

„Weil das mir zu weit geht mit dem Schnüffeln. [...] Ich erwarte Solidarität von meiner Kfz-Versicherung. Ich würde sogar mehr bezahlen, um nicht beschnüffelt zu werden.“

(männlicher Befragter, 61 J.)

Beispiele für Argumente für die Nutzung eines telematikbasierten Kfz-Haftpflichtversicherungstarifs



„Es geht letzten Endes um Gerechtigkeit. Jeder hat die gleiche Chance auf alles. Wer sich an die Regeln hält, sollte dadurch auch Vorteile genießen.“

(männlicher Befragter, 33 J.)

„Weil ich dadurch Vorteile habe, und mehr zu meiner Sicherheit beitrage und andere dadurch schütze.“

(weibliche Befragte, 27 J.)

„Würde es nur nutzen, weil ich profitieren würde und eine vorsichtige Fahrerin bin.“

(weibliche Befragte, 35 J.)

Beispiele für Argumente gegen die Nutzung eines telematikbasierten Krankenversicherungstarifs



„Weil ich keine Lust habe, mich von der Krankenkasse überwachen zu lassen. Es ist meine eigene Privatsphäre und es geht niemanden was an was ich, wie mache. Ich finde es eine Benachteiligung für chronisch Kranke oder ältere Menschen.“

(weibliche Befragte, 57 J.)

„Ich halte zum Solidaritätsprinzip. Viele haben aufgrund von beruflichen Verpflichtungen oder wegen Krankheit einfach keine Möglichkeit dazu. In dem Fall wäre das auch ziemlich unfair.“

(männlicher Befragter, 53 J.)

„Ich möchte die Werte nicht aufzeichnen und auch nicht, dass die von jemandem erfasst werden. Und alle sollten einen ähnlichen Tarif unabhängig von der persönlichen Konstitution haben.“

(weibliche Befragte, 63 J.)

Beispiele für Argumente für die Nutzung eines telematikbasierten Krankenversicherungstarifs



„Das Solidaritätsprinzip der Krankenversicherung für alle Versicherten sollte erhalten bleiben, selbstverschuldetes Fehlverhalten dagegen (wie z. B. Alkoholkonsum) sollte sanktioniert werden.“

(männlicher Befragter, 66 J.)

„Wer für seinen Körper nichts tut, der sollte auch dafür zahlen und wenn man alles berücksichtigt, dann sollte man einen günstigeren Tarif bekommen (allerdings immer dem Alter angemessen).“

(weibliche Befragte, 82 J.)

„Weil die Vorteile sind, wenn man etwas für seine Gesundheit tut, hat man auch bei der Krankenversicherung Vorteile. Das ist so eine Win-Win-Situation.“

(männlicher Befragter, 29 J.)

Mittels Regressionsanalyse wird untersucht werden (siehe D.II.2), von welchen Faktoren die Akzeptanz von Kfz- bzw. Gesundheits-Scores abhängt. Auch hier liegt die Vermutung nahe, dass insbesondere solche befragten Personen eine verhaltensbasierte Versicherung gutheißen, die davon profitieren würden (z.B. besonders gesunde oder sportliche Personen) und daher Einsparungen durch den Abschluss einer verhaltensbasierten Versicherung erwarten könnten.

1.3.2 Wie stehen die Menschen zu einer Veröffentlichung von Scores?

Unterschiedliche Kfz-Versicherungen werben damit, dass Scorewerte auf ihren Websites veröffentlicht werden, um besonders gute Autofahrer herauszustellen. Der Veröffentlichung von Scorewerten kommt insbesondere dann eine besondere Bedeutung zu, wenn der Wettbewerb zwischen Verbrauchern zur Steigerung der eigenen Performance verstärkt werden soll.

Eine freiwillige bzw. verpflichtende Veröffentlichung von Scores wird für die Bereiche Bonität, Kfz und Gesundheit von der deutlichen Mehrheit abgelehnt. Für den sensiblen Bereich Gesundheit ist die Ablehnungsrate erwartungsgemäß am höchsten. Die Ablehnungsraten werden vom Bonus-Malus-Framing nicht nennenswert beeinflusst. Die Ergebnisse sind in Tabelle D.5 zusammenfassend dargestellt.

i

FAZIT

Die Befragten lehnen eine Veröffentlichung von Scores mit großer Mehrheit ab. Dies gilt sowohl für eine Veröffentlichung auf freiwilliger Basis als auch für eine allgemeine Pflicht zur Veröffentlichung.

Tabelle D.5

Bereitschaft zur freiwilligen bzw. verpflichtenden (in Klammern) Veröffentlichung von Scores (in %).							
	JA			NEIN			VERWEIGERT / WEISS NICHT*
	gesamt	Bonus	Malus	gesamt	Bonus	Malus	
Bonität	12 (7)	--	--	87 (92)	--	--	1 (1)
Kfz	12 (10)	13 (13)	11 (8)	87 (90)	86 (87)	88 (92)	1 (0)
Gesundheit	5 (5)	4 (6)	7 (5)	95 (94)	96 (94)	93 (95)	0 (0)

Bonität: gesamt: n = 2.215; Kfz: gesamt: n = 1.104, Bonus: n = 536, Malus n = 568; Gesundheit: gesamt: n = 1.111, Bonus: n = 570, Malus: n = 541 (Quelle: ifas, 2018).

1.3.3 Wie stehen die Menschen zu Super-Scores und einer scorebasierten Sanktionierung?

Die Umfrage-Teilnehmer wurden auch nach ihrer Akzeptanz von Super-Scores (d. h. der Verknüpfung von Scores aus verschiedenen Bereichen) und ggf. scorebasierten Sanktionen in hypothetischen Szenarien befragt. Die Szenarien umfassten: 1) Die Verknüpfung zweier Scores (Gesundheit und Kfz), 2) die automatische Sanktionierung von potenziellem Fehlverhalten mit Anbindung an „höhere Autoritäten“ (Polizei, Arzt) sowie 3) eine stilisierte Version eines Sozialkreditsystems, das noch weitere Verknüpfungen zwischen verschiedenen Lebensbereichen herstellt als in 1) geschildert.

Der Verknüpfung von Daten aus den Bereichen Kfz und Gesundheit stimmen knapp ein Viertel bis knapp ein Drittel der Befragten zu. Auch die automatische Über-

mittlung von Geschwindigkeitsüberschreitungen an die Polizei findet bei gut einem Viertel der Befragten Zustimmung. Eine deutliche Abnahme der Akzeptanzrate findet sich bei der Aufzeichnung der Schrittzahl und einer damit assoziierten Beteiligung an Arztkosten bei Nichteinhaltung der Vorgaben und beim Sozialkreditsystem. Dieses könnte daran liegen, dass die Schrittzahl einen Bereich tangiert, der im weitesten Sinne dem sensiblen Bereich Gesundheit zuzuordnen ist und dass Personen deutlich ablehnender auf Scoring reagieren, wenn ihnen die potenziell negativen (monetären) Konsequenzen von Scoring (Beteiligung an den Arztkosten) bewusst gemacht werden. Zum anderen scheint eine allumfassende Erhebung und Verknüpfung von Daten im Sinne des Sozialkreditsystems sehr unerwünscht. Die Ergebnisse sind zusammenfassend in Tabelle D.6 dargestellt.

Tabelle D.6

Akzeptanz von Super-Scores und scorebasierten Sanktionen in der Gesamtgruppe (n = 2.215) der Befragten (in %).				
		JA	NEIN	VERWEIGERT / WEISS NICHT
Score-Verknüpfung	Kombitarif: Kfz+Gesundheit*	24	76	0
	Kombitarif: Gesundheit+Kfz**	31	68	1
Sanktionierung von Verhalten	Geschwindigkeitsüberschreitung: Meldung an Polizei	28	72	0
	Mangelnde Bewegung: Beteiligung an Arztkosten	8	91	0
Sozialkreditsystem	Fahrweise, Gesundheit, Rechnungen bezahlen, respektvolles Verhalten	9	91	0

*Die Befragten wurden in zwei Gruppen geteilt. Dieser Gruppe wurde zunächst ein Bonus-System im Kfz-Bereich vorgestellt. Im Anschluss wurden die Befragten danach befragt, ob sie sich vorstellen könnten, den Score mit einem Gesundheits-Score zu verknüpfen, um eine Reduktion des Tarifs zu erhalten. **Vice versa. In die Kategorie „ja“ fallen die Antworten „ja, mit Sicherheit“ sowie „ja, möglicherweise“; in die Kategorie „nein“ fallen die Antworten „nein, eher nicht“ sowie „nein, sicher nicht“. (Quelle: ifas, 2018).

Tabelle D.7

Akzeptanz von Super-Scores und scorebasierten Sanktionen nach Bonus- bzw. Malus-Framing (in %).				
		JA	NEIN	VERWEIGERT / WEISS NICHT
Score-Verknüpfung	Kombitarif: Kfz+Gesundheit*	K: 24 (24)	K: 76 (76)	K: 0 (0)
	Kombitarif: Gesundheit+Kfz**	G: 32 (30)	G: 68 (69)	G: 0 (1)
Sanktionierung von Verhalten	Geschwindigkeitsüberschreitung: Meldung an Polizei	K: 31 (24)	K: 69 (76)	K: 0 (0)
	Mangelnde Bewegung:	G: 6 (10)	G: 93 (89)	G: 0 (0)
	Beteiligung an Arztkosten			
Sozialkreditsystem	Fahrweise, Gesundheit, Rechnungen bezahlen, respektvolles Verhalten	K: 10 (10)	K: 90 (90)	K: 0 (0)
		G: 7 (7)	G: 92 (92)	G: 0 (0)

Die Zahlen ohne Klammern beziehen sich auf das Bonus- die in Klammern auf das Malus-Framing. In die Kategorie „ja“ fallen die Antworten „ja, mit Sicherheit“ sowie „ja, möglicherweise“; in die Kategorie „nein“ fallen die Antworten „nein, eher nicht“ sowie „nein, sicher nicht“. K = Befragte aus dem Split Kfz; G: Befragte aus dem Split Gesundheit.

Beispiele für Argumente gegen die Nutzung eines Kombitarifs (Kfz und Gesundheit)



„Das eine darf bei einer Versicherungsfrage nichts mit dem anderen zu tun haben; getrennte Betrachtung der Risikofaktoren bei Versicherungen.“

(männlicher Befragter, 56 J.)

„Das ist etwas unheimlich, wenn eine Versicherung all meine Gesundheitsdaten bekommt. Man weiß im Endeffekt nicht, wo sie wirklich landen.“

(weibliche Befragte, 36 J.)

„Die gesamte Aufzeichnung von Daten und persönlichem Leben lehne ich völlig ab.“

(männlicher Befragter, 63 J.)

Beispiele für Argumente für die Nutzung eines Kombitarifs (Kfz und Gesundheit)



„Ich bin immer für innovative Produkte und Dienstleistungen zu haben, wenn sie einen gewissen Mehrwert bieten. Hier werden Synergieeffekte genutzt, die eventuell ein großes Potenzial haben können. So voll und ganz bin ich mir da jedoch nicht sicher, ob das nicht doch zu weit geht.“

(männlicher Befragter, 50 J.)

„Weil ich das schon gut finde, wenn jeder bewusster am Straßenverkehr teilnimmt und bewusster lebt.“

(weibliche Befragte, 52 J.)

„Man hat dadurch Vorteile wie z. B. eventuelles Geld sparen. Für mich persönlich wäre das gut, weil ich die Voraussetzungen dafür erfülle.“

(weibliche Befragte, 56 J.)

Die Unterschiede in den Akzeptanzraten zu Super-Scores und scorebasierten Sanktionen nach Bonus- bzw. Malus-Framing sind sowohl im Vergleich zur Gesamtgruppe als auch im Vergleich zwischen Bonus und Malus allenfalls als marginal einzustufen (siehe Tabelle D.7).

In einer tiefergehenden Analyse wird mittels Regressionsanalyse untersucht, mit welchen Faktoren die Akzeptanz bzw. Ablehnung von Super-Scores und scorebasierten Sanktionen zusammenhängt (siehe D.II.2).

i

FAZIT

Die Befragten lehnen eine Verknüpfung von Scores mit teils sehr großer Mehrheit ab. Insbesondere der auf einem Score basierenden monetären Sanktionierung im Bereich des Gesundheitsverhaltens und der Verknüpfung von Scores, die den gesamten Lebensbereich umfassen, wird mit großer Ablehnung begegnet.

1.4 Die Bedeutung des Bonus-/Malus-Framings

Telematik-Tarife in der Kfz-Haftpflichtversicherung werden heutzutage üblicherweise als Bonus-System angeboten, d. h. als ein System, bei dem Verbraucher durch Nutzung des Tarifs bei gutem Fahrverhalten Vergünstigungen auf den Ausgangstarif erhalten. Dabei ist dieser Gewinn von der Höhe des Ausgangstarifs abhängig – je höher dieser im Vergleich zu einem Tarif ohne Telematik ist, desto schwieriger wird es, durch die Teilnahme am Telematik-Tarifprogramm insgesamt zu sparen. Darüber hinaus ist es denkbar, dass Versicherungen künftig auch explizit Malus-Bausteine in ihr Tarifsysteem einbauen, um Autofahrer noch stärker zu einer sichereren Fahrweise zu motivieren. Das Framing – also die Frage, ob Verbrauchern Scoring im Rahmen von Versicherungstarifen als Gewinn- oder Verlustsystem deklariert wird – ist daher möglicherweise von Bedeutung und nicht nur die Akzeptanz der Merkmalerhebung selbst, sondern die Akzeptanz des Scores insgesamt könnte erheblich dadurch geprägt sein.

Die Ergebnisse der Befragung zeigen, dass der Unterschied zwischen Bonus- und Malus-Framing zum Teil erheblich ausfällt. Es wird deutlich, dass allein aufgrund des Framings Zustimmungsraten höher oder niedriger ausfallen können. Die Differenz zwischen den Zustimmungsraten des Bonus- und Malus-Framings bei Merkmalen, die zur Tarifierung für neuartige Telematiktarife herangezogen werden bzw. potenziell herangezogen werden könnten, ist in Tabelle D.8 dargestellt. Insbesondere beim Merkmal Geschwindigkeit ist die Differenz zwischen Bonus und Malus-Framing besonders groß, während sich die Zustimmungsraten für das Merkmal Rauchen nicht durch das Framing beeinflussen lässt.



Es ist anzunehmen, dass beispielsweise Versicherungen, die Telematiktarife anbieten, solche Tarife eher positiv, also entsprechend des hier verwendeten Bonus-Framings annoncieren. Damit stehen u. U. auch dem Bonussystem innewohnende Verlustpotenziale bzw. die Gefahr des Missbrauchs im Zusammenhang mit der Datensammlung weniger im Bewusstsein der Verbraucher. Tatsächlich kann man durch Verhalten, das nicht konform mit den jeweiligen Vorgaben ist, jederzeit auch Punkte verlieren, so dass der Bonus abnimmt. Zudem ist derzeit unklar, wie sich ungescorete Basistarife langfristig entwickeln werden und ob Malus-Bausteine künftig Einzug in Telematik-Programme erhalten werden.

FAZIT

Die Akzeptanzraten dafür, ob ein Merkmal für die Tarifierung bei einer Kfz-Haftpflicht- bzw. Krankenversicherung als gerechtfertigt bewertet wird, variieren in Abhängigkeit von der Präsentation eines Bonus- oder Malus-Framings.

Tabelle D.8

Unterschiede in der Akzeptanz von Merkmalserfassung und neuartigen Scores nach Bonus- vs. Malus-Framing (Differenz Malus-Bonus in %-Punkten).			
		ÄNDERUNG*	
Merkmalserfassung*	Geschwindigkeit	↓	-31
	Gehen	↓	-18
	Krebsfrüherkennung	↓	-17
	Land-/Stadtfahrt	↓	-15
	Tag-/Nachtfahrt	↓	-14
	Schlafen	↓	-9
	Alkohol	↓	-8
	Gewicht	↓	-6
	Rauchen	↔	0
	Beschleunigung	↑	+4
	Handy	↑	+12
Neuartige und potenziell mögliche Scores**	Kfz	↓	-6
	Gesundheit	↓	-2

*Dargestellt ist die Änderung der Zustimmung für die beiden Antwortkategorien „voll und ganz“ sowie „eher gerechtfertigt“. **Dargestellt ist die Änderung der Zustimmung für die Kategorie „ja“

2. Multivariate Regressionsanalysen: Darstellung und Diskussion der Ergebnisse

Die Darstellung in diesem Abschnitt beschränkt sich auf die wesentlichen Ergebnisse der multivariaten Regressionsanalysen in vereinfachter Form. Diese sollen

aufzeigen, mit welchen Faktoren das Wissen über von Auskunfteien verwendete Verbrauchermerkmale oder die Akzeptanz für neuartige Telematik-Tarife in den Bereichen Kfz und Gesundheit zusammenhängen. Eine ausführliche Darstellung der Regressionsanalysen einschließlich der statistischen Kennwerte finden sich in infas (2018). Die Ergebnisse sind für multiples Testen nach der sequentiellen Bonferroni-Holm-Korrektur adjustiert (Holm, 1979).

i

Einflussfaktoren des Wissens über üblicherweise verwendete Merkmale des Bonitäts-Scorings (Tabelle D.1; Tabelle 2 in infas, 2018)

Tabelle 2 (Anzahl richtig genannter Bonitäts-Merkmale) aus infas (2018): Untersucht wurde mithilfe einer linearen Regression der Einfluss bestimmter sozio-demografischer Variablen sowie Einstellungen der Befragten auf die Anzahl korrekter Antworten zur Erfassung von Verbrauchermerkmalen der Bonitäts-Auskunfteien. Modelliert ist die abhängige Variable als stetige Variable mit der Anzahl richtig genannter Verbrauchermerkmale von 0 (kein Merkmal richtig benannt) bis 7 (alle Merkmale richtig benannt). Aus dem Regressionskoeffizienten – sofern signifikant – lässt sich ablesen, wie viele zusätzliche Antworten bestimmte Personengruppen im Verhältnis zu einer Referenzgruppe im Schnitt richtig beantworten.

Beispiel: Der Regressionskoeffizient der Variablen „Haushaltsäquivalenzeinkommen: unterdurchschnittlich“ i. H. v. $-0,15$ besagt, dass Befragte mit einem unterdurchschnittlichen Haushaltsäquivalenzeinkommen im Schnitt $0,15$ weniger richtige Antworten geben als Befragte mit einem durchschnittlichen Haushaltsäquivalenzeinkommen.

Für alle weiteren Analysen (logistischen Regressionen) aus den Tabellen 3–8 (infas, 2018) gilt: Der Abstand des Regressionskoeffizienten von 1 bringt die Stärke des Effekts zum Ausdruck. Ausgedrückt in Prozent bedeutet dies, dass beispielsweise ein Koeffizient von $1,78$ zum Ausdruck bringt, dass die Chance (Odds) des Eintretens eines Ereignisses um 78% größer relativ zur Referenzgruppe ist. Ein Regressionskoeffizient <1 bedeutet hingegen, dass die Wahrscheinlichkeit einer richtigen Antwort in dieser Teilgruppe geringer ist als in der Referenzgruppe.

Beispiel: Der Regressionskoeffizient der Variablen „Alter: 16-34 Jahre“ i. H. v. $1,8$ (Tabelle 1, infas, 2018) besagt – sofern signifikant – in Prozentwerten ausgedrückt, dass Personen im Alter von 16-34 Jahren im Vergleich zur Referenzgruppe (65 Jahre und älter) eine um 80% höhere Wahrscheinlichkeit haben die Frage, ob das endgültige Nicht-Bezahlungen von Rechnungen von vielen Auskunfteien als Merkmal erfasst wird, richtig zu beantworten.

2.1 Wissen bezüglich der Merkmalsverwendung beim Bonitäts-Scoring

Das Ausmaß des Wissens im Zusammenhang mit dem Bonitäts-Scoring ist vor allem mit dem Alter und dem Abruf des SCHUFA-Werts innerhalb der letzten fünf Jahre assoziiert: So ist die Summe der korrekten Antworten bei Befragten, die in den vergangenen fünf Jahren ihren SCHUFA-Wert erfragt haben, im Schnitt geringfügig höher als bei Personen, die den SCHUFA-Wert in diesem Zeitraum nicht erfragt haben. Zudem geben Befragte in der Altersklasse von 16 bis 64 Jahren im Schnitt etwa eine halbe korrekte Antwort mehr als Personen über 65 Jahre. Auf der Ebene der Einzelmerkmale, die beim Bonitäts-Scoring herangezogen werden, zeigt sich, dass die Wahrscheinlichkeit für eine richtige Antwort bei den Merkmalen laufender Kreditvertrag, Daten aus sozialen Netzwerken und Nationalität bei jüngeren Befragten (16-34 Jahre bzw. 35-49 Jahre) höher ist als bei älteren Befragten (65 Jahre und älter). Zudem liegt bei Personen, die innerhalb der letzten fünf Jahre ihren SCHUFA-Wert abgerufen haben, die Wahrscheinlichkeit eine richtige Antwort auf die Frage, ob Daten aus sozialen Netzwerken Verwendung bei der Bonitäts-Beurteilung finden, höher als bei Personen, die in diesem Zeitraum den SCHUFA-Wert nicht abgerufen haben. Bei jüngeren Befragten (16-34 Jahre) ist die Wahrscheinlichkeit einer richtigen Antwort auf die Frage, ob das Merkmal Nationalität Verwendung beim Bonitäts-Scoring findet, höher als bei älteren Befragten (65 Jahre und älter).

Insgesamt deuten die Ergebnisse darauf hin, dass Befragte insbesondere dann über die verwendeten bzw. nicht verwendeten Merkmale zur Beurteilung der Bonität Bescheid wissen, wenn eine Bonitäts-Auskunft aktiv eingeholt wurde bzw. wenn sie zu der Altersklasse gehören, die sich beispielsweise am ehesten mit der Aufnahme eines Immobilienkredits beschäftigt und sich in diesem Zusammenhang mit ihrem Bonitäts-Score befassen muss. Vor dem Hintergrund, dass kein systematischer Zusammenhang zwischen allgemeiner formaler Bildung und korrektem Antwortverhalten in Bezug auf die Merkmalsverwendung von Auskunfteien beobachtet werden konnte, ist diesbezügliches Wissen offenbar nicht hinreichend durch formale Bildung adressiert. So erscheint eine Scoring-spezifische Wissensvermittlung (Stichwort: Förderung von Scoring-Kompetenz) sowohl im Rahmen der formalen Bildung als auch auf anderen Weg sinnvoll zu sein.

2.2 Akzeptanz der Merkmalserfassung bei Kfz-Telematik (Tabelle D.3 und D.4; Tabelle 3 und Tabelle 4 in infas, 2018)

Das Alter der Befragten sowie die Größe des Wohnorts (Einwohnerzahl) spielen bei der Akzeptanz des Merkmals Fahren auf dem Land vor dem Hintergrund eines in Aussicht gestellten Bonus im Sinne einer Ersparnis beim Versicherungstarif eine Rolle: Bei älteren Personen (65 Jahre und älter) ist die Wahrscheinlichkeit für die Akzeptanz dieses Merkmals höher als bei Personen mittleren Alters (35-49 Jahre). Zudem ist die Wahrscheinlichkeit bei Personen, die in einer kleineren Stadt leben (20.000 bis 100.000 Einwohnern), für die Akzeptanz des Merkmals Fahren auf dem Land für die Tarifierung im Schnitt höher als bei Personen, die in einer Großstadt leben (mehr als 100.000 Einwohner). Hier deutet sich ein „Mitnahmeeffekt“ bei derjenigen Personengruppe an, die in einer kleineren Stadt (Wohnortsgröße bis 100.000 Einwohner) wohnt: Je kleiner die Stadt ist, in der eine Person wohnt, desto höher ist die Akzeptanz des Merkmals Fahren auf dem Land, für das ein Bonus in Aussicht gestellt wird.

Darüber hinaus zeigen die Ergebnisse, dass die Wahrscheinlichkeit für die Akzeptanz des Merkmals Geschwindigkeit für die Tarifierung bei jüngeren Befragten (16-34 Jahre) im Vergleich zu älteren Befragten (65 Jahre und älter) unter dem Malus-Framing höher ist.

2.3 Potenzielle Nutzung von Telematik-Tarifen und Kombi-Telematik-Tarifen (Tabelle D.6 und Tabelle D.7; Tabelle 7 in infas, 2018)

Für den Bereich der Kfz-Telematik zeigt sich, dass die potenzielle Nutzung eines solchen Tarifsystems ganz wesentlich vom Kontext und den in Aussicht gestellten Konsequenzen abhängt: Personen, die den Kfz-Telematik-Tarif als Bonus-System präsentiert bekommen haben, können sich die Nutzung eines Telematiktarifs eher vorstellen als diejenigen Personen, die es als Malus-System erhielten. Dies bedeutet, dass Kfz-Telematik-Tarife möglicherweise nur dann für einige Personengruppen attraktiv erscheinen, wenn die Möglichkeit einer Preisersparnis besteht. Das Framing des Telematik-Angebots scheint somit eine erhebliche Rolle hinsichtlich der möglichen Nutzung von Scoring-Angeboten in der Bevölkerung zu spielen.

Für den Bereich der Gesundheits-Telematik zeigt sich, dass sich Personen mit einer überdurchschnittlich ausgeprägten internalen Kontrollüberzeugung („Ich habe mein Leben selbst in der Hand“, „Wenn ich mich anstrengte, werde ich auch Erfolg haben“) eher die Nutzung eines Gesundheits-Telematiktarifs vorstellen können als diejenigen mit einer unterdurchschnittlichen internalen Kontrollüberzeugung. Demnach können sich also Personen, die der Meinung sind, ihr Leben selber beeinflussen zu können und dass das eigene Verhalten zum Erfolg führt, eher vorstellen, einen Gesundheits-Telematik-Tarif zu nutzen.

Darüber hinaus können sich Personen mit ausgeprägter Preisorientierung eher die Nutzung eines Kombitarifs vorstellen, das heißt eines Tarifs, bei dem weniger für den Kfz-Versicherungstarif gezahlt werden müsste, wenn dafür auch Daten über das Gesundheitsverhalten der Versicherung zur Verfügung gestellt werden würden als Personen mit geringerer Preisorientierung.

2.4 Akzeptanz von Telematik mit Sanktionierung (Tabelle 8 in infas, 2018)

Die Aufzeichnung der Schrittzahl mit automatischer Sanktionierung (finanzielle Beteiligung an Arztkosten) wird am ehesten von denjenigen Personen befürwortet, die sportlich aktiver (Sport mehrmals pro Woche) sind, verglichen mit Personen, die dieses eher nicht sind (Sport einmal pro Woche oder weniger). Dies deutet darauf hin, dass insbesondere Personen, die für einen bestimmten Bereich (z. B. Bewegung) eher keinen Nachteil für sich selbst zu erwarten haben, einer scoring-basierenden Sanktionierung in diesem Bereich eher zustimmen.

Eine Geschwindigkeitsaufzeichnung mit automatischer Sanktionierung (Übermittlung der Geschwindigkeitsüberschreitung an die Polizei) wird eher von denjenigen Personen befürwortet, die keine Punkte in Flensburg haben im Vergleich mit Personen, bei denen das der Fall ist. Scoring scheint demnach darüber hinaus besonders eher dann akzeptiert zu werden, wenn eine Person in dem Bereich, in dem sie (potenziell) gescort wird, nicht negativ auffällig geworden ist. So könnte man spekulieren, dass Personen, die ohnehin schon eine eher umsichtige Fahrweise aufweisen (Personen, die beispielsweise die vorgegebene Höchstgeschwindigkeit nicht überschreiten und auch nicht stark beschleunigen) und ein Bonus damit wahrscheinlicher ist, einen Kfz-Telematik-Tarif eher akzeptieren als Personen auf die dieses nicht zutrifft. Befürwortet wird dieser Score zudem von denjenigen Personen, die sich hauptsächlich mit dem öffentlichen Personennahverkehr fortbewegen im Vergleich mit Personen, die das Auto als Hauptfortbewegungsmittel nutzen. Demnach wird dieses Scoring eher von Personen befürwortet, die davon und den damit assoziierten Sanktionen vermutlich eher weniger betroffen wären.

Im Allgemeinen gilt somit für die hier berichteten hypothetischen Scores, dass diese eher befürwortet werden, sofern für die Befragten selbst eher keine negativen Konsequenzen dadurch zu erwarten sind.

3. Gesamtzusammenfassung und Fazit der Ergebnisse der Bevölkerungsbefragung

3.1 Zur Transparenz und Verständlichkeit von Scoring

Die Befragten sind mehrheitlich dagegen, eine Rückmeldung zu ihrem Score zu erhalten, die anderen (knapp die Hälfte) möchten grundsätzlich mit unterschiedlicher Häufigkeit darüber informiert werden. Denkbar wäre die gängige Praxis (über eine opt-in Option) dahingehend anzupassen, die automatische Übermittlung eines Scores immer dann zu veranlassen, wenn sich dieser stark verändert hat und für den Verbraucher mit (potenziellen) Konsequenzen einhergeht. Beispielsweise bei einer qualitativen Änderung des Scores, die mit dem (drohenden) Abstieg von einer Kategorie in die nächst darunterliegende einhergeht.

Die Akzeptanzraten dafür, ob ein Merkmal für die Tarifierung bei einer Kfz-Haftpflicht- bzw. Krankenversicherung als gerechtfertigt bewertet wird, variieren auch damit, ob dieses im Zusammenhang mit einem Bonus oder Malus präsentiert wird. Um Verbrauchern eine informierte Entscheidung über die Gewinn- und Verlustmöglichkeiten zu ermöglichen, wäre es wünschenswert, dass durch die Anbieter über die durchschnittliche Bonushöhe informiert und über Verlustmöglichkeiten – sofern relevant – aufgeklärt werden.

3.2 Zum Wissen über Scoring und Scoring-Kompetenz

Über alle Befragten hinweg ist das Wissen über die Verwendung von Merkmalen zur Beurteilung der Bonität durch etablierte Auskunfteien mäßig ausgeprägt.

Die Ergebnisse zeigen im Detail, dass das Wissen über die verwendeten Merkmale im Bereich des Bonitäts-Scorings neben dem Alter davon abhängt, ob eine Person in den vergangenen fünf Jahren eine Bonitäts-Selbstauskunft eingeholt hat. Evident wird, dass die formale Bildung alleine keinen relevanten Einfluss darauf zu haben scheint. Neben dem Fördern von scoring-bezogenem Wissen scheint auch die Vermittlung scoring-spezifischer Kompetenzen sinnvoll. Die Nutzung von (Selbst-) Auskünften erscheint beispielsweise als ein nützliches und vergleichsweise unaufwendiges Mittel hierfür. Dieses Mittel sollte gezielt genutzt werden, um Verbrauchern wichtige Informationen über verwendete Merkmale und ggf. auch deren Gewichtung leicht verständlich zu übermitteln, damit sie informiert am Scoring-Verfahren teilnehmen können.

3.3 Zu Telematik-freien Optionen

Insgesamt ist die Akzeptanz hinsichtlich der Merkmale, die für das präsentierte Kfz- bzw. Gesundheits-Scoring herangezogen werden sollen, niedrig. Dieses deutet darauf hin, dass die Mehrheit der Befragten bei den meisten der präsentierten verhaltensbasierten und situationsabhängigen Merkmalen dagegen ist, dass diese bei der Ausgestaltung von Versicherungstarifen Berücksichtigung finden; möglicherweise nicht zuletzt deshalb, weil diese teils sensible Bereiche wie die Gesundheit betreffen.

Im Detail zeigen die Ergebnisse, dass die Akzeptanz der Merkmalerfassung sowie die Akzeptanz von Telematik-Angeboten insgesamt bei bestimmten Personengruppen mit der individuellen Betroffenheit sowie dem Lebenskontext zusammenhängen:

- Nutzer öffentlicher Verkehrsmittel befürworten die automatische Weiterleitung der Geschwindigkeitsüberschreitung von Autofahrern an die Polizei,
- Menschen, die ohnehin sportlich sind, befürworten eher Gesundheits-Telematik,
- Personen, die Punkte in Flensburg haben, lehnen Kfz-Telematik eher ab.

Wenngleich die Akzeptanzraten bei spezifischen Personengruppen im Einzelnen höher ausfallen, ist die Akzeptanz hinsichtlich der Erfassung von Merkmalen für Telematik-Tarife grundsätzlich gering ausgeprägt. Darüber hinaus ist auch vor dem Hintergrund, dass die Faktoren, die mutmaßlich zu einer höheren Akzeptanz führen können, jedoch nicht von allen Personen immer selbst ohne weiteres beeinflussbar sind (z. B. Nutzung öffentlicher Verkehrsmittel oder unbeeinträchtigte körperliche Bewegung), erscheint eine Beibehaltung von Telematik-freien Optionen in der Tarifgestaltung begrüßenswert.

Unter Gerechtigkeitsgesichtspunkten kann es zudem als problematisch erachtet werden, dass Scoring-Angebote im Bereich der Telematik insbesondere bei Personen mit ausgeprägter Preisorientierung höhere Akzeptanz finden. Zwar ist es grundsätzlich zu begrüßen, wenn durch Scoring Tarifvergünstigungen ermöglicht werden – ein impliziter Zwang zur Nutzung von Telematik und damit einhergehend einer Preisgabe persönlicher Daten aufgrund potenzieller finanzieller Problemlagen wären aus Verbrauchersicht jedoch wenig wünschenswert. Auch aus diesen Gründen wäre daher zu überlegen, Verbrauchern stets eine Telematik-freie (datensparsame) Option anzubieten, um eine echte Wahlfreiheit zu gewährleisten.

3.4 Zu Super-Scores

Die Befragten lehnen eine Veröffentlichung von Scores, die sie charakterisieren, mit großer Mehrheit ab. Dies gilt sowohl für eine Veröffentlichung auf freiwilliger Basis als auch für eine allgemeine Pflicht zur Veröffentlichung. Eine Verknüpfung von Scores aus verschiedenen Lebensbereichen lehnen die Befragten mit teils sehr großer Mehrheit ab.

Insbesondere der auf einem Score basierenden monetären Sanktionierung im Bereich des Gesundheitsverhaltens und der Verknüpfung von Scores, die den gesamten Lebensbereich umfassen, wird mit großer Ablehnung begegnet.

Der Rechtsrahmen des Scorings



Der SVRV definiert Scoring als die Zuordnung eines Zahlenwertes zu einem Menschen zum Zweck der Verhaltensprognose oder der Verhaltenssteuerung. Die Bestimmung dieses Zahlenwertes erfolgt in der Regel auf der Grundlage einer breiten Datenbasis durch ein algorithmisches Verfahren (siehe A.I). Für so definiertes „Scoring“ bestehen zwar durchaus rechtliche Vorgaben. Diese Vorgaben sind aber in der Rechtsordnung an ganz unterschiedlichen Stellen verortet. Eine übergreifende Regelung des Scorings, gar ein „Scoring-Gesetz“, gibt es nicht.

Thematisch beschränkte Anforderungen an die Rechtskonformität des Scorings ergeben sich aus verschiedenen Normkomplexen. Je nachdem, wer ein Scoring im oben genannten Sinne durchführt und mit welchem Zweck dies geschieht, wer und in welchem Lebensbereich gescort wird und welche rechtlichen oder tatsächlichen Folgen sich an den ermittelten Score knüpfen, kommen verschiedene Rechtsgebiete und Vorschriften zur Anwendung und sind für die Rechtmäßigkeit des Scorings maßgeblich. Bereits an dieser Stelle ist zur Klarstellung vorzuschicken, dass auch das „Scoring“, wie es im Bundesdatenschutzgesetz (§ 31 BDSG) explizit geregelt ist und so die öffentliche Diskussion prägt, das hier betrachtete Phänomen nur ausschnittsweise erfasst (näher unter E.I.3). Scoring liegt quer zu etablierten Rechtsbereichen und ist als ein spezifisches regelungsbedürftiges Phänomen bislang nicht zum Gegenstand gesetzgeberischen Handelns geworden.

Nachfolgend werden zentrale Berührungspunkte zwischen Scoring und der Rechtsordnung in den Blick genommen. Zunächst wird das Augenmerk auf Scoring als Datenverarbeitungsvorgang gelegt, sodann werden bereichsspezifische Vorgaben für das Scoring beschrieben. Dies geschieht angesichts der Heterogenität der unter dem Begriff des Scorings zusammengefassten Erscheinungen notwendig in exemplarischer Form. Hierauf aufbauend wird das Potenzial des geltenden Rechts für die Bewältigung allgemeiner mit dem Scoring verbundener Probleme – betreffend die mathematisch-statistische Güte, die Transparenz und das Diskriminierungsverbot – erörtert. Abschließend soll ein kurzer Blick auf Aufsichtsstrukturen geworfen werden, an die für eine Durchsetzung weiterreichender Scoring-Anforderungen (siehe die in Kapitel F. formulierten Handlungsempfehlungen) angeknüpft werden könnte.

I. Datenschutzrechtliche Grundlagen

Am ehesten erfüllen Art. 22 der Datenschutz-Grundverordnung (DSGVO) und § 31 BDSG die Funktion einer allgemeinen Regelung des Scorings. Keine der beiden Vorschriften erfasst Scoring, wie es hier verstanden wird, indessen in seiner Gesamtheit. Die Definition des Profilings in der Datenschutz-Grundverordnung gibt für eine zielgerichtete Regulierung des Scorings nichts her, da mit ihr keine fest umrissenen Rechtsfolgen verbunden sind (dazu sogleich unter 1.).

Als Vorgang der Datenverarbeitung muss Scoring selbstverständlich auch den allgemeinen datenschutzrechtlichen Vorgaben genügen. Zu den Herausforderungen, die das Scoring für das Institut der datenschutzrechtlichen Einwilligung (siehe auch B.VIII.2), des Zweckbindungsgrundsatzes und des Grundsatzes der Datenminimierung bedeutet, sei verwiesen auf das dieses Gutachten ergänzende SVRV Working Paper „Verbraucher-Scoring aus Sicht des Datenschutzrechts“ (Domurath & Neubeck, 2018).

Nicht sicher abzuschätzen ist derzeit, welche Prägekräfte die in der Datenschutz-Grundverordnung niedergelegten Grundsätze für die Verarbeitung personenbezogener Daten (Art. 5 Abs. 1 DSGVO) im Bereich des Scorings entfalten werden. Die maßstabsetzenden Potenziale dieser Prinzipien (Frenzel, 2018, Art. 5 DSGVO Rn. 55 f.) dürfen jedenfalls nicht von vornherein als gering veranschlagt werden, wie etwa Untersuchungen zum Grundsatz der Datenverarbeitung nach Treu und Glauben – *fair data processing* – gemäß Art. 5 Abs. 1 Buchstabe a DSGVO aufzeigen (Maxwell, 2015; Hacker, 2017).⁸¹

1. Profiling (Art. 4 Nr. 4 DSGVO)

Art. 4 der DSGVO definiert in Form eines Kataloges zahlreiche datenschutzrechtliche Grundbegriffe. Art. 4 Nr. 4 DSGVO enthält eine Definition des Profiling.



Art. 4 Nr. 4 DSGVO. Begriffsbestimmungen.

Im Sinne dieser Verordnung bezeichnet der Ausdruck: (...) „Profiling“ jede Art der automatisierten Verarbeitung personenbezogener Daten, die darin besteht, dass diese personenbezogenen Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten, insbesondere um Aspekte bezüglich Arbeitsleistung, wirtschaftliche Lage, Gesundheit, persönliche Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel dieser natürlichen Person zu analysieren oder vorherzusagen.

1.1 Profiling als rechtsfolgenloser Tatbestand

Scoring im Sinne dieses Gutachtens lässt sich zwanglos unter den Begriff des Profilings fassen (Domurath & Neubeck, 2018; Schild, 2018, Art. 4 DSGVO Rn. 64; Martini, 2018 Art. 22 DSGVO Rn. 7). Bloß bleibt die Definition des Profilings in der Datenschutz-Grundverordnung folgenlos. An den Tatbestand des Profilings knüpft die Datenschutz-Grundverordnung keine bestimmten Rechtsfolgen (Veil, 2018, Art. 4 Nr. 4 DSGVO Rn. 1 und Art. 22 DSGVO Rn. 4).

⁸¹ In der vieldiskutierten „Google Spain“-Entscheidung des Gerichtshofs der Europäischen Union (EuGH, Urteil [GK] vom 13. Mai 2014 - Rs. C-131/12 -, [ECLI:EU:C:2014:317]) leitete dieser aus Art. 6 Abs. 1 Buchstabe c bis e der Datenschutzrichtlinie (Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr; die Vorschrift ist der Vorgängerrechtsakt zur Datenschutz-Grundverordnung) ab, „dass es eine unzulässige Verarbeitung sei, wenn eine Internetsuchmaschine bestimmte Ergebnisse im Rahmen einer Namenssuche anzeige, und begründete so das „Recht auf Vergessenwerden“ (Schantz, 2018, Art. 5 DSGVO Rn. 2).

Zwar wird der Begriff des Profilings von mehreren Bestimmungen der Datenschutz-Grundverordnung aufgegriffen.⁸² In jeder dieser Vorschriften könnte indessen der Begriff des Profilings entfallen, ohne dass sich am Regelungsumfang der Vorschrift etwas änderte (Veil, 2018, Art. 4 Nr. 4 DSGVO Rn. 3 f. und Art. 22 DSGVO Rn. 53).⁸³

Weiterhin nehmen mehrere Erwägungsgründe der Datenschutz-Grundverordnung auf das Profiling Bezug.⁸⁴ Bei den Erwägungsgründen handelt es sich zwar nicht um ein „Recht minderer Güte“, sondern um einen durch das unionale Primärrecht vorgesehenen (s. Art. 296 UAbs. 2 AEUV) vollwertigen Bestandteil des jeweiligen Unionsrechtsaktes. Allerdings werden durch Erwägungsgründe keine Rechte oder Pflichten begründet, so dass es auch mit Blick auf diese Erwägungen von Profiling missverständlich wäre, hier von einer Regelung zu sprechen.

Erklärbar ist dieser Befund damit, dass sich über die Rechtsfolgen des Profilings im europäischen Gesetzgebungsverfahren zur Datenschutz-Grundverordnung keine Einigkeit erzielen ließ. Strittig war nicht nur, was aus dem Tatbestand Profiling folgen sollte (Veil, 2018, Art. 4 Nr. 4 DSGVO Rn. 9 ff.; vgl. auch WP 29, 2013), sondern bereits auf die vorgelagerte Frage, auf welche Merkmale und Umstände abgehoben werden soll, um dem Phänomen Profiling überhaupt Konturen zu verleihen, waren im Laufe der Verhandlungen über die Datenschutz-Grundverordnung verschiedene Antworten gegeben worden (Veil, 2018, Art. 4 Nr. 4 DSGVO Rn. 5 f.). Womöglich hat die Entscheidung des Gesetzgebers, an das Profiling keine Rechtsfolgen zu knüpfen, den in Art. 4 Nr. 4 DSGVO dokumentierten formalen Konsens erst ermöglicht (SVRV, 2016).

1.2 Profiling als Abwägungsgesichtspunkt

„Dass das Profiling gleichwohl“ – das heißt trotz seiner rechtlichen Folgenlosigkeit – in der Datenschutz-Grundverordnung „definiert wird, hat eine lediglich politische Bedeutung und soll signalisieren, dass der Normgeber die mit verschiedenen Formen des Profilings verbundenen Herausforderungen zumindest erkannt hat.“ (Veil, 2018, Art. 4 Nr. 4 DSGVO Rn. 1). Über diese „Statement-Funktion“ einer Definition des Profilings hinausgehend wird man zudem annehmen dürfen, dass das Vorliegen einer Profiling-Datenverarbeitung überall dort eine abwägungsdirigierende Wirkung entfalten wird, wo die Interessen von Verarbeitern und Betroffenen auszutarieren sind. Denn die Datenschutz-Grundverordnung regelt die Zulässigkeit von Datenverarbeitungen und die Voraussetzungen für eine Ausübung von Betroffenenrechten an vielen Stellen durch Generalklauseln (Buchner, 2017; zu den Gründen für dieses datenschutzrechtliche Regelungsmodell siehe Petersen, 2000; Hoffmann-Riem, 1998).

Generalklauseln sind in einem besonderen Maße offen dafür, eine Vielzahl an situativ zu bestimmenden und zu gewichtenden Wertungsgesichtspunkten aufzunehmen. Sie verlangen den Ausgleich widerstreitender Interessen und Gemeinwohlbelange, geben das Ergebnis dieses Ausgleichs aber nicht vor. Generalklauseln leiten die Rechtsanwendung deshalb in einem vergleichsweise schwachen Maße an. Das betrifft sowohl die Anleitung der – um Compliance bemühten – datenschutzrechtlich Verantwortlichen (Art. 4 Nr. 7 DSGVO), die ihre Datenverarbeitungsprozesse und -strukturen rechtskonform ausgestalten wollen, als auch die Programmierung der Kontrolltätigkeit von Aufsichtsbehörden und Gerichten, die über die Rechtmäßigkeit solcher Datenverarbeitungsarrangements letztlich zu entscheiden haben.

Ist eine Datenverarbeitung unter den Begriff des Profilings im Sinne von Art. 4 Nr. 4 DSGVO zu fassen, ist es legitim, das Schutzbedürfnis der betroffenen Person mit erheblichem Gewicht in eine gesetzlich angeord-

⁸² Siehe Art. 13 Abs. 2 Buchst. f; Art. 14 Abs. 2 Buchst. g; Art. 15 Abs. 1 Buchst. h; Art. 21 Abs. 1 Satz 1 Halbsatz 2, Abs. 2 Halbsatz 2; Art. 35 Abs. 3 Buchst. a; Art. 47 Abs. 2 Buchst. e; Art. 70 Abs. 1 Buchst. f DSGVO.

⁸³ Soweit ersichtlich einzige Ausnahme: Der Europäische Datenschutzausschuss (Art. 68 ff. DSGVO) hat unter anderem die Aufgabe, Leitlinien usw. zur näheren Bestimmung der Kriterien und Bedingungen für die auf Profiling beruhenden Entscheidungen gemäß Art. 22 Abs. 2 DSGVO bereitzustellen (Art. 70 Satz 2 Buchstabe f DSGVO); nicht jede der von Art. 22 Abs. 2 DSGVO erfassten automatisierten Einzelfallentscheidungen beruht aber auf Profiling-Vorgängen. Ein Grund für die nach dem Textbefund der Vorschrift gegebene Einschränkung des Aufgabenbereichs des Ausschusses (die man nicht als Redaktionsversehen abtun kann, s. EG 72 Satz 2 DSGVO) ist nicht erkennbar.

⁸⁴ Siehe vor allem EG 71 DSGVO, weiter EG 60 Satz 3, EG 63 Satz 3, EG 70 Satz 1, EG 72 Satz 1 und 2, EG 91 Satz 2 DSGVO.

nete Abwägung aufzunehmen. Der Schutzbedürftigkeit der betroffenen Person, die der 71. Erwägungsgrund der Datenschutz-Grundverordnung ausdrücklich anerkennt, kann dadurch im Rahmen der Rechtsanwendung Rechnung getragen werden. Illustrieren lässt sich dies an Art. 6 Abs. 1 Satz 1 Buchstabe f DSGVO. Nach dieser Bestimmung ist eine Datenverarbeitung rechtmäßig, wenn „die Verarbeitung (...) zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich [ist], sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen“. Im Rahmen der von der Vorschrift geforderten Abwägung ist dem Interesse des Betroffenen, ohne seine Einwilligung keiner Profiling-Datenverarbeitung unterworfen zu sein, ein beträchtliches Gewicht einzuräumen (Buchner, 2018, Art. 4 Nr. 4 DSGVO Rn. 8).

1.3 Regulierungspotenziale des Scorings durch Art. 4 Nr. 4 DSGVO

Als Fazit ist festzuhalten, dass mit der Definition des Profilings in Art. 4 Nr. 4 DSGVO keine Regelung dieses Phänomens stattfindet, sondern lediglich die gesellschaftliche Bedeutung dieses Datenverarbeitungsvorgangs herausgestellt worden ist. Eine sachangemessene Regulierung des Scorings kann auf diesem Fundament nicht gegründet werden. Eine das gesetzliche Regelungsprogramm ergänzende einzelfallübergreifende Maßstabbildung durch den Gerichtshof der Europäischen Union und die nationalen Gerichte ist zwar durchaus möglich, ist aber angesichts der bisherigen relativ schwachen justiziellen Durchformung des Datenschutzrechts kein allzu realistisches Szenario.

2. Automatisierte Einzelfallentscheidung (Art. 22 DSGVO)

i

Art. 22 DSGVO. Automatisierte Entscheidungen im Einzelfall einschließlich Profiling.

(1) Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

(2) Absatz 1 gilt nicht, wenn die Entscheidung

- a) für den Abschluss oder die Erfüllung eines Vertrags zwischen der betroffenen Person und dem Verantwortlichen erforderlich ist,
- b) aufgrund von Rechtsvorschriften der Union oder der Mitgliedstaaten, denen der Verantwortliche unterliegt, zulässig ist und diese Rechtsvorschriften angemessene Maßnahmen zur Wahrung der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person enthalten oder
- c) mit ausdrücklicher Einwilligung der betroffenen Person erfolgt.

(3) In den in Absatz 2 Buchstaben a und c genannten Fällen trifft der Verantwortliche angemessene Maßnahmen, um die Rechte und Freiheiten sowie die berechtigten Interessen der betroffenen Person zu wahren, wozu mindestens das Recht auf Erwirkung des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunkts und auf Anfechtung der Entscheidung gehört.

(4) Entscheidungen nach Absatz 2 dürfen nicht auf besonderen Kategorien personenbezogener Daten nach Artikel 9 Absatz 1 beruhen, sofern nicht Artikel 9 Absatz 2 Buchstabe a oder g gilt und angemessene Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person getroffen wurden.

2.1 Der Aufbau der Vorschrift und ihre Rechtsfolgen

Art. 22 DSGVO statuiert Anforderungen an die Rechtmäßigkeit automatisierter Einzelfallentscheidungen. Die Vorschrift weist eine eindrucksvolle legislative Komplexität auf: sie kennt eine Grundsatzentscheidung (Absatz 1), die Ausnahme hierzu (Absatz 2), sodann eine Gegen Ausnahme (Absatz 4, erster Satzteil) und schließlich die Gegen-Gegenausnahme (Absatz 4, zweiter Satzteil).

Absatz 1 der Vorschrift statuiert den Grundsatz, dass einem Betroffenen ein Abwehrrecht zusteht, einer solchen Entscheidung dann nicht unterworfen zu werden, wenn sie ihm gegenüber rechtliche Wirkung entfaltet oder ihn in ähnlicher Weise erheblich beeinträchtigt. Absatz 2 der Vorschrift sieht drei Ausnahmen von diesem Grundsatz vor und gestattet automatisierte Einzelfallentscheidungen zur Vertragsdurchführung, auf gesetzlicher Grundlage oder mit Einwilligung. Der deutsche Gesetzgeber hat mit § 37 BDSG eine solche Grundlage für eine weitergehende Einschränkung geschaffen. Absatz 4 schränkt die Ausnahmetrias für bestimmte Fälle ein, in denen die automatisierte Entscheidung auf besonderen Kategorien personenbezogener Daten (Art. 9 Abs. 1 DSGVO) beruht. Ist nach diesem Regelungskomplex eine automatisierte Einzelfallentscheidung zulässig, ordnet Art. 22 DSGVO ergänzend an, dass angemessene Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen des Entscheidungsunterworfenen vorzusehen sind (Absatz 2 Buchstabe b, Absatz 3, Absatz 4).

An das Vorliegen einer automatisierten Einzelfallentscheidung nach Art. 22 DSGVO knüpft die Datenschutz-Grundverordnung zahlreiche Folgen (Zusammenstellung bei Veil, 2018: Art. 22 DSGVO Rn. 16 ff.). Zu diesen Folgen gehören spezifische Informationspflichten vor der Datenverarbeitung (Art. 13 Abs. 2 Buchstabe f DSGVO für die Datenerhebung beim Betroffenen, Art. 14 Abs. 2 Buchstabe g DSGVO für die Datenerhebung bei anderen

Personen) und Auskunftsrechte nach der Datenverarbeitung (Art. 15 Abs. 1 Buchstabe h DSGVO), siehe hierzu weiter E.III.1. Hinzu treten „system- und verfahrensbezogene Pflichten“ (Dreyer/Schulz, 2018, S. 32 ff.), zu denen insbesondere die Verpflichtung zur Durchführung einer Datenschutz-Folgenabschätzung (Art. 35 DSGVO), zum Erlass interner Datenschutzvorschriften (Art. 47 Abs. 1 Buchstabe e DSGVO) und zur Bestellung eines Datenschutzbeauftragten (Art. 37 DSGVO in Verbindung mit § 38 Abs. 1 Satz 2 BDSG und Art. 35 DSGVO) gehören.

2.2 Die „Entscheidung“ als zentrales Kriterium in Art. 22 DSGVO

Seine Potenziale zur Regulierung des Scorings kann der Art. 22 DSGVO nur entfalten, wenn für das Scoring die Tatbestandsvoraussetzung der „auf einer automatisierten Verarbeitung beruhenden Entscheidung“, die „rechtliche Wirkungen entfaltet oder in ähnlicher Weise erheblich beeinträchtigt“, erfüllt ist. Das auf diese Weise beschriebene Phänomen soll im Folgenden unter dem Begriff der „Entscheidung“ im Sinne des Art. 22 DSGVO zusammengefasst werden.

In welchen Fällen eine „Entscheidung“ im Sinne von Art. 22 DSGVO vorliegt, ist alles andere als evident (Gesellschaft für Informatik, 2018). Gerade am Beispiel des Scorings lässt sich diese Unsicherheit gut veranschaulichen. Schon die Zuordnung einer bestimmten Zahl x zu einer Person P lässt sich als eine Entscheidung bezeichnen – man denke etwa an Preisrichterentscheidungen im Turnen oder an die Vergabe von Schulnoten. Sähe man in jedem Zuordnungsvorgang dieser Art allerdings bereits eine Entscheidung im Sinne von Art. 22 DSGVO, wäre eine geradezu ubiquitäre Anwendbarkeit der Vorschrift geboten. Das Merkmal der Entscheidung würde jede Unterscheidungskraft einbüßen und keine tatbestandseinschränkende Funktion mehr erfüllen können.

Durch das Tatbestandsmerkmal der „Entscheidung“ trifft Art. 22 DSGVO die Unterscheidung zwischen Interaktionen, in denen ein Mensch über oder für einen anderen Menschen entscheidet einerseits, und strukturell parallelen Maschine-Mensch-Interaktionen andererseits; (nur) für letztere errichtet die Datenschutz-Grundverordnung das oben beschriebene besondere Regelungsregime mit Art. 22 DSGVO in seinem Zentrum. Allerdings fehlt dem Gesetzgeber eine klare Vorstellung davon, was

an den zu regelnden Maschinen-Entscheidungen eigentlich das problematische Moment sein soll (Dammann, 2016; Veil, 2018, Art. 22 DSGVO Rn. 3), das ihre legislative Einhegung rechtfertigen soll. Dass die Vorschrift „Ausdruck eines diffusen allgemeinen Unbehagens“ (Schulz, 2017, Art. 22 DSGVO Rn. 2; siehe auch B.VII.2) gegenüber von Computern getroffenen Entscheidungen darstellt, erscheint vor diesem Hintergrund als eine angemessene Charakterisierung.

Folge des Fehlens einer „erkennbaren Schutzstrategie“ (Veil, 2018, Art. 22 DSGVO Rn. 4) der Vorschrift ist, dass Reichweite und Interpretation des Merkmals der Entscheidung unsicher sind. Verstärkt wird die Unsicherheit dadurch, dass zu den von Art. 22 DSGVO erfassten Entscheidungen ausweislich des 71. Erwägungsgrundes der Datenschutz-Grundverordnung auch „Maßnahmen“ zu zählen sind – ein Begriff, der deutlich stärker als der der „Entscheidung“ die Herbeiführung faktischer Wirkungen (im Gegensatz zu rechtlichen) beim Betroffenen konnotiert (Veil, 2018, Art. 22 DSGVO Rn. 58; siehe aber Abel, 2018, der in der Wendung ein Fossil früherer Regelungsambitionen der Kommission sieht und demzufolge vor Überinterpretationen warnt). Deshalb kann der Anwendungsbereich von Art. 22 DSGVO jedenfalls nicht reduziert werden etwa auf die Abgabe einer elektronischen Willenserklärung (zu ihnen eingehend Wiebe, 2002) oder auf sonstige Umgestaltungen der Rechtslage des Betroffenen unmittelbar durch das Ergebnis eines Datenverarbeitungsvorgangs. Ebenso wenig kann eine lediglich formale Betrachtungsweise maßgeblich sein, nach der es darauf ankäme, ob ein Mensch der betroffenen Person die Entscheidung mitteilt oder für den Betroffenen als Urheber der Entscheidung erscheint. Diese Betrachtungsweise würde den Anwendungsbereich des Art. 22 DSGVO in gänzlich unsachgemäßer Art und Weise beschränken. Denn die betroffene Person ist nicht allein deshalb weniger schutzbedürftig, weil ihr das Ergebnis eines Datenverarbeitungsvorgangs durch einen Menschen mitgeteilt wird und nicht über, zum Beispiel, ein Computerinterface.

Die rechtswissenschaftliche Kommentarliteratur zu Art. 22 DSGVO (und ihrer Vorgängernormen in Art. 15 DSRL und § 6a BDSG a.F., siehe Bygrave, 2001) bietet zur Identifizierung jener Maschinenhandlungen, die als „Entscheidungen“ den Rechtsfolgen dieser Vorschrift unterworfen werden sollen, nur eine schwache Anleitung. Sie behilft

sich zur Erläuterung des Tatbestandsmerkmals mit darstellerischen Strategien, die eindeutige Fälle der Anwendbarkeit von Art. 22 DSGVO beschreiben. Dadurch wird dem Rechtsanwender die Möglichkeit zum analogen Argumentieren eröffnet und ein gewisses Maß an Vorhersehbarkeit in der Rechtsanwendung erreicht. Betont wird, dass es gerade nicht die „Entscheidungsunterstützung“ durch sogenannte *decision support systems* ist, die von der Vorschrift erfasst sein soll. Das heißt, immer dann, wenn es einen menschlichen Letztentscheider gibt, findet Art. 22 DSGVO keine Anwendung. Damit allerdings ist angesichts des komplexen und arbeitsteiligen Zusammenspiels maschineller und menschlicher Handlungen eher das Problem markiert als ein Lösungskriterium genannt.

Zahlreiche maschinelle Handlungen (zur Handlungsträgerschaft technischer Systeme siehe Schulz-Schaeffer, 2007) sind eingebettet in komplexe „Entscheidungsarchitekturen“ (Thaler & Sunstein, 2008, S. 89 ff.; Sunstein & Reisch, 2014, S. 157), an denen sowohl menschliche Entscheider als auch Maschinen beteiligt sind. Maschinen nehmen Priorisierungen vor, sortieren und klassifizieren und lenken so die Aufmerksamkeit menschlicher Entscheider (Wu, 2016; Mengden, 2018). Sie präformieren die Räume, in denen sich menschliche Handlungsautonomie verwirklicht, und strukturieren menschliche Entscheidungsroutrinen vor. Menschen treffen Entscheidungen üblicherweise nicht gleichsam vom leeren Blatt aus, sondern fügen mit der Entscheidung einem komplexen sozialen Gewebe einen weiteren Faden hinzu.

Eine tragfähige Interpretation der Tatbestandsvoraussetzung der Entscheidung muss das jeweilige „sozio-technische System“ (Ropohl, 2009, S. 58 f.) insgesamt in den Blick nehmen. Eine die Anwendbarkeit des Art. 22 DSGVO eröffnende „Entscheidung“ wird man demnach umso plausibler annehmen dürfen, erstens je weniger man erwarten kann, dass ein Mensch das Ergebnis des automatisierten Datenverarbeitungsvorgangs ändert oder für unmaßgeblich erachtet und zweitens je sozial erheblicher das Ergebnis des Datenverarbeitungsvorgangs ist. Denn nur dann ergibt die gesetzgeberische Festlegung Sinn, dem Betroffenen einen Anspruch auf Eingreifen einer Person seitens des Verantwortlichen einzuräumen (Art. 22 Abs. 3 DSGVO). Wenn dagegen das Ergebnis des Datenverarbeitungsvorgangs bloß als ein Gesichtspunkt in eine auf eine breite Tatsachenbasis ge-

stellte menschliche Entscheidung einget, trüge der statuierte Anspruch auf einen „human in the loop“ (Dreyer & Schulz, 2018, S. 36) zum Betroffenenenschutz nichts bei.

2.3 Wann liegt eine „Entscheidung“ im Sinne des Art. 22 DSGVO vor?

Um zu bestimmen, ob eine Entscheidung im Sinne von Art. 22 DSGVO vorliegt, sind die Entscheidungswege und -routinen der datenverarbeitenden Organisation übergreifend in den Blick zu nehmen. Es ist der Frage nachzugehen, auf welche Weise menschliche Entscheider die Ergebnisse von Datenverarbeitungsvorgängen in ihre eigenen anschließenden Entscheidungen einfließen lassen. Erreichen die Verfahren der „Herstellung von Entscheidungen“ (Hoffmann-Riem, 2016, S. 98 ff.) eine gewisse Komplexität, wird man Anhaltspunkte für und wider die Anwendbarkeit des Art. 22 DSGVO ausmachen können.

Als ein erstes Indiz gegen das Vorliegen einer automatisierten Entscheidung im Sinne des Art. 22 DSGVO ist es zu werten, wenn die Datengrundlage, auf der ein Computer tätig geworden ist, auch für den Menschen offenliegt. Das dem menschlichen Entscheider präsentierte Ergebnis des Datenverarbeitungsvorgangs ist dann kein kontextfreies und opakes Expertenurteil (zur Gefahr der Dekontextualisierung von Scores siehe oben B.I.5), sondern hat eher den Charakter einer „zweiten Meinung“. Macht sich ein menschlicher Entscheider von der Präformierung seines Entscheidungsraums mitunter tatsächlich frei und setzt sich über das Computervotum hinweg, liegt hierin ein gewichtiger Anhaltspunkt gegen das Vorliegen einer automatisierten Einzelfallentscheidung nach Art. 22 DSGVO vor. Ist einem menschlichen Entscheider organisationsintern die Befugnis eingeräumt, auf der Grundlage einer nicht weiter explizierten Gesamtwürdigung aller Umstände ein eigenes Urteil zu treffen oder atypischen Merkmalen des Einzelfalles Rechnung zu tragen, ist dies ebenfalls ein gegen die Anwendbarkeit von Art. 22 DSGVO sprechendes Moment. Anders wird man die einer bestimmten Regel gehorchende Abweichung des menschlichen Entscheiders von der maschinellen Vor-Entscheidung bewerten müssen. Hier arbeiten Maschine und Mensch arbeitsteilig an verschiedenen Aspekten der Letztentscheidung zusammen, so dass man die menschliche Entscheidung einer Anwendbarkeit von Art. 22 DSGVO nicht entgegenhalten

kann. Als ein Indiz für eine automatische Entscheidung im Sinne von Art. 22 DSGVO wird man es werten müssen, wenn Abweichungen vom maschinellen Urteil zwar möglich, aber mit erheblichen Haftungsrisiken verbunden sind. Insbesondere im medizinischen Bereich könnte auf dieser Grundlage dem Art. 22 DSGVO ein breiter Anwendungsbereich eröffnet werden.

2.4 Regulierungspotenziale des Scorings durch Art. 22 DSGVO

An diesen Maßstäben gemessen wird beim Scoring der Anwendungsbereich des Art. 22 DSGVO zwar oft, aber nicht definitionsgemäß eröffnet sein. Der Anwendungsbereich wird zumeist deshalb eröffnet sein, weil der Score, erstmal ermittelt, regelmäßig wie ein Etikett an der gescorten Person haften und in nachfolgenden Entscheidungen wie eine „objektive Tatsache“ über diese Person behandelt werden wird (siehe oben B.I.5). Das heißt, die empirische Grundlage für diese Score-Bildung wird als solche nicht mehr zum Gegenstand menschlicher Beurteilung gemacht werden – einmal gepackt, wird das Score-Paket nicht wieder aufgeschnürt. Im Gegensatz dazu wird dort, wo ein bestimmter Score zwar *prima facie* für einen menschlichen Entscheider beachtlich ist, aber dieser Entscheider mit einiger Häufigkeit auf die dem Score zugrunde gelegten Daten des Betroffenen durchgreift und sich ein eigenes Urteil bildet, der Anwendungsbereich von Art. 22 DSGVO nicht eröffnet sein.

3. Scoring von Wahrscheinlichkeitswerten (§ 31 BDSG)

§ 31 BDSG regelt die Anforderungen an die „Verwendung eines Wahrscheinlichkeitswerts über ein bestimmtes zukünftiges Verhalten einer natürlichen Person zum Zweck der Entscheidung über die Begründung, Durchführung oder Beendigung eines Vertragsverhältnisses mit dieser Person“ und definiert diese Handlung als „Scoring“. Der Gegenstand gesetzgeberischer Regelungstätigkeit ist also nicht das

§ 31 BDSG. Schutz des Wirtschaftsverkehrs bei Scoring und Bonitätsauskünften.

(1) Die Verwendung eines Wahrscheinlichkeitswerts über ein bestimmtes zukünftiges Verhalten einer natürlichen Person zum Zweck der Entscheidung über die Begründung, Durchführung oder Beendigung eines Vertragsverhältnisses mit dieser Person (Scoring) ist nur zulässig, wenn

1. die Vorschriften des Datenschutzrechts eingehalten wurden,
2. die zur Berechnung des Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich sind,
3. für die Berechnung des Wahrscheinlichkeitswerts nicht ausschließlich Anschriftendaten genutzt wurden und
4. im Fall der Nutzung von Anschriftendaten die betroffene Person vor Berechnung des Wahrscheinlichkeitswerts über die vorgesehene Nutzung dieser Daten unterrichtet worden ist; die Unterrichtung ist zu dokumentieren.

(2) Die Verwendung eines von Auskunftseien ermittelten Wahrscheinlichkeitswerts über die Zahlungsfähig- und Zahlungswilligkeit einer natürlichen Person ist im Fall der Einbeziehung von Informationen über Forderungen nur zulässig, soweit die Voraussetzungen nach Absatz 1 vorliegen und nur solche Forderungen über eine geschuldete Leistung, die trotz Fälligkeit nicht erbracht worden ist, berücksichtigt werden,

1. die durch ein rechtskräftiges oder für vorläufig vollstreckbar erklärtes Urteil festgestellt worden sind oder für die ein Schuldtitel nach § 794 der Zivilprozessordnung vorliegt,
2. die nach § 178 der Insolvenzordnung festgestellt und nicht vom Schuldner im Prüfungstermin bestritten worden sind,
3. die der Schuldner ausdrücklich anerkannt hat,
4. bei denen
 - a) der Schuldner nach Eintritt der Fälligkeit der Forderung mindestens zweimal schriftlich gemahnt worden ist,
 - b) die erste Mahnung mindestens vier Wochen zurückliegt,
 - c) der Schuldner zuvor, jedoch frühestens bei der ersten Mahnung, über eine mögliche Berücksichtigung durch eine Auskunft unterrichtet worden ist und
 - d) der Schuldner die Forderung nicht bestritten hat oder
5. deren zugrunde liegendes Vertragsverhältnis aufgrund von Zahlungsrückständen fristlos gekündigt werden kann und bei denen der Schuldner zuvor über eine mögliche Berücksichtigung durch eine Auskunft unterrichtet worden ist.

Die Zulässigkeit der Verarbeitung, einschließlich der Ermittlung von Wahrscheinlichkeitswerten, von anderen bonitätsrelevanten Daten nach allgemeinem Datenschutzrecht bleibt unberührt.

Prognoseverfahren als solches, sondern der Umgang mit Score-Werten im Rechtsverkehr.⁸⁵

Ob dem deutschen Gesetzgeber die Kompetenz zustand, die Datenschutz-Grundverordnung um eine nationale scoring-spezifische Sonderbestimmung zu ergänzen, ist in der Rechtswissenschaft umstritten (siehe etwa Buchner, 2018, § 31 BDSG Rn. 4). Zu einer Entscheidung des Gerichtshofs der Europäischen Union in dieser Frage könnte es kommen, wenn ein von § 31 BDSG verpflichteter Scorer sich durch die in der Vorschrift angeordneten Restriktionen nicht gebunden sieht. Insbesondere die Beschränkung der Möglichkeiten zum Geo-Scoring (§ 31 Abs. 1 Nr. 3 BDSG) könnten einen wirtschaftlichen Anreiz bieten, die Frage der Unionsrechtskonformität des § 31 BDSG vor Gericht zu tragen.

3.1 Scoring im Sinne des Gutachens und Scoring nach § 31 BDSG

Der gesetzgeberische Zugriff auf einen regulierungsbedürftigen Bereich nach § 31 BDSG unterscheidet sich vom Zuschnitt des Scorings, wie es in diesem Gutachten verstanden wird, in zweierlei Hinsicht: Erstens nimmt das Gutachten das Scoring-Verfahren insgesamt in den Blick, ohne dieses bereits definitorisch an einen bestimmten sozialen, wirtschaftlichen oder rechtlichen Verwendungszweck zu knüpfen. Zweitens bezieht das Gutachten auch solche Score-Wert-Bildungen in die Betrachtung mit ein, bei denen der Score-Wert nicht die Wahrscheinlichkeit eines bestimmten zukünftigen Verhaltens ausdrückt, sondern der Score das Verhalten beeinflussen soll.

3.2 Der Regelungsgegenstand des § 31 BDSG

§ 31 BDSG erfasst jene Scoring-Verfahren, in denen der ermittelte Score unmittelbar ein Wahrscheinlichkeitsurteil über das zukünftige Auftreten eines bestimmten Verhaltens darstellt. Die Vorschrift verwendet den in der statistischen Fachliteratur eher ungebräuchlichen Begriff des „Wahrscheinlichkeitswertes“, statt schlicht von einer „Wahrscheinlichkeit“ zu sprechen. Ein sachlicher Unterschied ist damit nicht verbunden. Die Abweichung von der Terminologie der Stochastik (ex-

emplarisch Christoph & Hackel, 2002) dürfte vielmehr dem Umstand geschuldet sein, dass der Begriff der „Wahrscheinlichkeit“ rechtlich gewissermaßen schon besetzt war. In anderen Vorschriften werden mit ihm nicht exakte Werte, sondern qualitative, nicht durch Zahlen ausgedrückte Urteile über Kausalverläufe bezeichnet (siehe etwa im Recht der sozialen Entschädigung: § 1 Abs. 3 BVG, § 81 Abs. 6 SVG; im Polizeirecht: § 56 Abs. 1 Nr. 2 BKAG; im Schadenersatzrecht: § 252 BGB). Mit dem Begriff des „Wahrscheinlichkeitswertes“ in § 31 BDSG will der Gesetzgeber an die begrifflichen Kategorien der Stochastik gerade anschließen statt sich von ihnen abzusetzen.

Die von § 31 BDSG erfassten Scoring-Verfahren ermitteln die Wahrscheinlichkeit in Form einer Zahl (zum Beispiel „ $p = 0,96$ “ oder „ $p = 96\%$ “) oder in Form von Darstellungen, die in Wahrscheinlichkeiten oder bestimmte Wahrscheinlichkeitsintervalle eindeutig übersetzbar sind. So macht es keinen Unterschied, wenn der Scorer das Ergebnis des Scoring-Verfahrens als, zum Beispiel, „Goldstatus“ bezeichnet, solange gleichzeitig feststellt, dass dieses Label für die Wahrscheinlichkeit eines bestimmten zukünftigen Verhaltens von $p \geq 0,99$ vergeben wird. Nicht erfasst von § 31 BDSG sind verhaltenssteuernde Scores, die keinen direkten Wahrscheinlichkeitsbezug aufweisen, etwa Gesundheits-Scores oder Fahrgüte-Scores, aber gleichwohl Prognosen enthalten (zu solchen Scores siehe unten 3.3.).

Der vom Scoring-Verfahren nach § 31 BDSG ermittelte prognostische Wahrscheinlichkeitswert muss sich auf ein bestimmtes zukünftiges Verhalten einer natürlichen Person beziehen. Das bedeutet, das „Ereignis“ (im Sinne der Stochastik, siehe Christoph & Hackel, 2002), dessen Wahrscheinlichkeit das Scoring-Verfahren ermittelt, muss juristisch als ein „Verhalten“ zu qualifizieren sein. Der Gesetzgeber wollte den Anwendungsbereich des § 31 BDSG „auf ein selbstbestimmtes Handeln“ beschränkt sehen, „so dass Ereignisse, die auf höhere Gewalt oder Fremdeinwirkung zurückgehen, ausscheiden (z. B. Blitzschlag, Diebstahl oder Erkrankung). Die Verfahren zur Tarifierung etwa von Lebens- oder Krankenversicherungen oder Versicherungen gegen Kfz-Diebstahl stellen daher kein Scoring im Sinne von § 28b [und jetzt: § 31 BDSG] dar.“ (BT-Drucks. 16/10529, S. 16).

⁸⁵ Das ist neu. § 28b BDSG a.F. ließ sich noch zwanglos so verstehen, dass das Prognoseverfahren selbst als „Scoring“ definiert werden soll.

Das selbstbestimmte Handeln, das § 31 BDSG erfassen soll, ist aber oft gerade eine vorhersehbare Reaktion auf Ereignisse, die der gescorten Person widerfahren. Zwischen „Opfer eines Diebstahls werden“ und „Diebstahlversicherungsschutz einfordern“ oder zwischen „krank werden“ und „Krankheitskosten geltend machen“ besteht ein enger Zusammenhang. Um den Anwendungsbereich des § 31 BDSG effektiv und vorhersehbar zu beschränken, erscheint die Anknüpfung an ein Tatbestandsmerkmal „Verhalten“ wenig geeignet. Dessen Trennschärfe wird mit steigender Leistungsfähigkeit von Scoring-Verfahren zunehmend prekär werden.

Die Schwierigkeiten einer Beschränkung des Anwendungsbereichs von § 31 BDSG auf Verhaltensprognosen zeigen sich bereits im Zentrum der Vorschrift, dem Bonitäts-Scoring. Die Entstehungsgeschichte des Scoring-Paragrafen § 28b BDSG a.F. stand ganz im Zeichen einer Regulierung der Tätigkeit der Auskunfteien. Deren mit dem Bonitäts-Score ausgedrückte Zielvariable ist aber mitnichten ein Verhalten des Schuldners, sondern das unpersönliche Auftreten eines Negativmerkmals (siehe C.III.3). Selbst im Wortlaut des § 31 BDSG ist die Unschärfe der Einschränkung auf ein „selbstbestimmtes Handeln“ (BT-Drucks. 16/10529, S. 16) der gescorten Person ablesbar. Der zweite Absatz regelt die Berechnung von Wahrscheinlichkeitswerten über die Zahlungsfähigkeit der gescorten Person: Diese ist aber unter Umständen ebensowenig Ergebnis des selbstbestimmten Handelns, von dem die Gesetzesbegründung spricht, wie die Erkrankung, die nach dem Willen des Gesetzgebers jedenfalls aus dem Anwendungsbereich von § 31 BDSG herausfallen soll.

3.3 Prognostisches Scoring jenseits von § 31 BDSG

Gleich, wie weit man „Ereignisse“ in der Terminologie der Stochastik als „Verhalten“ qualifiziert (siehe oben 3.2), erschöpft § 31 BDSG das Feld des prognostischen Scorings nicht. Auch solche Scoring-Verfahren ermöglichen Prognosen über zukünftige Verhaltensweisen, deren Score kein präzises Wahrscheinlichkeitsurteil darstellt. Denn gerechtfertigt ist die Annahme, dass

die Verhaltensweisen der gescorten Person, die in der Vergangenheit zur Bildung eines Scores in bestimmter Höhe beigetragen haben, von dieser Person auch in Zukunft an den Tag gelegt werden.⁸⁶ Man nehme zum Beispiel einen Gesundheits-Score an, bei dem Punkte für eine heterogene Vielzahl an gesundheitsförderlichen Verhaltensweisen verteilt werden (Bewegung, Alkoholabstinenz, ...). Hier ermöglicht ein hoher Score durchaus die generalisierende Prognose, dass die der Score-Berechnung zugrunde liegenden Verhaltensweisen auch in Zukunft in entsprechendem Ausmaß vorliegen werden. Vor diesem Hintergrund ist auch die Qualifizierung eines solchen Scores als prognostisch gerechtfertigt. Maßstäbe zur regulatorischen Bewältigung solcher Scores enthält § 31 BDSG allerdings nicht. Diese Vorschrift beschränkt sich ausweislich ihres Wortlauts auf solche Scores, die einem bestimmten Verhalten einen Wahrscheinlichkeitswert zuordnen.

3.4 Der datenschutzrechtliche Regulierungszugriff des § 31 BDSG

§ 31 BDSG adressiert Scoring lediglich als ein datenschutzrechtliches Problem. Damit fallen Datenverarbeitungsvorgänge, in die keine personenbezogenen Daten eingehen, von vornherein aus dem Regelungsbereich der Vorschrift heraus (§ 1 Abs. 1 BDSG, siehe auch die Überschrift von Teil 2 Kapitel 1 BDSG: „Rechtsgrundlagen der Verarbeitung personenbezogener Daten“). Der spezifisch datenschutzrechtliche Zugriff des § 31 BDSG bedeutet, dass Anforderungen an die Qualität und Fairness von Scoring-Verfahren gesetzgeberisch nur „indirekt“ bestimmt sind, nämlich als Anforderungen an die Verarbeitung der Daten, die zur Ermittlung von Score-Werten herangezogen werden.

Ein solcher spezifisch datenschutzrechtlicher Regulierungszugriff auf das Scoring ist keine Selbstverständlichkeit. Vor einer Kontrastfolie – dem Recht der Eigenmittelausstattung von Kreditinstituten – gewinnt er schärfere Konturen. Eine unzureichende Ausstattung eines Kreditinstituts mit Eigenmitteln betrifft nicht nur dessen eigenen wirtschaftlichen Bestand. Sie kann systemische Risiken verursachen und gravieren-

⁸⁶ Auch die vermeintlich präzisen, wahrscheinlichkeitsauswerfenden Scoring-Verfahren des § 31 BDSG beruhen auf dieser Annahme; dies ist auch nicht verblüffend, ist jede Modellbildung doch – trivialerweise – darauf verwiesen, für ihre Aussagen über die Zukunft aus dem in der Vergangenheit liegenden Datenbestand zu schöpfen.

de Folgen für die Stabilität und Funktionsfähigkeit des Finanzmarktes zeitigen. Deshalb ist die zureichende Ausstattung von Kreditinstituten mit Eigenmitteln ein wesentlicher Bestandteil der Regulierung des Finanzsystems. In welchem Ausmaß das Kreditgeschäft von Kreditinstituten mit Eigenmitteln zu unterlegen ist, hängt auch von den Ausfallrisiken der ausgegebenen Kredite ab. Ein Kredit mit geringem Ausfallrisiko bedarf einer geringeren Eigenkapitalunterlegung als einer mit höherem Ausfallrisiko. Deshalb sind die Rating-Systeme, d. h. die Modelle, nach denen das zur Risikounterlegung erforderliche Eigenkapital zu bemessen ist, ein intensiv regulierter Bereich (Paraschiakos, 2017; Wundenberg, 2012). Die zur Risikobemessung verwendeten Rating-Systeme müssen gesetzlich definierten Qualitätsanforderungen genügen. Die gesetzlichen Fundamente des Eigenmittelregulierungsrechts – siehe insbesondere die sogenannte Eigenkapitalrichtlinie⁸⁷ und die Kapitaladäquanzverordnung⁸⁸ – sind eine Spezialismaterie von beträchtlicher Komplexität.

Mit § 10 Abs. 2 KWG enthält das Gesetz über das Kreditwesen eine Regelung, die dem § 31 BDSG bis in die Formulierungen des Normtextes hinein ähnelt. Die Vorschrift gestattet es Kreditinstituten, personenbezogene Daten zu erheben und zu verwenden, um Rating-Systeme zu entwickeln und zu betreiben. Insoweit ergeben sich keine konzeptionellen Unterschiede zu § 31 BDSG, der Scoringern eine vergleichbare Erlaubnis für Scoring-Systeme einräumt. Ein wesentlicher Unterschied liegt aber darin, dass neben § 10 Abs. 2 KWG das soeben skizzierte umfangreiche Qualitätsgewährleistungsrecht für Rating-Systeme besteht. § 10 Abs. 2 KWG erlaubt datenschutzrechtlich, was gesetzlich an anderer Stelle gefordert ist. Solche anderweitigen gesetzlichen Qualitätsanforderungen bestehen im Bereich des Scorings indessen nicht.⁸⁹ Sollte man dem von § 31 BDSG erfassenden Scoring eine vergleichbare wirtschaftliche oder so-

ziale Relevanz beimessen wie der funktionsadäquaten Ausstattung von Kreditinstituten mit Eigenmitteln, so läge es nahe, auch den § 31 BDSG um ein nicht im Datenschutzrecht verankertes Scoring-Qualitätsgewährleistungsrecht als Seitenstück zu ergänzen. Dieses sähe sich auch nicht den Einwänden ausgesetzt, die gegen die Unionsrechtskonformität des § 31 BDSG erhoben werden (siehe oben 3.1).

3.5 Die Struktur des § 31 BDSG

§ 31 BDSG enthält in seinem Absatz 1 Regelungen für die Verwendung jeglicher prognostischer Score-Werte.⁹⁰ Absatz 2 der Vorschrift enthält ergänzende Bestimmungen über das Bonitäts-Scoring, also die Berechnung der Wahrscheinlichkeit von „Zahlungsfähig- und Zahlungswilligkeit einer natürlichen Person“.

Zunächst stellt die Vorschrift die Maßgeblichkeit des allgemeinen Datenschutzrechts klar. Jedes Scoring muss selbstverständlich den allgemeinen datenschutzrechtlichen Anforderungen genügen (§ 31 Abs. 1 Nr. 1 BDSG). Insbesondere müssen personenbezogene Daten, auf denen die Ermittlung von Score-Werten beruht, in Übereinstimmung mit Art. 6 Abs. 1 Satz 1 DSGVO erhoben worden sein und das Scoring-Verfahren muss den in Art. 5 Abs. 1 DSGVO niedergelegten Grundsätzen genügen. Das heißt, dass die Lektüre allein des § 31 BDSG ein äußerst unvollständiges Bild vermittelt über den Schutz des Verbrauchers beim Scoring. Die weitreichenden scoring-unspezifischen Schutzinstrumente des allgemeinen Datenschutzrechts sind auch beim Scoring zu beachten, etwa das grundsätzliche Erfordernis einer Einwilligung der betroffenen Person in die Erhebung der zur Score-Berechnung herangezogenen Daten (Art. 6 Abs. 1 Satz 1 Buchstabe a DSGVO).

87 Richtlinie 2013/36/EU des Europäischen Parlaments und des Rates vom 26. Juni 2013 über den Zugang zur Tätigkeit von Kreditinstituten und die Beaufsichtigung von Kreditinstituten und Wertpapierfirmen, zur Änderung der Richtlinie 2002/87/EG und zur Aufhebung der Richtlinien 2006/48/EG und 2006/49/EG.

88 Verordnung (EU) Nr. 575/2013 des Europäischen Parlaments und des Rates vom 26. Juni 2013 über Aufsichtsanforderungen an Kreditinstitute und Wertpapierfirmen und zur Änderung der Verordnung (EU) Nr. 646/2012.

89 Sehr indirekt wird die Qualität von Scoring-Dienstleistungen vom Gewerbeamt adressiert. § 38 GewO (Überwachungsbedürftige Gewerbe) sieht für Auskunfteien (Absatz 1 Satz 1 Nr. 2) eine generelle, anlassunabhängige Prüfung der Zuverlässigkeit des Gewerbetreibenden vor (siehe näher Schönleiter, 2009: § 38 GewO Rn. 4 f., 19 ff.) und intensiviert damit die behördliche Gewerbeaufsicht gegenüber dem Normalfall (§§ 1, 14, 35 GewO). Zum Verhältnis der von § 38 Abs. 1 Satz 1 Nr. 2 GewO einerseits und § 31 BDSG andererseits erfassten Regelungsgegenstände siehe Overbeck, 2016, S. 11 f.

90 Im Gesetzgebungsverfahren zu § 28b BDSG a.F. hatte der Bundesrat gefordert, die Vorschrift insgesamt auf Verträge zu beschränken, bei denen ein kreditorisches Ausfallrisiko besteht (BR-Drucks. 548/1/08, S. 11 f.); die Bundesregierung trat dem mit Erfolg entgegen, weil es für eine entsprechende Beschränkung des Anwendungsbereichs keine „sachliche Rechtfertigung“ gebe (BT-Drucks. 16/10581, S. 1); siehe Domurath & Neubeck, 2018.

Darüber hinaus macht § 31 BDSG zwei scoring-spezifische Arten von Vorgaben, denen die Berechnung von Scores genügen muss. Die Vorschrift formuliert einerseits normative Anforderungen an Scoring, andererseits qualitative.

Genuin normative Belange adressiert § 31 BDSG, soweit die Vorschrift auf die Frage antwortet, welche Daten in die Berechnung eines Scores einfließen dürfen. Zum Beispiel darf ein Bonitäts-Score nicht an die bloße Tatsache anknüpfen, dass ein Schuldner eine fällige Forderung nicht beglichen hat. Es müssen vielmehr besondere Umstände – etwa eine zivilgerichtliche Verurteilung zur Zahlung oder ein Anerkenntnis – hinzutreten, damit an die Nichtbegleichung einer Forderung angeknüpft werden darf. Weiterhin dürfen bestimmte Daten, nämlich die gesetzgeberisch für besonders sensibel erachteten Anschriftendaten, nur neben weiteren Daten für die Score-Wert-Berechnung verwendet werden, nicht hingegen allein genommen (siehe § 31 Abs. 1 Nr. 3 BDSG). Bestimmte Daten dürfen nur dann verwendet werden, wenn bei ihrer Gewinnung oder ihrer Heranziehung zur Score-Wert-Berechnung spezifizierte Verfahrens- und Formvorschriften eingehalten worden sind. Dies betrifft wiederum die Anschriftendaten (§ 31 Abs. 1 Nr. 4 BDSG), weiterhin – für das Bonitätsscoring – Daten über nichterfüllte Forderungen (§ 31 Abs. 2 Satz 1 Nr. 1 bis 5 BDSG). Daten, die nicht in Übereinstimmung mit den maßgeblichen Bestimmungen zur Datengewinnung erhoben worden sind, dürfen nicht zum Scoring verwendet werden, unabhängig davon, ob sie die Prognoseleistung eines Scoring-Verfahrens verbessern oder nicht. Alle diese Anforderungen haben damit nicht lediglich instrumentellen Charakter, um eine möglichst hohe Score-Güte zu gewährleisten. Vielmehr adressieren sie übergeordnete Gerechtigkeitsbelange. In diesen Anforderungen finden somit Wertvorstellungen über gesellschaftlich angemessene Formen des Scorings ihren Niederschlag.

Zweitens stellt § 31 BDSG Anforderungen an die Qualität des Verfahrens der Score-Wert-Berechnung. Hier antwortet die Vorschrift auf die Frage: Welchen Qualitätsanforderungen müssen prognostische Scoring-Verfahren genügen? Die normtextliche Grundlage für diese Qualitätsanforderungen ist § 31 Abs. 1 Nr. 2 BDSG. Die Bestimmung fordert, dass „die zur Berechnung des Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich sind“. Damit wird das prognostische Scoring-Verfahren auf einen empirisch nachvollziehbaren Rationalitätsstandard verpflichtet (näher unten E.III.1).

3.6 Regulierungspotenziale des Scorings durch § 31 BDSG

§ 31 BDSG hat das Potenzial, als Blaupause einer übergreifenden sachangemessenen Regulierung des Scorings zu dienen.⁹¹ Das betrifft nicht so sehr die normativen Vorgaben über die zulässigerweise heranziehbaren Daten. Hier ist die Vorschrift, auch in ihrem allgemein gefassten Absatz 1, erkennbar auf die Regulierung des Bonitäts-Scorings zugeschnitten. Die angeordneten Beschränkungen der Datenbasis sind durch den Blick auf diesen Wirtschaftsbereich gelenkt. Ein allgemeines Scoring-Recht ist diesen Teilen des § 31 BDSG nicht zu entnehmen. Anders verhält es sich mit der die Scoring-Qualität betreffenden Verpflichtung auf Wissenschaftlichkeit und Erheblichkeit. Hier birgt das geltende Recht den Nukleus eines allgemeinen „Produktsicherheitsrechts“ für Scoring-Verfahren. Dies setzt voraus, dass an die in § 31 Abs. 1 Nr. 2 BDSG enthaltenen Anforderungen weiter ausgebaut werden, als dies bisher geschehen ist (siehe weiter unten E.III.1).

⁹¹ Auch über den beschränkten Bereich von Scoring-Verfahren hinausgehend hat der Sachverständigenrat für Verbraucherfragen in seinem Gutachten „Verbraucherschutz 2.0“ das regulatorische Potenzial des § 31 BDSG (entspr. § 28b BDSG a.F.) herausgearbeitet und festgestellt, dass diese Regelung einen brauchbaren Ausgangspunkt für die Regulierung selbstlernender Algorithmen enthält (SVRV, 2016, S. 6 – dort die Lösungsoption II.1 – sowie zur Begründung S. 60 f.).

II. Bereichsspezifische Regelungen

Neben die allgemeinen datenverarbeitungsrechtlichen Vorgaben des Art. 22 DSGVO und des § 31 BDSG treten bereichsspezifische Regelungen. Sie adressieren bestimmte Sachprobleme, die das Scoring in einzelnen Wirtschaftsbereichen oder Vorgängen des Rechtslebens aufwirft. Allerdings sind diese Regelungen nicht in dem Bewusstsein geschaffen worden, Facetten oder spezielle Ausprägungen des Scorings zu regulieren. Sie fügen sich deshalb nicht in ein System der Scoring-Regulierung, sondern ergeben notwendigerweise ein fragmentiertes Gesamtpanorama. Die nachfolgenden Ausführungen sind deshalb nicht erschöpfend in dem Sinne, dass es keine weiteren rechtlichen Vorgaben gäbe, an denen die „Zuordnung eines Zahlenwertes zu einem Menschen zum Zwecke der Verhaltenssteuerung oder Verhaltensprognose“ produktiv gemessen werden könnte. Ausgeklammert bleibt beispielsweise die genuin grund- und menschenrechtliche Dimension des Scorings, obwohl Scoring geeignet ist, Persönlichkeitsrechte der gescorten Personen in erheblicher Weise zu beeinträchtigen. Hier gibt es offenkundig Forschungsbedarf (vgl. auch Gesellschaft für Informatik, 2018).

Regelungen vereinbart werden. Andere Bestimmungen können nach Absatz 1 Satz 2 in Verbindung mit Absatz 1 Satz 1 unwirksam sein.

Ein Scoring-System kann Bestandteil eines Vertrages zwischen Scorer und gescorter Person sein. Der ermittelte Score hat Einfluss auf die Rechte und Pflichten innerhalb des durch Scoring dynamisierten Vertragsverhältnisses. Ein anschauliches Beispiel hierfür ist ein Score-Wert, nach dem sich die Höhe eines Entgelts bemisst – man denke hier etwa an die Bestimmung einer Versicherungsprämie nach dem erfassten und in einen Score übersetzten Verhalten des Versicherten.

Wie jeder Vertragsinhalt haben solche Score-Klauseln – unter den gesetzlich bestimmten allgemeinen Umständen (siehe §§ 305, 310 BGB) – die Vorgaben des Rechts der Allgemeinen Geschäftsbedingungen zu beachten. Das in diesen Bestimmungen niedergelegte gestufte Kontrollregime zur Einbeziehung von Klauseln in den Vertrag, zu ihrer Auslegung und zu ihrer inhaltlichen Zulässigkeit enthält Instrumente, die, zumindest in bestimmten Hinsichten, Schutz vor unfairem Scoring gewährleisten können. Deshalb ist das Recht der Allgemeinen Geschäftsbedingungen ein wesentlicher Referenzpunkt, an dem die Zulässigkeit und Grenzen von Scoring-Klauseln erörtert werden (Rudkowski, 2017; Brömmelmeyer, 2017; Klimke, 2015).

Schutz davor, einer ungewollten Verstrickung des Kunden in ein Scoring-System auf vertraglicher Grundlage entgegenzuwirken, ergeben sich aus den Pflichten des AGB-Verwenders, auf die Allgemeinen Geschäftsbedingungen hinzuweisen und der anderen Partei die Möglichkeit zur Kenntnisnahme zu verschaffen (§ 305 Abs. 2 Nr. 1 und 2 BGB). Sind die Klauseln, die das Scoring-System begründen, als überraschend zu bewerten, werden sie nicht Vertragsbestandteil (§ 305c Abs. 1 BGB). Die faktische Bedeutung dieses Schutzinstruments dürfte derzeit als gering zu veranschlagen sein. Erstens setzen Unternehmen Scoring-Systeme heute als Mittel ein, um Kunden für die eigenen Produkte zu interessieren und sie an das Unternehmen zu binden (siehe C.III). Weiterhin bauen Scoring-Systeme auf einer technischen Infrastruktur auf, deren Einrichtung ohne Mitwirkung des Kunden kaum möglich sein dürfte. Eine „überraschende“ vertragliche Grundlegung des Scoring-Systems erscheint deshalb fernliegend.

1. Recht der Allgemeinen Geschäftsbedingungen

i

§ 307 BGB. Inhaltskontrolle.

(1) Bestimmungen in Allgemeinen Geschäftsbedingungen sind unwirksam, wenn sie den Vertragspartner des Verwenders entgegen den Geboten von Treu und Glauben unangemessen benachteiligen. Eine unangemessene Benachteiligung kann sich auch daraus ergeben, dass die Bestimmung nicht klar und verständlich ist.

(2) (...)

(3) Die Absätze 1 und 2 sowie die §§ 308 und 309 gelten nur für Bestimmungen in Allgemeinen Geschäftsbedingungen, durch die von Rechtsvorschriften abweichende oder diese ergänzende

Ein wirkungsvolles Instrument zur Regulierung des Scorings in Hinblick auf den Belang der Transparenzsicherung enthält § 307 Abs. 1 Satz 2 BGB. Die Vorschrift gebietet eine Inhaltskontrolle Allgemeiner Geschäftsbedingungen am Maßstab der Frage, ob diese den Vertragspartner des AGB-Verwenders unangemessen benachteiligen. Eine solche unangemessene Benachteiligung kann sich bereits daraus ergeben, dass die betreffende Vertragsbedingung nicht „klar und verständlich“ ist (eingehend zu den Dimensionen dieses Gebots Micklitz, 2014). Die Vorschrift bezieht sich – anders als die inhaltliche Kontrolle von AGB-Klauseln im Übrigen – auch auf den sogenannten kontrollfreien Bereich, das heißt auf die Hauptleistungsversprechen und Preisklauseln (Wurmnest, 2016, § 307 BGB Rn. 12 ff., 16 ff.). Für das Versicherungsrecht erlaubt § 307 Abs. 1 Satz 2 in Verbindung mit Abs. 3 Satz 2 BGB damit etwa die Kontrolle der Vertragsklausel, die bestimmt, dass die Prämie oder ein auf die Prämie gewährter Rabatt überhaupt variabel ausgestaltet ist (Rudkowski, 2017).

Schließlich eröffnet das Recht der Allgemeinen Geschäftsbedingungen über § 307 Abs. 1 Satz 1, Abs. 2 BGB (sowie über § 308 und § 309 BGB) die Möglichkeit, Vertragsklauseln unter dem Gesichtspunkt ihrer inhaltlichen Ausgewogenheit zu kontrollieren. Hiervon erfasst sind etwa Klauseln in Versicherungsverträgen, die die Anpassungsmodalitäten von Prämien definieren; diese fallen nicht in den nach § 307 Abs. 3 Satz 1 BGB kontrollfreien Bereich (Rudkowski, 2017). § 307 Abs. 1 Satz 1 BGB verlangt eine Abwägung der Interessen des Verwenders einerseits und der Gegenpartei andererseits (Wurmnest, 2016, § 307 BGB Rn. 33). In dieser Abwägung können scoring-spezifische Individualbelange der gescorten Person Berücksichtigung finden. So dürfte etwa das Interesse an der Durchführung von Scoring-Verfahren, die unter Qualitätsaspekten beanstandungswürdig sind, vergleichsweise gering zu gewichten sein. Weiterhin kann auch die Breite der erfassten Datenbasis, die zur Ermittlung eines Score-Wertes herangezogen wird, ein maßgeblicher Belang sein. Die Beeinträchtigung des Allgemeinen Persönlichkeitsrechts der gescorten Person fällt umso intensiver aus, je umfassender Daten über sie erhoben werden (vgl. Rudkowski, 2017).

2. Versicherungsvertragsrecht und Versicherungsaufsichtsrecht

Der Versicherungsvertrag ist ein vergleichsweise intensiv reguliertes „Rechtsprodukt“ (Dreher, 1991) und die Versicherungswirtschaft ein engmaschig beaufsichtigter und in erheblichem Maße durch EU-Vorgaben strukturierter Wirtschaftszweig (Überblick bei Dreher, 2018, Einl.). Die oft erhebliche, mitunter sogar existenzielle Bedeutung des Versicherungsvertrages für den Versicherten und die Systemrelevanz der Versicherungsunternehmen für die Stabilität und Funktionsfähigkeit der Finanzmärkte legitimieren grundsätzlich ein strenges versicherungsrechtliches Rechtsregime.

Die Frage nach der rechtlichen Zulässigkeit von Scoring-Verfahren „in der Versicherung“ muss differenziert beantwortet werden. Ihre Antwort hängt ab vom jeweils in den Blick genommenen Versicherungszweig, denn deren rechtliche Grundlagen unterscheiden sich erheblich.

Zu beachten ist weiterhin, dass Scoring-Methoden innerhalb eines Versicherungszweiges in ganz unterschiedlicher Weise in das Versicherungsvertragsverhältnis integriert werden können (vgl. Deutsche Aktuarvereinigung e.V., 2017). Die dynamische Anpassung von Prämien in Abhängigkeit von Score-Werten ist hier zwar das anschaulichste Beispiel, aber derzeit nicht das praxisrelevanteste (siehe C.III.1).

Die Anpassung von Prämienrabatten, die Bemessung der Höhe von Überschussbeteiligungen oder von Beitragsrückerstattungen sowie die Bestimmung von Vergünstigungen bei Dritten sind weitere Bereiche, in denen Scoring-Verfahren zum Einsatz kommen können. Insbesondere kann auch die Individualisierung der Versicherungsvertragsinhalte zum Abschluss des Versicherungsvertrages als Scoring aufgefasst werden; die Berücksichtigung etwa des Alters des Versicherten hat eine lange Tradition. Eine pauschale Beurteilung der versicherungsrechtlichen Zulässigkeit des Scorings verbietet sich deshalb und würde der Komplexität des Wirtschaftsbereiches nicht im Ansatz gerecht werden (Bitter & Uphues, 2017).

Allgemeine versicherungsaufsichtsrechtliche Regelungen stehen einer Einführung von Scoring-Elementen in der Versicherung nicht fundamental entgegen. Ihnen kann durch eine angemessene Gestaltung der Wirtschaftstätigkeit der Versicherungsunternehmen Rechnung getragen werden. So ist die Überlassung der für die Durchführung von Scoring erforderlichen Telematik-Komponenten grundsätzlich nicht als versicherungsfremdes und deshalb am Maßstab des § 15 VAG gemessen unzulässiges Geschäft zu qualifizieren (Klimke, 2015). Das Verbot von Sondervergütungen nach § 48b VAG ist zu beachten, steht aber Scoring-Elementen im Versicherungsverhältnis nicht kategorisch entgegen. Die sonderdatenschutzrechtliche Regelung des § 213 Abs. 1 VVG zählt die Stellen auf, bei denen der Versicherer Gesundheitsdaten über den Versicherungsnehmer zu erheben berechtigt ist: Ärzte, Krankenhäuser und sonstige Krankenanstalten, Pflegeheime und Pflegepersonen, anderen Personenversicherer und gesetzliche Krankenkassen sowie Berufsgenossenschaften und Behörden. Die Anbieter von Wearables oder Gesundheits-Apps (dazu bereits Adam & Micklitz, 2016) für das Smartphone sind hier nicht genannt. Die Vorschrift dürfte gleichwohl kein gravierendes Hindernis für die Etablierung von Scoring-Elementen in der Krankenversicherung darstellen, denn erstens ist die Vorschrift abdingbar (Rixecker, 2016, § 213 VVG Rn. 28) und zweitens lässt sich plausibel vertreten, dass eine Datenerhebung im Falle des Zugriffs des Versicherers auf die durch Wearables oder Apps erzeugten Gesundheitsdaten nicht bei Dritten erfolgt, sondern beim Versicherten.

Das vorliegende Gutachten betrachtet intensiv die Kfz-Versicherung und die Krankenversicherung als mögliche Arenen des Scorings.⁹² In den Blick rücken damit zwei Endpunkte eines regulatorischen Spektrums. Für die Kfz-Versicherung legt das geltende Recht der Integration von Scoring-Elementen kaum versicherungsrechtsspezifische Hindernisse in den Weg. Die private Krankenversicherung ist dagegen wegen ihrer existentiellen Bedeutung für die Versicherten (Schüffner & Franck, 2018, § 47 Rn. 118) und ihrer sozialpolitischen Relevanz in ein enges regulatorisches Korsett geschnürt.

Für eine Nutzung von Scoring-Verfahren während eines laufenden Versicherungsverhältnisses lässt das geltende Recht hier wenig Raum.

Die Kfz-Versicherung ist in der Einführung von Scoring-Verfahren weitestgehend frei. Das Leitprinzip eines solidarischen Schadensausgleichs im Kollektiv ist dieser Versicherung fremd. Die Versicherung ist bestimmt durch das versicherungstechnische Äquivalenzprinzip, das Prämie und übernommenes Risiko in Beziehung zueinander setzt. Einen faktischen „Solidaritätseffekt“ mag man in dem Umstand erkennen, dass innerhalb einer Versicherungsperiode Schäden notwendigerweise nur bei manchen Personen aus dem in einer bestimmten Risikogruppe zusammengefassten Versichertenkollektiv eintreten, während andere verschont bleiben. Hierbei handelt es sich aber „um einen Ausgleich zufälliger Schwankungen im Schadenverlauf sowie in der Zeit und nicht um den Ausgleich systematischer Unterschiede in der Schwere der individuellen Risiken“ (Bitter & Uphues, 2017, S. 3 f.). Eine Intensivierung erfährt dieser Solidaritätseffekt dadurch, dass die technischen Möglichkeiten zur aktuarischen „Fragmentierung“ (Looschelders, 2015) der Versichertengesamtheit in einzelne Risikogruppen nicht völlig ausgereift sind. Dadurch kommt es zu einem Ausgleich zwischen Personen mit höherer und solchen mit geringerer Schadenswahrscheinlichkeit, die in einer Risikogruppe zusammengefasst sind.

Vor diesem Hintergrund ist es nicht überraschend, dass für eine Einführung von fahrverhaltensbezogenen Telematik-Tarifen keine strukturellen rechtlichen Hindernisse gesehen werden (Klimke, 2015; Schumann, 2017; Koch, 2017, K. AKB 2015 Rn. 9 ff.). Als problematisch könnte sich allerdings die Vereinbarkeit von Telematiktarifen mit den versicherungsvertragsrechtlichen Bestimmungen zur Gefahrerhöhung (§§ 23 ff. VVG) erweisen (näher Lüttringhaus, 2018). Der Verwendung der Telematik-Daten in der rechtlichen Bewältigung des eingetretenen Versicherungsfalls stehen keine grundsätzlichen Bedenken entgegen (Klimke, 2015).

⁹² Zum Scoring (ohne Verwendung dieses Begriffs) in der Lebens- und Berufsunfähigkeitsversicherung siehe Brömmelmeyer, 2017.

„Die“ private Krankenversicherung ist ein hochdifferenziertes Feld unterschiedlicher Versicherungsarten (Vollversicherung und verschiedene Zusatzversicherungen, siehe auch § 192 VVG), die auf unterschiedlichen kalkulatorischen Grundlagen fußen (Prämienkalkulation nach Art der Lebensversicherung und nach Art der Schadenversicherung, siehe Kalis, 2018, § 44 Rn. 209 ff.) und die, in unterschiedlichem Ausmaß, eine auch sozialpolitische Funktion erfüllen. Auf der Hand liegt die sozialpolitische Bedeutung bei der substitutiven Krankenversicherung (§ 146 VAG, § 195 VVG), die an die Stelle der gesetzlichen Krankenversicherung treten kann und deshalb besonderen Vorgaben genügen muss. Generelle Aussagen über die Zulässigkeit von Scoring stoßen hier mehr noch als im Versicherungsrecht im Übrigen auf eine große Vielfalt denkbarer Versicherungsprodukte und Einsatzbereiche für Scoring-Methoden. Im Herzen der Scoring-Debatte steht die Vorstellung einer Zukunft, in der Versicherungsprämien an das gescorte Gesundheitsverhalten geknüpft sind. Jedenfalls hier errichten das geltende Versicherungsvertrags- und Versicherungsaufsichtsrecht wirksame Barrieren.

Wenn die Krankenversicherungsprämie nach Art der Lebensversicherung berechnet wird – dies ist eine Voraussetzung dafür, dass sie als substitutive Krankenversicherung fungieren kann (s. § 146 Abs. 1 VAG) –, steht einer auf Scoring basierenden Prämienanpassung § 203 VVG entgegen.⁹³ Der erste Absatz der Vorschrift bestimmt abschließend (Voit, 2018, § 203 VVG Rn. 6), nach welchen Kriterien die Prämie zu bemessen ist. Der zweite Absatz der Vorschrift bestimmt in Verbindung mit § 208 Satz 1 VVG die Voraussetzungen und das Verfahren der Prämienanpassung. Diese Vorschrift ist halbzwingend (Boetius, 2017, § 203 VVG Rn. 51), das heißt, dass von ihr nicht zulasten des Versicherungsnehmers abgewichen werden kann. § 203 VVG gestattet es weder bei der Prämienbestimmung noch der Prämienanpassung, das individuelle Gesundheitsverhalten bzw. einen auf diesem basierenden Score zur Grundlage der Prämiengestaltung zu machen.⁹⁴ Denn § 203 Abs. 1 Satz 1 und Abs. 2 Satz 4 VVG nehmen das Prämienbemessungs-

recht des Versicherungsaufsichtsgesetzes in Bezug. Die Bestimmungen stellen insoweit einen Gleichlauf der Rechtsgebiete her und transformiert die aufsichtsrechtlichen Maßgaben in versicherungsvertragsrechtliche (Boetius, 2017, § 203 VVG Rn. 3 und 42). Versicherungsaufsichtsrechtlich enthält § 6 Abs. 1 und § 10 Abs. 1 KVAV Maßgaben zur Prämienkalkulation in der privaten Krankenversicherung. Die Bestimmung schließt aus, das Gesundheitsverhalten während der Vertragslaufzeit als prämiensbildenden Faktor zu berücksichtigen. Bei der Prämienbemessung zu berücksichtigen sind das Alter und der Leistungsumfang der Krankenversicherung (siehe aber zu den Besonderheiten des Basistarifs § 152 VAG und dazu Voit, 2018, § 203 VVG Rn. 11 ff.). Vorerkrankungen können im Wege des Risikozuschlags in der Prämienkalkulation berücksichtigt werden (Voit a.a.O., Rn. 6); davon abgesehen besteht für die Prämienindividualisierung kein Raum.

Zu beachten ist weiterhin § 194 Abs. 1 Satz 2 VVG. Die Vorschrift ordnet an, dass die allgemeinen Bestimmungen über die Gefahrerhöhung (§§ 23 ff. VVG) keine Anwendung finden. Dieser Ausschluss der Anwendbarkeit betrifft nicht nur die Krankenversicherungen mit Prämienkalkulation nach Art der Lebensversicherung, deren Prämienbemessung sich ohnehin nach Maßgabe des § 203 VVG vollzieht (Kalis, 2017, § 194 VVG Rn. 24), sondern auch die Krankenversicherung nach Art der Schadenversicherung. „Nach Vertragsschluss eintretende Änderungen im Gesundheitszustand des Versicherten lassen das einmal erteilte Leistungsversprechen ebenso unberührt wie die Höhe der Prämie“ (Kalis a.a.O.). Das Risiko nachträglicher Gefahrerhöhungen trägt der Versicherer und nicht der Versicherungsnehmer (s. BVerwG, Urteil vom 5. März 1999 – 1 A 1/97; BGH, Urteil vom 9. Mai 2012 – IV ZR 1/11).

Plausibel lässt sich weiterhin annehmen, dass die Regelung in § 146 Abs. 2 Satz 1 in Verbindung mit § 138 Abs. 2 VAG bestimmten score-basierten Tarifen entgegensteht, wenn diese sich nicht selbst tragen und damit

⁹³ Allerdings hängt die Anwendbarkeit des § 203 VVG nicht davon ab, dass die betreffende private Krankenversicherung tatsächlich die Voraussetzungen des § 146 VAG erfüllt, siehe Voit, 2018, § 203 VVG Rn. 5.

⁹⁴ Siehe aber die Individualisierungsmöglichkeit nach § 203 Abs. 1 Satz 2 VVG, wonach außer „bei Verträgen im Basistarif nach § 152 des Versicherungsaufsichtsgesetzes (...) der Versicherer mit Rücksicht auf ein erhöhtes Risiko einen angemessenen Risikozuschlag oder einen Leistungsausschluss vereinbaren“ kann.

indirekt zulasten der im üblichen Tarif Versicherten gehen (Brömmelmeyer, 2017).

Grundsätzlich möglich ist eine Bemessung von Beitragsrückerstattungen nach Maßgabe von Scoring-Elementen. Hierbei handelt es sich um eine Fortentwicklung jener versicherungsvertraglichen Regelungen, die eine erfolgsunabhängige Beitragsrückerstattung von der Nichtinanspruchnahme von Versicherungsleistungen in einem vergangenen Abrechnungszeitraum abhängig machen (Boetius, 2017, § 203 WG Rn. 322; zu den restriktiveren Voraussetzungen in der gesetzlichen Krankenversicherung siehe § 53 Abs. 2 und 9 SGB V). Die innere Konsistenz eines in diesem Sinne zweigeteilten Regelungsregimes mit strikten Vorgaben zur Prämienanpassung und vergleichsweise weichen Möglichkeiten zur Berücksichtigung von Scores im Übrigen kann man kritisch beurteilen. Denn bei wirtschaftlicher Betrachtungsweise kann sich die Beitragsrückerstattung als ein Äquivalent zur Prämienindividualisierung darstellen.

3. Sozialversicherungsrecht der gesetzlichen Krankenversicherung

Die gesetzliche Krankenversicherung ist eine in Selbstverwaltung organisierte öffentliche Einrichtung, die sich auf das grundgesetzliche Sozialstaatsprinzip gründet. Gemeinsam mit den übrigen Zweigen der Sozialversicherung dient sie der Absicherung gegenüber elementaren Lebensrisiken und ist somit ein Instrument der Sozialpolitik. Ihr grundlegendes Strukturmerkmal ist das Solidarprinzip (Kingreen, 2003). Die Mitgliedschaft in der gesetzlichen Krankenversicherung ist für zahlreiche Personengruppen als eine auf Gesetz beruhende Pflichtmitgliedschaft ausgestaltet. Die Beiträge zur gesetzlichen Krankenversicherung bemessen sich grundsätzlich nach der wirtschaftlichen Leistungsfähigkeit (§ 241 SGB V), die Leistungen nach dem individuellen medizinischen Bedarf (Butzer, 2001; Kingreen, 2003). Die Bemessung der Beiträge nach dem individuellen Risiko des Versicherten ist weder bei Begründung der Mitgliedschaft noch im laufenden Mitgliedschaftsverhältnis zulässig.

Für an das Gesundheitsverhalten anknüpfende individualisierte Anreizsysteme im Wege des Scorings bleibt in einem solchen System wenig Raum (vgl. Landessozialgericht Hessen, Urteil vom 4. Dezember 2008 – L 1 KR 150/08 KL -, Juris.). Derzeit gibt es dementsprechend auch keine komplexen Scoring-Verfahren in diesem Bereich.⁹⁵ Allerdings stellen die in § 65a SGB V vorgesehenen Bonusprogramme für individuelles gesundheitsförderliches Verhalten (Bundesversicherungsamt, 2016) ein System dar, das man als Proto-Scoring bezeichnen sollte. Innerhalb der gesetzlichen Krankenversicherung handelt es sich hierbei um eine besonders legitimationsbedürftige Ausnahme. Das Bundesversicherungsamt beurteilt sie insgesamt skeptisch (Bundesversicherungsamt, 2018; siehe auch C.III.3).

⁹⁵ Nichts gesagt ist damit zu den Grenzen, die das Verfassungsrecht einer Umgestaltung des Sozialversicherungsrechts weg vom Solidarprinzip und hin zu einer stärker akzentuierten Individualisierung zieht. Zwar fordert das grundgesetzliche Sozialstaatsprinzip (Art. 20 Abs. 1, Art. 28 Abs. 1 Satz 1 GG) eine Schaffung sozialer Sicherungssysteme gegen die Wechselfälle des Lebens (BVerfGE 28, 324 [348 ff.], 45, 376 [387], 68, 193 [209]; siehe dazu Axer, 2000), eine Bindung an die – ohnehin vorkonstitutionellen – Strukturprinzipien der gesetzlichen Krankenversicherung ergibt sich daraus allerdings nicht.

III. Bausteine eines Regulierungsrechts des Scorings

Die Durchsicht datenverarbeitungsrechtlicher und bereichsspezifischer Regelungen mit Bezug zum Scoring hat gezeigt, dass sie zwar kein allgemeines Scoring-Recht ausformen, aber zum Phänomen Scoring nicht berührungslos stehen. Das geltende Recht enthält Instrumente, mit denen bestimmte mit dem Scoring verbundene soziale Herausforderungen bewältigt werden können.

Im Folgenden soll für drei mit dem Scoring verbundene Herausforderungen – der Score-Güte, der Transparenz des Scoring-Verfahrens und des Diskriminierungsverbotes – skizziert werden, wie weit das geltende Recht Vorgaben macht. Weil die beschriebenen Herausforderungen nicht beschränkt sind auf einzelne Wirtschaftsbereiche (etwa die Krankenversicherung) oder besondere Situationen des Rechtslebens (etwa den Verbrauchervertragsschluss), soll dabei das Hauptaugenmerk auf jenen Regelungen liegen, welche die identifizierten Probleme in möglichst allgemeiner Form adressieren.

Beim Scoring handelt es sich um einen Datenverarbeitungsvorgang, so dass die Vorgaben der Datenschutz-Grundverordnung und des Bundesdatenschutzgesetzes besonders ergiebig sind, um regulatorische Ziele auf breiter Front zu verwirklichen. Bezogen auf den Schutz vor Diskriminierung ist das Allgemeine Gleichbehandlungsgesetz in den Blick zu nehmen. Dieses ist zwar in seinem Anwendungsbereich sachlich beschränkt (§§ 2, 19 AGG) und offenbart in vielen legislativen Details den Kompromisscharakter des letztlich verwirklichten Diskriminierungsschutzes. Allerdings enthält es nicht nur Legaldefinitionen der unmittelbaren und mittelbaren Benachteiligung (§ 3 Abs. 1 und 2 AGG), sondern eignet sich auch als Modell, anhand dessen das allgemeine Prüfprogramm zur Feststellung rechtlich verbotener Ungleichbehandlungen verdeutlicht werden kann.

1. Regulierung des „Wie“ versus Regulierung des „Ob“ des Scorings

Die eingangs genannten Herausforderungen der mathematisch-statistischen Güte, der Transparenz und des Diskriminierungsverbotes betreffen das Phänomen Scoring, *wenn es stattfindet*. Nicht weiter vertieft wird demgegenüber die vorgelagerte Frage danach, in welchen Rechtsbeziehungen überhaupt gescort werden darf. Die Frage nach dem statthaften Anwendungsbereich für Scoring-Verfahren kann nur differenziert nach Lebens- und Wirtschaftsbereichen beantwortet werden.

Die oben skizzierten Regelungsregime des Scorings im Bereich des Versicherungswesens zeigen auf, dass bereichsspezifische Vorschriften einer Anwendung von Scoring-Methoden entgegenstehen können. Ein solches Scoring-Hindernis kann sich dadurch ergeben, dass sie die Kriterien, nach denen eine Privatperson handeln und entscheiden darf, abschließend festgelegt sind – man denke hier an die oben skizzierten Vorgaben zur Bemessung von Prämien(änderungen) in der privaten Krankenversicherung. Das Recht kann indessen Scoring auch dadurch verhindern, dass es eine Heranziehung gerade von den Erkenntnissen, die aus Scoring-Verfahren gewonnen worden sind, untersagt. In den in diesem Gutachten betrachteten Wirtschaftsbereichen gibt für ein entsprechendes Regelungsmodell keine gesetzlichen Beispiele. Es steht jedoch zu erwarten, dass mit den fortschreitenden technischen Möglichkeiten zum Einsatz von Scoring-Methoden und ihrer wirtschaftlichen Verbreitung der Gesetzgeber oder (etwa in Durchsetzung der Horizontalwirkung des Allgemeinen Persönlichkeitsrechts) die Rechtsprechung einer Verwendung spezifisch von Scoring-Methoden in bestimmten Fällen grundsätzlich einen Riegel vorschieben wird.

2. Scoring-Regulierung und Algorithmen-Regulierung

Die Regulierung des „Wie“ des Scorings fügt sich ein in den Kontext der Regulierung von Algorithmen, deren Möglichkeiten derzeit intensiv diskutiert werden (siehe oben B.I.4). Demzufolge hat der SVRV bereits in seinem Gutachten „Verbraucherrecht 2.0“ ausgeführt, dass die Scoring-Vorschrift im Bundesdatenschutzgesetz⁹⁶ einen legislativen Ausgangspunkt für die Regulierung von Algorithmen enthält (SVRV, 2016; ebenso Härting, 2015). Hieran knüpfen die nachfolgenden Überlegungen an.

Am Sonderfall der Scoring-Regulierung kann aufgezeigt werden, wie das Recht qualitative Anforderungen (für sie steht exemplarisch die Gütegewährleistung) und ethische Anforderungen (dafür steht exemplarisch das Diskriminierungsverbot) an Algorithmen formulieren und durchsetzen kann. Die für die Erstattung dieses Gutachtens eingeholte Studie der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V. über „Technische und rechtliche Betrachtungen algorithmischer Entscheidungsfindung“ (Gesellschaft für Informatik, 2018) ist vor diesem Hintergrund mittragendes Element der nachfolgenden Ausführungen (vgl. auch Gigerenzer, Wagner & Müller 2018).

3. Gewährleistung definierter Score-Güte

Zahlreiche Scoring-Verfahren dienen dem Zweck, eine bestimmte Prognoseleistung zu erbringen. Der Score-Wert ist das Urteil über die Wahrscheinlichkeit eines zukünftigen Verhaltens eines Menschen. Diese Leistung kann ein Scoring-Verfahren besser oder schlechter erfüllen. Prognostische Scoring-Verfahren haben also eine

„Gütedimension“ (zur tatsächlichen Lage siehe C.III.3). Der Gesetzgeber verhält sich zu der Frage der Scoring-Güte zwar nicht indifferent, die in § 31 Abs. 1 Nr. 2 BDSG niedergelegten „Wissenschaftlichkeits“-Anforderungen gewährleisten allerdings nur ein Mindestmaß an Güte.

§ 31 Abs. 1 Nr. 2 BDSG bestimmt, dass „die zur Berechnung des Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich“ sein müssen. Damit adressiert die Vorschrift zwei⁹⁷ unterschiedliche Probleme, die beide ihren normtextlichen Anker in der Wendung der „genutzten Daten“ finden. Denn mit den „genutzten Daten“ können entweder die *allgemein* für die Score-Wert-Berechnung herangezogenen Datenarten gemeint sein („Alter“, „Anzahl der Girokonten“, „Adresse“). Oder es kann sich bei den „genutzten Daten“ um die für eine konkrete Score-Wert-Berechnung über eine bestimmte Person herangezogene Datengrundlage („50 Jahre“, „3“, „Schlossstraße 1“) handeln. Die regulatorischen Mittel, um einerseits Mängeln der einzelfallunspezifischen Scoring-Formel und andererseits den einzelfallbezogenen Input-Daten zu begegnen, unterscheiden sich.

3.1 Richtigkeit der konkreten Input-Daten

Das Anliegen, dass die dem Scoring zugrunde gelegten Daten die wahre Sachlage nicht zutreffend beschreiben (siehe oben B.V.), zielt auf die Gewährleistung einer hohen Datenqualität. Das Problem ist von großer praktischer Relevanz (siehe C.III.3). Tatsächlich dürfte es sich bei der Fehlerhaftigkeit personenbezogener Daten der gescorten Person um *das* zentrale Alltagsproblem im Hinblick auf Beschwerden handeln (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014, mit Berichten mehrerer Verbraucherzentralen über das Bonitäts-Scoring), wofür auch die Schwerpunkte der Informationstätigkeit mehrerer Verbraucherzentralen über das Bonitäts-Scoring sprechen (siehe exemplarisch Verbraucherzentrale Bre-

⁹⁶ Damals bezogen auf § 28b BDSG a.F., der mit dem hier behandelten, am 25. Mai 2018 in Kraft getretenen § 31 BDSG n.F. strukturgleich ist.

⁹⁷ Darüber hinaus ist die Voraussetzung der Nachweisbarkeit von eigenständiger Bedeutung, statuiert sie doch eine Dokumentationspflicht des Scorers, siehe BT- Drucks. 16/10529, S. 16 und Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe, 2014. Dieser Aspekt bleibt im Folgenden außer Betrachtung.

men, 2016; Verbraucherzentrale Niedersachsen, 2015; Verbraucherzentrale Nordrhein-Westfalen, 2018).

Die Gewährleistung der Verarbeitung lediglich richtiger personenbezogener Daten betrifft allerdings nicht nur das Scoring, sondern grundsätzlich jeden Datenverarbeitungsvorgang. Das Problem deshalb im Wesentlichen durch das in § 31 Abs. 1 Nr. 2 BDSG niedergelegte Verbot der Verarbeitung unrichtiger Daten (und deshalb zugleich unerheblicher Daten im Sinne des § 31 Abs. 1 Nr. 2 BDSG, siehe Overbeck, 2016) angehen zu wollen, erscheint als ein verengter und deshalb nicht sachgerechter Problemzuschnitt. Perspektivisch wird es eines Qualitätsgewährleistungsrechts für Datenbestände bedürfen. Eine normative Verankerung für ein solches Rechtsregime findet sich bereits heute im Grundsatz der Datenrichtigkeit (Art. 5 Abs. 1 Buchstabe d DSGVO, siehe dazu Pötters, 2018, Art. 5 DSGVO Rn. 24). Die Konturen dieses Rechtsgebiets und der Pflichten, die die Verantwortlichen in Hinblick auf die Richtigkeit der von ihnen verarbeiteten Daten zu erfüllen haben, sind derzeit indessen noch kaum entwickelt (Hoeren, 2016). Ein rechtlich einfaches Mittel der Problembewältigung sind sicher die Auskunfts- und Korrekturansprüche des Betroffenen (Art. 16 Satz 1 DSGVO, siehe weiter Domurath & Neubeck, 2018). Insoweit leidet das Datenschutzrecht allerdings an einem erheblichen Mobilisierungsdefizit (Härting, 2015; Spindler, Thorun & Wittmann, 2017).

3.2 Wissenschaftlichkeit des Scoring-Verfahrens

§ 31 Abs. 1 Nr. 2 BDSG gebietet, dass Scoring-Verfahren bestimmten wissenschaftlichen Standards genügen müssen (siehe bereits oben B.IV.1). Der Gesetzgeber schließt damit jedenfalls eine Heranziehung solcher Datenarten zum Zwecke des Scorings aus, die zur Prognoseleistung eines Scoring-Verfahrens nichts beitragen können (Domurath & Neubeck, 2018). Wo nicht einmal eine Korrelation aufgezeigt werden kann zwischen einer bestimmten Datenart und dem Ereignis, dessen Eintrittswahrscheinlichkeit ermittelt werden soll, steht § 31 Abs. 1 Nr. 2 BDSG der Verwendung dieser Datenart entgegen.

Mitunter wird der Versuch unternommen, über den Nachweis einer Korrelation hinausgehende Anforderungen an die Erheblichkeit der verwendeten Daten zu

konturieren. Diese Anforderungen so zu formulieren, dass sie in der praktischen Rechtsanwendung handhabbar sind, erweist sich als ein schwieriges Unterfangen. Das gilt zum Beispiel für die mitunter vertretene Beschränkung der zulässigerweise heranziehbaren Datenarten auf solche, die „vertragsrelevant“ sind (Nachweise bei Domurath & Neubeck, 2018). Dabei soll es sich um solche Datenarten handeln, die in besonders „unmittelbarer“ und „direkter“ Weise auf die Wahrscheinlichkeit des betreffenden Verhaltens einen Einfluss haben (siehe auch Buchner, 2018, § 31 BDSG Rn. 8: Die Vorschrift verpflichtet dazu, „plausibel und nachvollziehbar zu belegen, dass die verwendeten Daten für die Scorewert-Berechnung von unmittelbarer Relevanz sind.“). Wie der Kreis dieser Datenarten jeweils verlässlich zu bestimmen ist, bleibt unklar.

Das in § 31 Abs. 1 Nr. 2 BDSG niedergelegte „Korrelations“-Erfordernis zieht es nach sich, dass sich derjenige, der ein prognostisches Scoring-Verfahren betreibt, bei dessen Gestaltung nicht auf statistisch ungeprüfte Alltagsplausibilitäten verlassen darf. Das Scoring bedarf insoweit einer „empirischen Härtung“. Diese Anforderung ist durchaus nicht selbstverständlich. Denn ein allgemeines Gebot, sich nur rational fundiert zu äußern, besteht für Träger des Grundrechts der Meinungsfreiheit (Art. 5 Abs. 1 Satz 1 GG) auch dann nicht, wenn sie Tatsachenbehauptungen kommunizieren. Insoweit stellt sich das Rationalitätserfordernis des § 31 Abs. 1 Nr. 2 BDSG bereits als ein durchaus erheblicher gesetzgeberischer Eingriff dar, der allerdings mit Blick auf die soziale Erheblichkeit (siehe oben A.) kommunizierter Wahrscheinlichkeits-Scores plausibel rechtfertigen lässt.

Dass ein Verfahren wissenschaftlich im Sinne des § 31 Abs. 1 Nr. 2 BDSG ist, gewährleistet aber für sich genommen nicht, dass es seine Prognoseleistung mit einer Zuverlässigkeit erbringt, die für den jeweiligen Lebensbereich angemessen ist. Denn jedes Verfahren, das eine bessere Prognoseleistung erbringt als ein Münzwurf, kann Ergebnis der fachgerechten Anwendung statistischer Methoden sein und als solche eine beachtliche und beifallswürdige wissenschaftliche Leistung darstellen. Ob ein Verfahren aber in einem bestimmten Lebensbereich angewandt werden sollte oder nicht, ist damit noch nicht entschieden. Bestimmte Qualitätskriterien sind mit der Verpflichtung auf Wissenschaftlichkeit nicht verbunden. Hier enthält das Rechtsregime des prognos-

tischen Scoring eine regulatorische Leerstelle, die etwa in Kontrast mit dem oben skizzierten Recht der Eigenmittelausstattung von Kreditinstituten (siehe E.I.3.4) besonders markant sichtbar wird. § 31 Abs. 1 Nr. 2 BDSG ist damit zwar kein zahnlöser Tiger, aber doch einer mit Beißhemmungen.

4. Gewährleistung von Transparenz und Verständlichkeit

Die Datenschutz-Grundverordnung erklärt den Belang der „Transparenz“ ausdrücklich zu einem Grundsatz, dem jede Verarbeitung personenbezogener Daten zu genügen hat. Art. 5 Abs. 1 Buchstabe a Var. 3 DSGVO bestimmt, dass personenbezogene Daten in einer „für die betroffene Person nachvollziehbaren Weise verarbeitet werden“ müssen. Programmatisch entfaltet wird dieses Transparenzprinzip in den Erwägungsgründen 39, 58 und 60 zur Datenschutz-Grundverordnung. Der ausschweifende Normtextbefund darf nicht darüber hinwegtäuschen, dass die Abstraktionshöhe des Transparenzgrundsatzes beträchtlich bleibt. Welche genauen Pflichten den Verantwortlichen in Bezug auf je konkrete Datenverarbeitungsvorgänge tatsächlich treffen (Art. 5 Abs. 2 DSGVO), bleibt in hohem Maße unsicher (siehe bereits oben vor E.I.1 sowie etwa Roßnagel, 2018; Wachter, Mittelstadt & Floridi, 2017; Selbst & Powles, 2017). Eine gewisse Konkretisierung des Pflichtenprogramms leisten die Art. 12 bis 15 DSGVO.

i

Art. 12 DSGVO. Transparente Information, Kommunikation und Modalitäten für die Ausübung der Rechte der betroffenen Person

Der Verantwortliche trifft geeignete Maßnahmen, um der betroffenen Person alle Informationen gemäß den Artikeln 13 und 14 und alle Mitteilungen gemäß den Artikeln 15 bis 22 und Artikel 34, die sich auf die Verarbeitung beziehen, in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache zu übermitteln; dies gilt insbesondere für Informationen, die sich speziell an Kinder richten. (...)

i

Art. 13 DSGVO. Informationspflicht bei Erhebung von personenbezogenen Daten bei der betroffenen Person (ähnlich Art. 14 DSGVO. Informationspflicht, wenn die personenbezogenen Daten nicht bei der betroffenen Person erhoben wurden)

(...) Zusätzlich zu den Informationen gemäß Absatz 1 stellt der Verantwortliche der betroffenen Person zum Zeitpunkt der Erhebung dieser Daten folgende weitere Informationen zur Verfügung, die notwendig sind, um eine faire und transparente Verarbeitung zu gewährleisten: (...) das Bestehen einer automatisierten Entscheidungsfindung einschließlich Profiling gemäß Artikel 22 Absätze 1 und 4 und – zumindest in diesen Fällen – aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung für die betroffene Person.

i

Art. 15 DSGVO. Auskunftsrecht der betroffenen Person

Die betroffene Person hat das Recht, von dem Verantwortlichen eine Bestätigung darüber zu verlangen, ob sie betreffende personenbezogene Daten verarbeitet werden; ist dies der Fall, so hat sie ein Recht auf Auskunft über diese personenbezogenen Daten und auf folgende Informationen: (...) das Bestehen einer automatisierten Entscheidungsfindung einschließlich Profiling gemäß Artikel 22 Absätze 1 und 4 und – zumindest in diesen Fällen – aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung für die betroffene Person.

Die Vorschriften statuieren umfangreiche Informationspflichten, die gegenüber der betroffenen Person zu erfüllen sind (WP 29, 2018), und räumen dieser Person weitreichende (und grundrechtlich fundierte, Art. 8 Abs. 2 Satz 1 GRCh) Auskunftsrechte ein. Aber auch diese Vorschriften lassen in der Rechtsanwendung erhebliche Spielräume. Dies hat zwei Gründe.

Erstens haben Belange, die dem Transparenzprinzip entgegenlaufen, in der Datenschutz-Grundverordnung ebenfalls Anerkennung gefunden und sind demzufolge bei der Auslegung der wertungsoffenen Begriffe des Transparenzregimes zu berücksichtigen. Erwägungsgrund 63 macht dies deutlich: „Eine betroffene Person sollte ein Auskunftsrecht hinsichtlich der sie betreffenden personenbezogenen Daten, die erhoben worden sind, besitzen und dieses Recht problemlos und in angemessenen Abständen wahrnehmen können, um sich der Verarbeitung bewusst zu sein und deren Rechtmäßigkeit überprüfen zu können. (...) Dieses Recht sollte die Rechte und Freiheiten anderer Personen, etwa Geschäftsgeheimnisse oder Rechte des geistigen Eigentums und insbesondere das Urheberrecht an Software, nicht beeinträchtigen. Dies darf jedoch nicht dazu führen, dass der betroffenen Person jegliche Auskunft verweigert wird.“ Es ist erkennbar, dass hier ein Regelungsproblem benannt, aber nicht gelöst wird. Konsens über die Frage, wie der Belang des Geschäftsgeheimnisschutzes und Auskunftsrechte nach geltendem Datenschutzrecht austariert sind, vermag die Datenschutz-Grundverordnung nicht zu stiften.

Zweitens definiert die Datenschutz-Grundverordnung das den Verantwortlichen treffende Pflichtenprogramm an einer für die Scoring-Transparenz zentralen Stelle in markant unanschaulicher Art und Weise. Art. 13 Abs. 2 Buchstabe f, Art. 14 Abs. 2 Buchstabe g und Art. 15 Abs. 1 Buchstabe h DSGVO bestimmen jeweils Informationen bzw. Auskünfte über „die involvierte Logik“ (*the logic involved; la logique sous-jacente*) einer automatisierten Entscheidungsfindung im Sinne von Art. 22 DSGVO. Man könnte vermuten, dass sich in den Wissenschaften, deren Gegenstand Algorithmen sind, der Begriff der „Logik“ auf eine bestimmte Beschreibungsperspektive auf einen Algorithmus beziehe, und das Recht diesen außerrechtlichen Begriff in Bezug nehmen und für eine juristische Rezeption (exemplarisch Klement, 2006; Mathis, 2017) bereitlegen würde. Diese Vermutung geht fehl. Was die „involvierte Logik“ einer automatisierten Entscheidung sein könnte, davon haben Mathematiker, Informatiker und Softwareingenieure eine nicht weniger vage Vorstellung als Juristen.

An der intensiv geführten Debatte (siehe oben B.I.3) über die Offenlegung der in den Bonitäts-Score der SCHUFA eingehenden Merkmale und ihrer Gewichtung lässt sich die fehlende normative Anleitung durch das Transparenzregime der Datenschutz-Grundverordnung anschaulich machen. Nimmt man an, dass die Ermittlung eines SCHUFA-Scores eine Entscheidung im Sinne des Art. 22 DSGVO darstellt, bleibt gleichwohl offen, welche Informationen über das Zustandekommen des Scores von einer Beschreibung der „involvierten Logik“ des Scoring-Verfahrens erfasst sind (Nachweise der Ansichten über die Reichweite der Vorschriften bei Wischmeyer, 2018; restriktiv etwa Paal & Hennemann, 2018, Art. 13 DSGVO Rn. 31; extensiv etwa Bäcker, 2018, Art. 13 DSGVO Rn. 54). Zum Teil wird in ausdrücklicher Bezugnahme auf die SCHUFA-Entscheidung des Bundesgerichtshofs eine weiterreichende Auskunftspflicht angenommen, als sie nach der bisherigen Rechtslage bestand: „In Bezug auf die Reichweite der Information über die ‚involvierte Logik‘ bleibt abzuwarten, ob die vom BGH gebilligte Geheimhaltung der Score-Formel (...) haltbar bleibt. Ohne Kenntnis der Score-Formel ist es der betroffenen Person nämlich kaum möglich, Fehler bei der Berechnung des Score-Wertes aufzudecken und zu korrigieren (...). Andererseits kann die unbeschränkte Offenlegung der Score-Formel das Geschäftsmodell von Auskunftfeiern beeinträchtigen (...).

Aufgrund der analogen Anwendung von Art. 15 Abs. 4 und der darin angeordneten Abwägung von Grundrechten und -freiheiten wird man jedoch die vom BGH gebilligte strikte Geheimhaltung der Score-Formel nicht mehr aufrechterhalten können, wenn deren Kenntnis für die betroffene Person notwendig ist, um fehlerhafte Berechnungen festzustellen und korrigieren zu lassen. Vielmehr wird es auf den Einzelfall ankommen, sodass in bestimmten Fällen sowohl die Score-Formel als auch deren zugrundeliegende Parameter durchaus Gegenstand einer Auskunft sein können.“ (Schmidt-Wudy, 2018, Art. 15 DSGVO Rn. 78.3). Die Vorsicht, mit der der Kommentator hier seine Rechtsauffassung äußert, ist illustrativ für die markant schwache normative Anleitung durch die Art. 13 bis 15 DSGVO (siehe aber Heuzeroth & Seibel, 2018). Hinter der – ihrerseits weit von Eindeutigkeit entfernten – normativen Anleitung durch § 34 BDSG a.F., auf dessen Grundlage der Bundesgerichtshof den Auskunftsanspruch gegen die SCHUFA konturierte, bleibt die jetzige Rechtslage zurück.

Vor diesem Hintergrund verwundert es nicht, dass über die Reichweite der Transparenzanforderungen aus der Datenschutz-Grundverordnung kontrovers diskutiert wird. Kristallisationspunkt der Debatte ist die Frage, ob die Datenschutz-Grundverordnung der betroffenen Person ein „Recht auf Erklärung“ einer automatisierten Einzelfallentscheidung einräume. Diese international geführte Diskussion bemüht sich um einen Brückenschlag zwischen einerseits den Transparenzanforderungen der Datenschutz-Grundverordnung und andererseits der lebhaft geführten Diskussion über die Möglichkeiten, komplexe algorithmische Entscheidungssysteme für Menschen verständlich zu machen (siehe bereits oben B.I.4 sowie Gesellschaft für Informatik, 2018; Selbst & Powles, 2017; Selbst & Barocas, 2018; Wachter, Mittelstadt & Floridi, 2017).

Nicht zu verkennen ist zwar, dass die Datenschutz-Grundverordnung in ihren Transparenzanforderungen an ihre unionsrechtliche Vorgängerin, die Datenschutz-Richtlinie, anschließt. Dies legt eine sehr zurückhaltende Interpretation der Transparenzanforderungen aus den Art. 13 bis 15 DSGVO nahe (Wischmeyer, 2018). Dann wären die nach den Vorschriften geschuldeten Informationen in hohem Maße allgemein gehalten und beschränkten sich auf eine lediglich oberflächliche Darstellung der Programmfunktionen. Andererseits mag diese zurückhaltende Konturierung der datenschutzrechtlichen Transparenzanforderungen Ausdruck des Umstandes sein, dass derzeit noch diskutiert wird, wie die Herstellung von Transparenz konkret möglich ist (siehe bereits oben B.I.4). Nicht im rechtlichen Dürfen, sondern im technischen Können hat der Transparenzdiskurs derzeit sein Zentrum (siehe Selbst & Barocas, 2018; Burrell, 2016; Lipton, 2016). Die technizistisch anmutende, aber in ihrer Substanz vage Beschreibung der geschuldeten Transparenz – „involvierte Logik“, „Tragweite“, „angestrebte Auswirkungen“ – mag sich deshalb als besonders rezeptionsoffen für zukünftige Entwicklungen der Wissenschaft erweisen.

5. Gewährleistung des Diskriminierungsverbots

i

§ 1 AGG. Ziel des Gesetzes

Ziel des Gesetzes ist, Benachteiligungen aus Gründen der Rasse oder wegen der ethnischen Herkunft, des Geschlechts, der Religion oder Weltanschauung, einer Behinderung, des Alters oder der sexuellen Identität zu verhindern oder zu beseitigen.

i

§ 3 AGG. Begriffsbestimmungen

(1) Eine unmittelbare Benachteiligung liegt vor, wenn eine Person wegen eines in § 1 genannten Grundes eine weniger günstige Behandlung erfährt, als eine andere Person in einer vergleichbaren Situation erfährt, erfahren hat oder erfahren würde. (...)

(2) Eine mittelbare Benachteiligung liegt vor, wenn dem Anschein nach neutrale Vorschriften, Kriterien oder Verfahren Personen wegen eines in § 1 genannten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können, es sei denn, die betreffenden Vorschriften, Kriterien oder Verfahren sind durch ein rechtmäßiges Ziel sachlich gerechtfertigt und die Mittel sind zur Erreichung dieses Ziels angemessen und erforderlich.

5.1 Diskriminierungshandlung und Diskriminierungswirkung

Das geltende Antidiskriminierungsrecht hat konzeptionelle Schwierigkeiten, das Problem des „diskriminierenden Scorings“ (siehe oben B.II) zu erfassen. Denn es kontrolliert typischerweise, ob die Gründe, die Menschen oder Institutionen für ihre Handlungen angeben, antidiskriminierungsrechtlich legitim sind. Ob ein Handlungsgrund antidiskriminierungsrechtlich zu beanstanden ist oder nicht, lässt sich in zwei Schritten ermitteln:

Im ersten Schritt ist zu fragen, ob der Beweggrund für das auf seine Rechtskonformität zu prüfende Verhalten antidiskriminierungsrechtlich aufmerksamkeitsbedürftig ist. Das kann daran liegen, dass für das Verhalten ein in § 1 AGG genanntes Merkmal mit ausschlaggebend war. Zum Beispiel: Der Vertragsschluss wird „wegen der ethnischen Herkunft“ abgelehnt (siehe § 3 Abs. 1 AGG). Genauer zu prüfen sind aber auch solche Verhaltensweisen, die ihren Grund in dem Anschein nach neutralen Kriterien finden, wenn diese Kriterien Personen wegen eines in § 1 AGG genannten Merkmals in besonderer Weise benachteiligen. Zum Beispiel: Der Vertragsschluss wird „wegen unzureichender Kenntnisse der deutschen Sprache“ abgelehnt (siehe § 3 Abs. 2 AGG). Im zweiten Schritt wird geprüft, ob die Verwendung des verdächtigen Kriteriums in der konkreten Situation gerechtfertigt ist. Am Ende dieser Prüfung steht fest, ob eine verbotene Diskriminierung stattgefunden hat oder nicht.

Verbotenerweise zu diskriminieren bedeutet also, aus verbotenen Gründen zu handeln (eingehend Schramm, 2013). Das Antidiskriminierungsrecht ist „input-orientiert“. Es richtet seine Aufmerksamkeit auf das Zusammenspiel von bestimmten Entscheidungskriterien und deren Zulässigkeit. Im Bereich des Scorings kann dieses Rechtsanwendungsverfahren auf das falsche Gleis führen. Zum Beispiel knüpft ein Verkäufer, der seinem Gegenüber wegen eines niedrigen Scores etwa ein Vertragsangebot vorenthält, nicht an ein geschütztes Merkmal an, sondern eben an den Score. An diesem Ablehnungsgrund ändert auch die Tatsache nichts, dass für die Score-Berechnung beispielsweise das Geschlecht des Gegenübers von Bedeutung ist. Das Vorliegen einer antidiskriminierungsrechtlich aufmerksamkeitsbedürftigen Ungleichbehandlung könnte man unter Hinweis auf diesen Umstand bestreiten (so Moos & Rothkegel, 2016; siehe auch C.III.5): Abgelehnt wird der Vertragsschluss ja nicht *wegen* des Geschlechts, sondern *wegen* des unzureichenden Score-Wertes. Zwar kann man versuchen, dieses Phänomen mithilfe der Regeln zur mittelbaren Diskriminierung antidiskriminierungsrechtlich in den Griff zu bekommen. Allerdings werden die Schwierigkeiten dadurch nicht beseitigt.

In die Ermittlung eines Score-Wertes gehen sehr viele, womöglich sogar unüberschaubar viele, Einzelmerkmale ein. Praktisch nicht zu handhaben wäre eine Prüfung, die jedes der in einen Score eingehenden Merkmale ein-

zeln in den Blick nimmt und seine antidiskriminierungsrechtliche Zulässigkeit bewertet. Darüber hinaus liefe eine solche Prüfung Gefahr, bloße Scheinkorrelationen aufzudecken und als rechtfertigungsbedürftig auszuweisen, wiewohl deren Vorkommen in hinreichend großen Datenmengen geradezu unvermeidbar ist.

Für komplexe Scoring-Verfahren ist deshalb die „input-orientierte“ durch eine „impact-orientierte“ antidiskriminierungsrechtliche Analyse zu ergänzen. Das heißt, in den Blick zu nehmen sind nicht nur die Kriterien, nach denen Entscheidungen getroffen werden, sondern auch die Wirkungen, die Entscheidungen zeitigen. Der europäische Gesetzgeber hat für den Umstand, dass das auf Handlungsgründe bezogene Antidiskriminierungsrecht im Falle komplexer Datenverarbeitungsprozesse – und damit auch im Falle des Scorings – an seine Grenzen stößt, ein feines Formulierungsgespür bewiesen. In Erwägungsgrund 71 der Datenschutz-Grundverordnung spricht er nicht von „Diskriminierungen“ durch Datenverarbeitungsvorgänge, sondern von „diskriminierenden Wirkungen“.

Rechtlich anknüpfen kann eine wirkungszentrierte Analyse an den Umstand, dass § 3 Abs. 2 AGG nicht nur von „dem Anschein nach neutrale[n] Vorschriften“ und „Kriterien“ spricht, sondern auch von „Verfahren“. Über den Begriff des Verfahrens löst sich das Antidiskriminierungsrecht von der Kontrolle einzelner Handlungsgründe. Es öffnet sich für Betrachtungen, die auch komplexe und schwer zu durchschauende „Praktiken“ (Block, 2018, § 3 AGG Rn. 69) als antidiskriminierungsrechtlich problematisch ausweisen können. Weder das nationale Antidiskriminierungsrecht noch die zugrundeliegenden europäischen Richtlinien verlangen hinsichtlich der Frage des Anknüpfungspunktes, dass einzelne Kriterien eine Benachteiligung nach sich ziehen, sondern erlauben auch die Gesamtbetrachtung mehrerer Vorschriften oder ganzer Verfahren (Schiek, 2007, § 3 AGG Rn. 33). Ansatzpunkte für einen Diskriminierungsschutz, der nicht bei Handlungsgründen für Entscheidungen, sondern bei den Resultaten von unverständlich operierenden Systemen ansetzt, finden sich auch in der Rechtsprechung des Gerichtshofs der Europäischen Union. Das deutsche Antidiskriminierungsrecht wird durch diese Rechtsprechung in großem Ausmaß mitgestaltet. Sie hat Entlohnungssysteme als geschlechterdiskriminierend bewertet, auch ohne dass dafür einzelne diskrimi-

nierende Faktoren herauspräpariert werden mussten (EuGH, Urteil vom 27. Oktober 1993 – Rs. C-127/92 –, Enderby [ECLI:EU:C:1993:859]).

Der Bezugspunkt einer „impact-orientierten“ Betrachtung von Scoring-Verfahren sind nicht die einzelnen Merkmale, die in die Ermittlung des Score-Wertes eingehen, sondern die Wirkungen des Scoring-Verfahrens. Das Scoring-Verfahren selbst ist das „Verfahren“ im Sinne von § 3 Abs. 2 AGG, dessen antidiskriminierungsrechtliche Unbedenklichkeit zu gewährleisten ist (Hacker, 2018).

5.2 Herausforderungen wirkungsorientierten Diskriminierungsschutzes

Die Schwierigkeiten, der diskriminierenden Belastung bestimmter Personengruppen durch Scoring-Verfahren abzuwehren, können erheblich sein. Noch vergleichsweise leicht adressierbar sind solche Diskriminierungswirkungen, die auf einer mangelhaften technischen Gestaltung des Scoring-Verfahrens basieren. Wenn dieses für unterschiedliche Personengruppen Qualitätsunterschiede aufweist (siehe oben B.II.3), die ohne zusätzliche Kosten vermeidbar gewesen wären, ist das Scoring-Verfahren antidiskriminierungsrechtlich zu beanstanden (Hacker, 2018). Auch zusätzliche Kosten können dem Scorer auferlegt werden, wenn dadurch ein größeres Maß an Diskriminierungsfreiheit erreichbar ist (Hacker, 2018). Wo allerdings die diskriminierenden Effekte eines Scoring-Verfahrens zugleich dessen Prognosekraft verbessern, ist dem Scorer die Möglichkeit eröffnet, diese diskriminierenden Effekte zu rechtfertigen (näher Hacker, 2018).

Hinzu treten Schwierigkeiten, die diskriminierenden Effekte eines Scoring-Verfahrens zu beweisen (Hacker, 2018). Einem Kläger, der Diskriminierung vermutet, fehlen die Vergleichsdaten, um seine Vermutung zu untermauern (Hacker, a.a.O., Hildebrandt, 2015) und diskriminierende Effekte eines Scoring-Verfahrens darzulegen. Und selbst aus einer Vogelperspektive, aus der zahlreiche Datensätze überblickt werden können, fällt es schwer, die diskriminierenden Wirkungen eines Scoring-Verfahrens zu identifizieren. Denn sieht man von Alter und Geschlecht einmal ab, sind Daten zu den Merkmalen, anhand deren klassischerweise diskriminiert wird, regelmäßig nicht verfügbar: „Bereits ihre Erhebung wäre problematisch, weil niemandem zuge-

mutet werden darf, seine sexuelle Orientierung oder seine Religion zu offenbaren. Die Feststellung ethnischer Zugehörigkeit wird das grundlegende Problem auf, ob es überhaupt ‚objektive‘ Faktoren zur statistischen Erfassung dafür gibt, sieht man einmal von der Staatsangehörigkeit ab.“ (Grünberger, 2013, S. 664; vgl. auch Art. 9 DSGVO).

Eine schlichte Ergänzung des sachlichen Anwendungsbereiches des Allgemeinen Gleichbehandlungsgesetzes um „Verfahren automatisierter Entscheidungsfindung“ wäre nicht ausreichend, dem Problem des diskriminierenden Scorings zu begegnen. Potenziale zu dessen Lösung enthält das Datenschutzrecht. Der Grundsatz einer Datenverarbeitung nach „Treu und Glauben“ (*fairness, loyauté*), den Art. 5 Abs. 1 Buchstabe a DSGVO als Rechtsgebot statuiert, ist hierfür ein normativer Anker (Hacker, 2018). Im Falle der behebbaren diskriminierenden Qualitätsunterschiede von Scoring-Verfahren ist zudem der Grundsatz der Richtigkeit (Art. 5 Abs. 1 Buchstabe d DSGVO) betroffen (Hacker a.a.O.). Ist das Datenschutzrecht über diese Rechtsgrundsätze für den normativen Belang des Diskriminierungsschutzes geöffnet, bieten Auskunftsrechte nach Art. 15 Abs. 1 Buchstabe h DSGVO sowie Datenschutz-Folgeabschätzungen (Art. 35 DSGVO) und datenschutzbehördliche Aufsichtsinstrumente Möglichkeiten, dem Problem des diskriminierenden Scorings zu begegnen (näher Hacker, 2018).

IV. Aufsicht

Die rechtlichen Anforderungen an das Scoring betreffen verschiedene Scoring-Aspekte, prominent etwa Transparenz, Qualität und die Achtung des Diskriminierungsverbots. Die Anforderungen unterscheiden sich danach, wer wen zu welchem Zweck und auf welche Weise scort. Mit dieser heterogenen Vielfalt materieller Vorgaben korrespondiert eine Vielzahl von Rechtsverwirklichungsmechanismen. Einen Eindruck von der Vielfalt denkbarer institutioneller Arrangements zur Rechtsdurchsetzung vermittelt schon allein das Datenschutzrecht, das für die Formulierung rechtlicher Anforderungen an das Scoring eine zentrale Rolle einnimmt (siehe Schantz & Wolff, 2017, S. 295 ff.). Zu ihnen zählen die Rechtsschutzmöglichkeiten des individuell nachteilig betroffenen Verbrauchers oder Verbandsklagemöglichkeiten etwa der Verbraucherzentralen, aber auch Vorgaben an die Unternehmensorganisation (umfassend Spindler, 2011, exemplarisch etwa zum Datenschutz S. 270 ff.) oder Mechanismen staatlicher Aufsicht über die Rechtskonformität des Scorings.

Die staatliche Aufsicht ist das zentrale Instrument zur Durchsetzung der oben beschriebenen Qualitätsanforderungen an das prognostische Scoring nach § 31 BDSG. Über die Einhaltung der Qualitätsanforderungen an die Modelle zur Ermittlung von Kreditausfallrisiken führt die BaFin Aufsicht (siehe oben E.III.3). Im Bereich der Bonusprogramme der gesetzlichen Krankenversicherungen (siehe oben E.II.3) bestehen gesetzliche Grundlagen, die eine besonders intensive Aufsicht ermöglichen (Ullrich, 2018, § 65a SGB V Rn. 7). Es ist möglich, das materielle Recht, über dessen Einhaltung die genannten Stellen wachen, zu ergänzen und damit den Auftragsauftrag der betreffenden Stellen zu erweitern.

Das Potenzial hoheitlicher Aufsicht im Bereich des Scorings ist schon deshalb als vergleichsweise groß zu veranschlagen, weil sich die Fragen des Geheimnisschutzes (etwa im Sinne eines Geschäftsgeheimnisses) und nachteiliger gesellschaftlicher Folgen von Transparenz (siehe oben B.I.1) nicht stellen. Denn mit einer Kenntnis der aufsichtführenden staatlichen Stelle ist der Kreis derjenigen, die Einblick in die Funktion des betreffenden Scoring-Algorithmus haben, nicht auf eine unüberschaubare Allgemeinheit erweitert. Die kontrollierenden staatlichen Funktionsträger können ihrerseits auf Geheimhaltung verpflichtet werden und sind dies auch in der Regel (siehe etwa § 30 und § 29 Abs. 2 VwVfG).

So stellt sich das Problem behördlicher Aufsicht als ein Problem angemessener personeller und sachlicher Ausstattung der jeweils zuständigen Aufsichtsbehörden dar. Diese sind in die Lage zu versetzen, auch komplexe materiellrechtliche Prüfungen durchzuführen (siehe bereits SVRV, 2016; weiter Gesellschaft für Informatik, 2018).

Handlungs- empfehlungen



Gegenstand und Ziel

Als Scoring bezeichnen wir die Zuordnung eines Zahlenwertes zu einem Menschen zum Zweck der Verhaltensprognose oder Verhaltenssteuerung. Die Bestimmung dieses Zahlenwertes erfolgt in der Regel auf der Grundlage einer breiten Datenbasis durch ein algorithmisches Verfahren.

Scoring ist nicht gleichzusetzen mit Entscheidung. Ein Score bildet immer nur die Grundlage für eine Entscheidung, die von verantwortlichen Menschen auf der Basis des Scores getroffen wird (Einzelfallprüfung, etwa bei der Kreditvergabe von Banken).

Scoring ist keineswegs ein Phänomen der digitalen Welt. Traditionell werden Menschen nicht nur hinsichtlich ihrer Kreditwürdigkeit, sondern auch in Bildungseinrichtungen „gescored“, die Noten vergeben. Die Digitalisierung vervielfacht nun die Möglichkeiten, Personen einzuschätzen und zu bewerten. Damit vervielfacht sie sowohl die Risiken als auch die Chancen für Verbraucher: Denn formalisiertes Scoring ist auditier- bzw. überprüfbar; Scoring hat das Potenzial, individuelle Fehler und unerwünschte Diskriminierung von Entscheidungen zu reduzieren. Es könnte insofern informellen Entscheidungsprozessen durch Einzelne im Grundsatz überlegen sein.

Der Begriff Algorithmus kommt ebenfalls aus der analogen Welt und reicht von einfachen Entscheidungsregeln und mathematischen Verfahren bis hin zu Künstlicher Intelligenz und Deep Neural Networks. Algorithmen durchdringen mehr und mehr Lebensbereiche (z. B. personalisierte Preise im Onlinehandel, Ergebnisreihung durch Suchmaschinen, künftig selbstfahrende Autos). Das Gutachten betrachtet nur einen kleinen Ausschnitt von Scoring-Algorithmen, nämlich Scoring in den drei Themenfeldern Bonität, Gesundheit und Kfz-Versicherung. Unsere Analysen und Empfehlungen sind jedoch in modifizierter Form auch auf andere Bereiche anwendbar, etwa Personalauswahl-Verfahren und Arbeitnehmer-Scoring (People Analytics).

Das Potenzial digitalen Scorings kann nur ausgeschöpft werden, wenn eine Reihe von gesellschaftlich legitimen Erwartungen erfüllt sind. Nur dann wird es von Verbrauchern auch akzeptiert werden und nur dann wird es ihnen zunutze sein. Wenn diese Bedingungen erfüllt sind, sprechen wir von „verbrauchergerechtem Scoring“. So

müssen Verwechslungen der Betroffenen im Rahmen des Datenschutzes ausgeschlossen bzw. minimiert werden; Widerspruch gegen den eigenen Score-Wert muss auf einfache Weise effektiv möglich sein. Es darf weder direkt noch indirekt auf Basis von geschützten Merkmalen (wie etwa dem Geschlecht) ungerechtfertigt diskriminiert werden. Scoring zur Vorhersage muss belegen, wie gut die Qualität der verwendeten Merkmale sowie die Vorhersagen wirklich sind. Solch prädiktive Kraft sollte zudem über verschiedene sozioökonomische Gruppen hinweg stabil sein. Zudem darf das kommunizierte Ziel eines Scorings die Betroffenen nicht in die Irre führen. Vorhersagen müssen das Scoring-Ziel erfüllen und sollten nicht leichtfertig auf andere Bereiche angewendet werden als den, für den ein Score berechnet wurde. Vor allem aber muss Scoring für die Betroffenen zu verstehen sein.

Die Bedingungen verbrauchergerechten Scorings zu benennen, zu begründen und Vorschläge für die Umsetzung zu machen, ist Aufgabe des vorliegenden Gutachtens.

Dabei bedarf nicht jedes Scoring-Verfahren in gleichem Umfang einer tiefgehenden Prüfung. Im Mittelpunkt stehen solche Scoring-Verfahren, die für das Wohl der Verbraucher erheblich sind. Besonders relevant sind sensible Lebensbereiche, etwa, wenn der einem Menschen zugeordnete Score direkte Auswirkungen darauf hat, ob dieser einen Kredit bekommt und zu welchen Bedingungen, wenn der Scorewert die Höhe der Versicherungsbeiträge beeinflusst oder das Gesundheitsverhalten gezielt steuert. Bislang wurden weniger sensible Bereiche (wie Betrugserkennung im Online-Handel) von den hoch relevanten nicht deutlich abgegrenzt. Eine solche Abstufung sollte jedoch vorgenommen werden, um die notwendige Regulierungstiefe und -dichte eines verbrauchergerechten Scorings angemessen zu bestimmen.

Die nachfolgenden Empfehlungen bauen auf den vier Abschnitten des Gutachtens auf. Abschnitt B beleuchtet Begriffe, Strukturen und Funktionsweisen von Scoring, stellt den Stand der internationalen Forschung dar und diskutiert die relevanten Handlungsfelder. Es folgen drei Abschnitte mit vertieften Analysen, zwei davon bauen auf eigenen empirischen Erhebungen auf, auf die in den folgenden Empfehlungen auch immer wieder Bezug genommen wird. Abschnitt C enthält die Ergebnisse einer

Befragung von Unternehmen, die Bonitäts-Scores anbieten, Scores für Kfz-Haftpflicht (Telematiktarife) oder Boni für Krankenversicherte nutzen. Abschnitt D enthält die Ergebnisse einer repräsentativen Bevölkerungsbefragung zu Kenntnissen über und Einstellung zu Scoring. Kapitel E stellt Umfang (und Lücken) des Rechtsrahmens des Scoring sowohl innerhalb der materiellen Regeln als auch in Bezug auf die aufsichtsrechtlichen Kompetenzen dar. Auch dieses Kapitel ist unmittelbar für die Empfehlungen wichtig.

1. Scoring für den Verbraucher verständlich machen

Problem:

In der Bevölkerung in Deutschland besteht in Bezug auf Daten, Algorithmen, Geschäftsentscheidungsprozesse und deren Zusammenspiel ein hoher Aufklärungs- und Erklärungsbedarf. So werden beispielsweise heutzutage Kreditentscheidungen durch die Aufsichtsbehörden sehr umfassend reguliert und verschiedene Empfehlungen dieses Gutachtens sind somit bereits weitgehend adressiert. Gleichwohl ist dies in der Bevölkerung wenig bis gar nicht bekannt. Es herrscht Unsicherheit und Verunsicherung. Die Notwendigkeit, gemeinsam mit allen Beteiligten Informations- und Aufklärungsarbeit zu leisten, kann nicht genug betont werden.

Die Ergebnisse der repräsentativen Bevölkerungsbefragung zeigen im Detail, dass das Wissen über die verwendeten Merkmale im Bereich des Bonitäts-Scorings neben dem Alter davon abhängt, ob eine Person in den vergangenen fünf Jahren eine Selbstauskunft eingeholt hat. Formale Bildung für sich genommen scheint dagegen keinen relevanten Einfluss auf das Scoring-Wissen zu haben. Interessant ist auch, dass die Befragten mehrheitlich dagegen sind, eine Rückmeldung zu ihrem Score-Wert zu erhalten; knapp die Hälfte möchte grundsätzlich (mit unterschiedlicher Häufigkeit) darüber aktiv informiert werden.

Handlungsempfehlungen:

1. Der SVRV empfiehlt den Datenschutzbehörden, die in der DSGVO (vgl. Art. 15 Abs. 1 Buchst. h) festgeschriebenen Verständlichkeitsanforderungen für Scoring und score-basierte Geschäftsprozesse zu operationalisieren. Maßstab für Verständlichkeit sollte der Durchschnittsverbraucher sein. Soweit Scoring eine für den einzelnen Verbraucher nicht mehr nachvollziehbare Komplexität erreichen sollte, wäre sicherzustellen, dass Scoring nicht nur für Aufsichtsbehörden, sondern auch zumindest für Verbraucherorganisationen und Nicht-Regierungsorganisationen nachvollziehbar ist.
2. Scoring-Anbieter sollen den Verbrauchern die für sie wesentlichen Merkmale, auf deren Basis sie gescort werden, sowie möglichst auch deren Gewichtung auf verständliche und nachvollziehbare Weise offenlegen. Allerdings dürfen Geschäftsgeheimnisse dadurch nicht verletzt werden. Festzustellen, welche Merkmale für Verbraucher wesentlich sind, kann nicht nur Aufgabe des Gesetzgebers sein, sondern sollte auch Aufgabe von Verbraucherorganisationen oder der „Marktwächter“ der Verbraucherzentralen sein. Eine vollständige Offenlegung der Scores und ihrer Eigenschaften gegenüber Aufsichtsbehörden ist in jedem Fall notwendig (vgl. das SVRV-Gutachten Digitale Souveränität).

Ein Teil des SVRV spricht sich für eine weitreichende Scoring-Transparenz aus. Er ist der Auffassung, dass stets sämtliche in einen Score eingehenden Merkmale gegenüber dem Verbraucher offenzulegen sind und ihr relatives Gewicht in der Score-Berechnung anzugeben ist. Etwaige Geheimhaltungsinteressen der Scoring-Anbieter und -Nutzer haben insoweit gegenüber dem Auskunftsinteresse der Verbraucher zurückzustehen, wobei das Geschäftsgeheimnis der Entwicklung und Programmierung eines Scores gewahrt bleiben sollte.

3. Ein besseres Verständnis der Verbraucher bezüglich Scoring wird durch die bloße Offenlegung nicht automatisch sichergestellt. Dafür sind

verschiedene Maßnahmen notwendig, etwa: beispielhafte Darstellungen von Scores und deren Abstufungen für verschiedene Merkmale; didaktische visuelle Aufbereitung (etwa durch Verbraucherorganisationen); allgemein Verbesserung der Scoring-bezogenen Kompetenzen der Verbraucher (Skillset). Die Verständlichkeit der Scores durch Verbraucher sollte nicht nur von Fachleuten behauptet, sondern empirisch überprüft werden.

4. Verbraucher haben jetzt schon einen Anspruch (vgl. Art. 13 Abs. 2 lit. f, 15 Abs. 1 Buchst. h DSGVO) darauf, in Textform auf verständliche Weise adressatengerecht darüber informiert zu werden, dass sie gescort werden. Doch fehlt es an einer Konkretisierung. Unternehmen, Aufsichtsbehörden und Verbraucherorganisationen sollten gemeinsam Standards für Score-Anbieter entwickeln, damit Relevanz und Verständlichkeit gewährleistet werden können. Der SVRV empfiehlt zudem, Verbrauchern mitzuteilen, wie ihr persönlicher Score bezüglich der Verteilung aller Score-Werte in der Bevölkerung zu interpretieren ist (z. B. dass er „im oberen Drittel“ aller Score-Werte liegt).
5. Eine zeitnahe kostenlose Rückmeldung bei größerer Veränderung eines individuellen Scores (z. B. das Abrutschen in eine schlechtere Kategorie) ist zu gewährleisten – zumindest als wählbare Option. Dieser Mitteilung sind Grenzen gesetzt: Um eine Veränderung eines Scorewertes erkennen zu können, muss der Score-Anbieter alte Score-Werte speichern. Dies wird für viele praktische Anwendungen (wie bei Betrugserkennung oder der Bestimmung anzubietender Zahlungsoptionen) nicht der Fall sein. Die Berechnung eines Scores erfolgt bei Versicherern und Banken anlassbezogen. Damit wird keine Historie aufgebaut und so wird die Veränderung beim nächsten „Event“ nicht klar. Damit kann sich die Forderung nur auf die Institutionen beziehen, die kontinuierlich Daten sammeln, etwa Bonitäts-Scorer und das Fahreignungsregister des Kraftfahrt-Bundesamtes in Flensburg („Verkehrssünderkartei“), das solche Mitteilungen bereits versendet.

2. Scoring-Wissen und Kompetenzen fördern

Problem:

Verbraucher wissen wenig über Scoring. Dies liegt keineswegs nur an Intransparenz auf Seiten der Scoring-Anbieter und -Nutzer. Wissen ist jedoch wesentlich, damit informierte Verbraucher ihre Prüf- und Einspruchsmöglichkeiten wahrnehmen und systematische Fehlerquellen aufdecken können. Aufsichtsbehörden, Verbraucherorganisationen und deren Marktwächter sollen die Verbraucher entlasten.

Handlungsempfehlungen:

Wie bereits im SVRV-Gutachten „Digitale Souveränität“ empfohlen wurde, sollten staatliche Einrichtungen, Bildungsträger, NGOs, Verbraucherschutzorganisationen und Verbraucherschutzprojekte

1. Wissen zu Grundaspekten vermitteln, die alle Bereiche des Scorings und seine Verwendung in bestimmten Geschäftsfeldern betreffen. Hierzu sollte die Bundesregierung in ihrer Digitalisierungsstrategie innerhalb der laufenden Legislaturperiode Informations- und Diskussionsmaterial erarbeiten, um die Kompetenzen von Verbrauchern, Multiplikatoren und Entscheidungsträgern zu verbessern. Das Grundprinzip und die Qualitätsaspekte von Scoring sowie die Formen von Ungleichbehandlung und deren Ursachen gehören ebenso zum Grundwissen wie die einem Betroffenen zustehenden Rechte.
2. Kompetenzen fördern, um informierte Entscheidungen hinsichtlich der Teilnahme an einem Scoring-Verfahren zu treffen. Hierunter fällt die Kompetenz, Scoring-Angebote sowie Alternativen zu finden, kritisch zu prüfen, zu bewerten (z. B. werden die für den Verbraucher relevanten Informationen offengelegt) und zu nutzen.

3. Diskriminierung prüfen und offenlegen

Problem:

Scoring ist – wie informelle Entscheidungssysteme – nie perfekt, sondern kann immer auch mit Fehlern behaftet sein, systematischen oder unsystematischen. Ungerechtfertigte Diskriminierung ist ein solcher systematischer Fehler, da eine Gruppe von Verbrauchern fälschlicherweise systematisch niedrigere (höhere) Werte als eine andere Gruppe erhält. Die Marktbefragung innerhalb dieses Gutachtens zeigt beispielsweise, dass die meisten Auskunftsteilnehmer Geo-Scoring verwenden. Merkmale nach dem allgemeinen Gleichbehandlungsgesetz wie Geschlecht oder Alter werden von einigen Auskunftsteilnehmern für die Scores herangezogen. Bei Kfz-Telematiktarifen könnten bestimmte Verbrauchergruppen benachteiligt werden – wie z. B. Schichtarbeiter oder Stadtbewohner, durch schwer beeinflussbare Merkmale wie Zeit und Ort einer Fahrt.

Wir stellen Defizite in der Abstimmung des Antidiskriminierungsrechts mit der Datenschutzgrundverordnung fest, sowie vor allem in der Durchsetzung. Diese Durchsetzungsprobleme resultieren in erster Linie aus der schwierigen empirischen Nachweisbarkeit algorithmischer Diskriminierung auf individueller oder Gruppenebene. Dabei ist nicht zu verkennen, dass der Vergleich der Score-Verteilung zwischen verschiedenen geschützten Gruppen bisweilen schwer bis gar nicht möglich ist, weil Datenschutz-Vorgaben (siehe Art. 9 DSGVO) der Erhebung und Speicherung bestimmter Merkmale entgegenstehen können.

Handlungsempfehlungen:

1. Der SVRV empfiehlt, die Auskunftsansprüche von Verbrauchern zu stärken, wie sie in Art. 15 Abs. 1 Buchst. h DSGVO angelegt sind. Insbesondere sollten Verbraucher (sofern dies durch den Anbieter ermittelbar ist) erfahren können, wie die Scores zwischen verschiedenen Gruppen mit unterschiedlichen geschützten Merkmalen verteilt

sind, um so eine mögliche algorithmische Diskriminierung dem Grunde nach belegen zu können.

2. Der SVRV empfiehlt zudem die Stärkung der Aufsichtsbehörden (siehe Handlungsempfehlung 7).
3. Der SVRV spricht sich ferner für ein Verbandsklagerecht für Antidiskriminierungssachverhalte im Bereich des Scorings aus.

4. Telematikfreie Option sicherstellen

Problem:

Die Marktbefragung im Rahmen dieses Gutachtens zeigt, dass das Potenzial einer Ausweitung von Telematiktarifen in der Kranken- und Kfz-Versicherung gegeben ist, also die Nutzung von Online-Informationen über das Gesundheits- und Fahrverhalten. Die technischen Möglichkeiten zur Erfassung weiterer Daten entwickeln sich fortwährend (wie eCall-Technologie, Vernetzung von Automobilen, Wearables). Anbieter haben potenziell Interesse daran, immer mehr Merkmale einzubeziehen (z. B. im Gesundheitsbereich aus der elektronischen Patientenakte).

Wenn bestimmte Versicherte benachteiligt werden, weil ihnen der Erwerb von Vergünstigungen im Rahmen von verhaltensbasierten Tarifen, z. B. aufgrund körperlicher Einschränkungen, erschwert oder verwehrt wird, kann das Solidaritätsprinzip gefährdet sein. Versicherungstarife könnten in Abhängigkeit individueller Leistungsfähigkeit auseinanderdriften und die Wahlfreiheit eingeschränkt werden (z. B. von Personen mit hoher Datenschutzpräferenz). Die Verletzung des Solidaritätsprinzips ist in Deutschland im Kern durch die Sozialversicherungen wirksam ausgeschlossen; fragwürdig sind allerdings Bonusprogramme der gesetzlichen Krankenversicherungen.

Die Bevölkerungsbefragung hat ergeben, dass die Akzeptanz hinsichtlich der verhaltensbasierten Merkmale, die für das präsentierte Kfz- bzw. Gesundheitsscoring herangezogen werden sollen, insgesamt niedrig ist.

Handlungsempfehlungen:

1. Der SVRV empfiehlt, die Aufrechterhaltung telematikfreier Optionen im Versicherungsbe- reich (insbesondere Kfz-Haftpflichtversicherung, Krankenversicherung) gesetzlich zu garantieren. Insbesondere:
2. Versicherungsnehmern, die keinen Telematiktarif nutzen, dürfen keine substanziellen Nachteile im Vergleich zu Versicherungsnehmern mit Telematiktarif entstehen.
3. Telematiktarife sind nach Auffassung der Mehrheit des SVRV so auszugestalten, dass sie sich wirtschaftlich selbst tragen und dadurch nicht indirekt zu Lasten der Versicherungsnehmer gehen, die keinen Telematiktarif nutzen. Da insbesondere im Bereich der Krankenversicherungen Solidaritätsziele eine Rolle spielen, müsste sichergestellt werden, dass niedrigere Telematiktarife, die nur durch die Entscheidung von überdurchschnittlich gesunden Versicherten für diese Tarife entstehen und nicht die Ausgaben der Versicherer wesentlich reduzieren, nicht zulässig wären.

5. Score-Qualität gewährleisten

Problem:

Der deutsche Gesetzgeber fordert die Anwendung anerkannter „mathematisch-statistischer Verfahren“ (§ 31 BDSG), konkretisiert diese Vorgabe aber nicht. Auch die DSGVO enthält keine konkreten Vorgaben. Ein wissenschaftliches Verfahren, wie etwa multiple Regression, garantiert jedoch keineswegs automatisch die Qualität eines Scores. Außerdem sind komplizierte und schwer verständliche Verfahren nicht unbedingt besser als einfache und verständliche Regeln oder gar der Zufall. Deshalb müssen Anbieter ein hohes Qualitätsniveau für Scoring-Verfahren sicherstellen, und zwar nicht nur für die angewandte Methode, sondern auch für das Ziel des Scorings. Problematisch ist etwa, wenn bei einem „Gesundheits-Score“ massiv gesundheitsschädliches Verhalten wie Rauchen nur mit geringem Gewicht eingeht, wohingegen Aktivitäten, deren Gesundheitsnutzen nicht belegt ist, mit hohem Gewicht eingehen. Ein Score muss das messen, was er vorgibt zu messen, und er muss die weiteren in diesen Empfehlungen genannten Kriterien erfüllen.

Handlungsempfehlungen:

1. Der SVRV empfiehlt, ein anspruchsvolles Qualitätsleitbild auf der Basis von Beispielen guter Praxis (Best Practices) zu entwickeln. An bestehende Initiativen zur Qualitätssicherung algorithmischer Prozesse soll angeknüpft werden. Dieses Qualitätsleitbild soll in Zusammenarbeit zwischen Wirtschaft, den Aufsichtsbehörden, den Marktwächtern der Verbraucherzentralen und den Verbraucherorganisationen erarbeitet und aktualisiert werden (Ausarbeitung, Implementierung, Monitoring).
2. Scoring-Anbieter in den sensiblen Bereichen sollten verpflichtet werden, gegenüber den Aufsichtsbehörden darzulegen und im Detail überprüfbar zu machen, wie gut das Verfahren tatsächlich ist. Nur so ist es möglich, die Score-

Verfahren auf ihre Verbrauchergerechtigkeit hin zu testen. Diese Verpflichtung soll für Scores, die Verhalten prognostizieren, mittels statistischer Gütemaße (z. B. Falsch-Positiv Rate, Hit-Rate, Gini-Koeffizient, „Area under the ROC“) für die Gesamtpopulation und auch für relevante Teilpopulationen (nach Geschlecht, Alter, Bildung usw.) gelten. Dadurch können auch Diskriminierung erkennbar und fragwürdige Score-Qualitäten sichtbar gemacht werden.

3. Die nicht-adäquate Kennzeichnung des Ziels eines Scoring-Verfahrens ist bereits nach bestehender Rechtslage unzulässig. Zur Aufdeckung falsch gekennzeichnete Scores können neben Aufsichtsbehörden (siehe auch Empfehlung 7) aufgrund ihrer Expertise auch Verbraucherorganisationen oder die Marktwächter der Verbrauchzentralen beitragen.
4. Alle Stellvertreter-Merkmale, wie sie zum Beispiel beim Geo-Scoring verwendet werden, bedürfen einer besonderen Rechtfertigung (Stichwort Kausalzusammenhang) und Kontrolle durch die zuständige Aufsichtsbehörde. Die Verwendung von Proxy-Variablen sollte minimiert werden. Wenn sie eingesetzt werden, sollte ihr Zusammenhang zum Zielmerkmal inhaltlich plausibel begründet sein.

6. Datenqualität sichern

Problem:

Die Marktbefragung innerhalb dieses Gutachtens zeigt, dass alle Auskunftsteile Verbrauchern Korrekturmöglichkeiten bieten, wenn diese feststellen, dass falsche oder veraltete Informationen über sie vorliegen. Best Practice stellen hier weiterhin die Unternehmen in der Befragung dar, die laut eigenen Angaben kontinuierliche Datenüberwachung und Einhaltung von Löschrufen praktizieren, um die Qualität ihres Datenbestandes zu wahren.

Die Datenqualität ist ein Teilbereich von Data Governance, einem Framework, das in immer mehr Organisationen und Unternehmen umgesetzt wird. Dies unterstreicht die Bedeutung von Daten in der heutigen Zeit, nicht nur der personenbezogenen Informationen, und reflektiert die Notwendigkeit, mit dem Rohstoff unserer Zeit verantwortungsvoll und angemessen umzugehen. Die Data Governance definiert zum Beispiel die Verantwortlichkeiten für alle Entscheidungen, die Daten und deren Nutzen betreffen, sowie die relevanten Infrastrukturen und Sicherheitsaspekte.

Datenqualität ist keine singuläre Eigenschaft, sondern sie setzt sich aus diversen Dimensionen zusammen, beispielsweise der Korrektheit, der Vollständigkeit und der Aktualität von Informationen. Datenqualität ist insbesondere ein relativer Begriff. Eine gute Datenqualität liegt vor, wenn die „Fitness for use“ gewährleistet ist. Abhängig von der Anwendung sind die Anforderungen durchaus unterschiedlich zu definieren. Für hochqualitative Scoring-Verfahren muss eine sehr gute Datenqualität vorliegen.

Auf der einen Seite ist es unerlässlich, dass die Entwicklung eines Scoring-Verfahrens auf Daten von höchstmöglicher Qualität durchgeführt wird. Die Daten müssen einen hohen Grad an Korrektheit haben, sollten angemessen aktuell sein und einen Umfang aufweisen, der die Schätzung des Verfahrens erlaubt. Ferner ist die Repräsentativität der Entwicklungs- bzw. Trainingsdaten sicherzustellen: Die Daten müssen Zusammensetzung, Struktur und Verhalten des Personenkreises wiederspielen, auf den das Scoring-Verfahren angewandt werden soll.

Selbst das qualitativ hochwertigste Scoring-Verfahren wird bei Anwendung nicht-korrekte oder sogar falsche Ergebnisse liefern, wenn die zur Verfügung stehenden Eingabedaten von mangelhafter Qualität sind. Beispielsweise wurde im Rahmen der Marktstudie bei Kfz-Telematik von Fällen berichtet, in denen falsche Fahrdaten zustande kommen, weil das Kartenmaterial veraltet ist, oder weil es Störungen beim GPS-Signal oder beim Übertragen der Daten gab. Für die Anwendung von Scoring-Verfahren beim Bonitäts-Scoring liegt das Hauptaugenmerk im Hinblick auf Datenqualität auf der korrekten Zuordnung von Information zu Person (Entity Recognition), Korrektheit der Informationen, Aktualität der Informationen und Vollständigkeit.

Handlungsempfehlungen:

1. Eine angemessene Datenqualität bei der Score-Entwicklung ist zu gewährleisten und den Aufsichtsbehörden zu dokumentieren.
2. Scoring-Anbieter und -Nutzer sollten sich selbst verpflichten, ihre Data Governance und speziell ihr Datenqualitätsmanagement nach Maßgabe des Qualitätsleitbildes zu verbessern.
3. In der Anwendung der Verfahren ist die Korrektheit, Vollständigkeit und Aktualität der Daten sicherzustellen.
4. Zur Überprüfung der Datenbasis durch die Verbraucher selbst hat der SVRV bereits in seinem Gutachten „Digitale Souveränität“ die Option eines Daten-Dashboards skizziert, um ein Verbraucher-zentriertes Datenmanagement zu ermöglichen. Der SVRV bekräftigt seine Empfehlung, eine solche Option zu prüfen. Dazu sollten aktuelle Entwicklungen für ein sicheres Identitätsmanagement über Blockchain-basierte Systeme einbezogen werden, die dem Verbraucher die Möglichkeit geben, sicher und eindeutig die eigenen Identitätsdaten zu verwalten.
5. Der SVRV empfiehlt die Qualität der in relevanten Scoring-Verfahren eingesetzten Daten, insbesondere auch die Entity Recognition, zügig zu erforschen und ggf. zu verbessern, falls nötig

durch gesetzliche Vorgaben. Es ist sicherzustellen, dass der beim Anbieter für eine bestimmte Person ermittelte Score tatsächlich der korrekten Person zugeordnet wird. Der Minimierung von Identitätsverwechslungen dient die Verpflichtung, dass Anbieter die betreffende Person über die Tatsache informieren, dass sie gesort wird (siehe Handlungsempfehlung 1). Hier ergibt sich offenkundig ein Konflikt zwischen dem Interesse von Scoring-Anbietern und Nutzern und dem Datenschutz. Der SVRV empfiehlt deswegen, dass die Datenethikkommission der Bundesregierung die Möglichkeiten einer verbesserten Entity Recognition diskutiert und konkrete Empfehlungen erarbeitet.

7. Aufsicht verbessern

Problem:

Die zuständigen Aufsichtsbehörden sind mit ihrer gegenwärtigen personellen und technischen Ausstattung nicht annähernd in der Lage, ihren Prüfaufträgen in den in diesem Gutachten genannten Bereichen und ihrem Umfang tatsächlich nachzukommen. Juristische Kompetenz reicht nicht aus, vielmehr sind insbesondere Informatik- und Statistik-Kompetenz gefragt. Dies gilt in besonderem Maße, wenn den betreffenden Behörden die in diesem Gutachten geforderten zusätzlichen Aufgaben übertragen würden.

Handlungsempfehlungen:

1. Der SVRV empfiehlt der Bundesregierung zu prüfen, wie eine Digitalagentur (siehe auch SVRV-Gutachten „Verbraucherrecht 2.0“) als Kompetenzzentrum agieren kann, das den einzelnen Aufsichtsbehörden hilft, ihre Aufgaben wahrzunehmen. Hierbei wäre zum Beispiel die Form eines Bundesinstituts vorstellbar, das methodische Kompetenzen zur Qualitätssicherung bereithält, die auch für nicht-digitale Zwecke eingesetzt werden könnten.

2. Die zuständigen Aufsichtsbehörden sind strukturell und teilweise durch Verbesserungen der Entgelte von Spezialisten, vor allem im Statistik- und Informatik-Bereich, in die Lage zu versetzen, den beschriebenen Aufgaben nachgehen zu können. Die Entwicklung der BaFin (Bundesanstalt für Finanzdienstleistungsaufsicht) in den letzten Jahren kann als Vorbild dienen. Den zuständigen Aufsichtsbehörden sind die für die oben genannten zusätzlichen Aufgaben sowie für das Testen konkreter Scorings notwendigen erheblichen finanziellen Mittel zur Verfügung zu stellen.
3. Zur raschen Umsetzung der vorliegenden Empfehlungen schlägt der SVRV vor, unverzüglich nach Abschluss der Datenethikkommission auf der Ebene der Bundesregierung eine Task Force (etwa beim Bundeskanzleramt) einzurichten, die auf der Basis bereits existierender Verfahren (z. B. der BaFin) Leitlinien für die Ausarbeitung eines Qualitätsleitbilds vorlegt.

ziellen Gründen Daten aus verschiedensten Lebensbereichen in einer Datenbank und in einem Unternehmen zu vereinen. Dies könnte sogar auf freiwilliger Basis geschehen, wenn sich Menschen dadurch individuelle Vorteile versprechen. Dieses Potenzial existiert ganz konkret bei Versicherungsunternehmen, die verschiedene Versicherungssparten betreiben. Datenbestände könnten hier zusammengelegt und im Zusammenhang ausgewertet werden. Wer gesund lebt, könnte zum Beispiel eine günstigere Berufsunfähigkeits-Versicherung angeboten bekommen.

Im Übrigen zeigt die Bevölkerungsbefragung, dass bis zu einem Drittel der erwachsenen Bevölkerung in Deutschland sich die Zusammenführung von Kfz- und Gesundheits-Scores vorstellen kann; die Befragten geben an, dies sei „gerecht“ oder sie würden davon persönlich profitieren.

Handlungsempfehlungen:

Der SVRV empfiehlt, die Entwicklung in China und in anderen Ländern, wo mit Super-Scores experimentiert wird, sorgfältig zu verfolgen und zu analysieren. Insbesondere ist ein öffentlicher Diskurs über die sich damit verändernden gesellschaftlichen Werte und Gestaltungsoptionen notwendig.

Die Entwicklung von Super-Scores durch internationale kommerzielle Anbieter kann auch in Deutschland relevant werden. Gesetzgeber und Aufsichtsbehörden sollten sich darauf vorbereiten zu prüfen, ob Maßnahmen ergriffen werden sollten und können, damit Super-Scores nicht auch in Deutschland kommerziell angeboten werden.

Der SVRV empfiehlt, konkret zu prüfen, inwieweit bestehende Instrumente der DSGVO (insbesondere Zweckbindungsgrundsatz und Koppelungsverbot) auch zur Verhinderung von Super-Scores einsetzbar sind.

8. Super-Scores verhindern

Problem:

Die chinesische Regierung plant bis 2020 jedem Bürger einen Sozialenkredit-Wert zuzuordnen. Dieser Super-Score wird derzeit mit modernster Computertechnologie von kommerziellen Firmen entwickelt und soll die Vertrauenswürdigkeit in finanzieller, rechtlicher, sozialer und politischer Hinsicht messen. Bürger mit einem hohen Score sollen Vorteile erhalten, wie bei einem Vielflieger-Programm, und jene mit niedrigem Score dürfen etwa keine schnellen Züge benutzen oder ihre Kinder nicht auf die besten Schulen senden. Dieser Super-Score soll das moralische Verhalten der Bevölkerung verbessern, Korruption reduzieren und „Harmonie“ herstellen.

Auch wenn eine staatliche Einführung eines Sozialkredit-Systems in Deutschland nicht absehbar ist, wächst durch Datenhändler und Möglichkeiten der De-Anonymisierung von Datensätzen das Potenzial, aus kommer-

I. Liste der Befragungsteilnehmer (Marktbefragung)

Tabelle I.1: Auskunfteien

	NAME DER WIRTSCHAFTSAUSKUNFTEI	ANMERKUNGEN
1	Creditreform Boniversum GmbH	Teil der Creditreform Unternehmensgruppe; ehemals CEG Creditreform Consumer GmbH.
2	Infoscore Consumer Data GmbH	Gehört zu Arvato Financial Solutions, ein Finanzdienstleister des Bertelsmann-Konzerns.
3	Schufa Holding AG	

Tabelle I.2: Kfz-Versicherer

	NAME DES VERSICHERUNGSANBIETERS	NAME DES TELEMATIK-TARIFS
1	Allianz Deutschland	BonusDrive
2	Generali <i>Antwortete auch für Tochterunternehmen</i>	Generali Mobility
3	AachenMünchener	Telematik-Baustein
4	Cosmos Direkt	BetterDrive
5	HDI	Ryd (ehemals TankTaler)
6	HUK-Coburg <i>Antwortete auch für Tochterunternehmen</i>	Smart Driver
7	HUK24	Smart Driver
8	Itzehoer <i>Antwortete auch für Tochterunternehmen</i>	Secure Drive
9	AdmiralDirekt	Telematik Spar-Option
10	Sparkassen Direktversicherung	Pilotprojekt S-Drive (beendet 2015)

Tabelle I.3 Krankenversicherer

	NAME DES VERSICHERUNGSANBIETERS	TYP
1	AOK – Die Gesundheitskasse für Niedersachsen	GKV
2	AOK – Die Gesundheitskasse in Hessen	GKV
3	AOK Baden-Württemberg	GKV
4	AOK Bayern – Die Gesundheitskasse	GKV
5	AOK Bremen/Bremerhaven	GKV
6	AOK Nordost – Die Gesundheitskasse	GKV
7	AOK NordWest – Die Gesundheitskasse	GKV
8	AOK PLUS – Die Gesundheitskasse für Sachsen und Thüringen	GKV
9	AOK Rheinland/Hamburg – Die Gesundheitskasse	GKV
10	AOK Rheinland-Pfalz/Saarland – Die Gesundheitskasse	GKV
11	AOK Sachsen-Anhalt – Die Gesundheitskasse	GKV
12	atlas BKK ahlmann	GKV
13	Bertelsmann BKK	GKV
14	Betriebskrankenkasse PricewaterhouseCoopers	GKV
15	BKK Deutsche Bank AG	GKV
16	BKK EVM	GKV
17	BKK exklusiv	GKV
18	BKK Faber-Castell & Partner	GKV
19	BKK firmus	GKV
20	BKK Freudenberg	GKV
21	BKK Melitta Plus	GKV
22	BKK PFAFF	GKV

	NAME DES VERSICHERUNGSANBIETERS	TYP
23	BKK ProVita	GKV
24	BKK Public	GKV
25	BKK RWE	GKV
26	BKK Salzgitter	GKV
27	BKK Verkehrsbau Union (VBU)	GKV
28	BKK Voralb HELLER*INDEX*LEUZE	GKV
29	BKK Wirtschaft & Finanzen	GKV
30	BKK ZF & Partner	GKV
31	BKK24	GKV
32	Bosch BKK	GKV
33	Brandenburgische BKK	GKV
34	DAK-Gesundheit	GKV
35	hkk Krankenkasse	GKV
36	IKK classic	GKV
37	IKK Nord	GKV
38	Kaufmännische Krankenkasse – KKH	GKV
39	Knappschaft	GKV
40	Metzinger BKK	GKV
41	mhplus Betriebskrankenkasse	GKV
42	pronova BKK	GKV
43	R+V Betriebskrankenkasse	GKV
44	Siemens-Betriebskrankenkasse (SBK)	GKV

	NAME DES VERSICHERUNGSANBIETERS	TYP
45	Südzucker BKK	GKV
46	Techniker Krankenkasse	GKV
47	VIACTIV Krankenkasse	GKV
48	Allianz Deutschland AG	PKV
49	Barmenia Krankenversicherung AG	PKV
50	Bayerische Beamtenkrankenkasse AG	PKV
51	Krankenversorgung der Bundesbahnbeamten	PKV
52	Central Krankenversicherung AG	PKV
53	Continental Krankenversicherung AG	PKV
54	Debeka Krankenversicherungsverein AG	PKV
55	HanseMercur Krankenversicherung AG	PKV
56	HUK-COBURG-Krankenversicherung AG	PKV
57	LIGA Krankenversicherung katholischer Priester VVaG	PKV
58	Nürnberger Krankenversicherung AG	PKV
59	Pax-Familienfürsorge Krankenversicherung AG im Raum der Kirchen	PKV
60	SONO Krankenversicherung AG	PKV
61	Süddeutsche Krankenversicherung AG	PKV
62	UKV Union Krankenversicherung AG	PKV

II. Fragebögen (Marktbefragung)

Umfrage zu Bonitäts-Scoring von Auskunfteien

Unsere Recherche hat ergeben, dass Ihr Unternehmen Bonitätsprüfungen von Privatpersonen anbietet. Wir würden uns freuen, wenn Sie dazu die nachfolgenden Fragen beantworten würden.

1. FRAGEN ZUM GESCHÄFTSMODELL

- 1.1 Was zeichnet das Geschäftsmodell Ihres Unternehmens besonders aus (z. B. in Hinblick auf eingesetzte Technik, Datengrundlage, statistisches Modell, spezifische Scores für einzelne Branchen/Unternehmen etc.)?

- 1.2 Gibt es Voraussetzungen, die Geschäftskunden erfüllen müssen, um von Ihrem Unternehmen Auskunft über die Bonität eines Verbrauchers zu erhalten? Falls ja, welche?

- 1.3 Bitte nennen Sie uns den Namen des Bonitäts-Scores, den Ihr Unternehmen anbietet. Falls Ihr Unternehmen mehrere Bonitäts-Scores anbietet, nennen Sie uns bitte den Namen des gebräuchlichsten Scores bzw. des Scores, auf den Sie sich im Folgenden beziehen.

1. FRAGEN ZUM GESCHÄFTSMODELL

1.4 Mit welcher Maßnahme können Verbraucher ihren Bonitäts-Score am ehesten verbessern?

1.5 In welcher Form teilen Sie Privatpersonen das Ergebnis Ihres Bonitäts-Scorings mit?
Mehrfachnennungen sind möglich.

- Punktestand
- Wahrscheinlichkeitswert
- Symbole (z. B. Ampel, Farbskalen)
- Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher
- Sonstiges: _____

1.6 In welcher Form teilen Sie Geschäftskunden das Ergebnis Ihres Bonitäts-Scorings mit?
Mehrfachnennungen sind möglich.

- Punktestand
- Wahrscheinlichkeitswert
- Symbole (z. B. Ampel, Farbskalen)
- Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher
- Sonstiges: _____

1.7 Auf welcher Skala (z. B. 1–100) bewegt sich der Bonitäts-Score, den Ihr Unternehmen errechnet?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

Verbraucher und die allgemeine Öffentlichkeit können an der Score-Berechnung und der Erhebung der ihr zugrunde liegenden Daten ein Interesse haben. Welche Auskünfte geben Sie auf die folgenden Fragen:

2.1 Aus welchen Quellen bezieht Ihr Unternehmen die Daten zur Berechnung des Bonitäts-Scores?

2.2 Bitte geben Sie an, ob die Berechnung des Bonitäts-Scores in Ihrem Unternehmen selbst oder durch einen externen Dienstleister erfolgt:

- Im Unternehmen selbst
- Bei einem externen Dienstleister, und zwar: _____
- Bitte geben Sie auch den Sitz des externen Dienstleisters an: _____

2.3 Wie viele Merkmale fließen immer in die Berechnung des Bonitäts-Scores ein?

2.4 Welche Merkmale fließen immer in die Berechnung des Bonitäts-Scores ein?

2.5 Welche Merkmale können darüber hinaus (z. B. für branchenspezifische Scores) in die Berechnung des Bonitäts-Scores einfließen?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

Verbraucher und die allgemeine Öffentlichkeit können an der Score-Berechnung und der Erhebung der ihr zugrunde liegenden Daten ein Interesse haben. Welche Auskünfte geben Sie auf die folgenden Fragen:

2.6 Fließen Daten zur Wohngegend (Geodaten) in die Berechnung des Bonitäts-Scores mit ein?

- Ja, immer
- Ja, unter folgenden Umständen: _____
- Falls ja, welche? _____
- Nein, nie

2.7 Fließen Daten aus Onlineverhalten (z. B. aus sozialen Netzwerken) in die Berechnung des Bonitäts-Scores mit ein?

- Ja, immer
- Ja, unter folgenden Umständen: _____
- Nein, nie

2.8 Anhand welcher Kriterien wird entschieden, welche Merkmale für die Berechnung des Bonitäts-Scores relevant und aussagekräftig sind?

2.9 Anhand welcher Kriterien wird entschieden, mit welcher Gewichtung Merkmale in die Berechnung des Bonitäts-Scores einfließen?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

Verbraucher und die allgemeine Öffentlichkeit können an der Score-Berechnung und der Erhebung der ihr zugrunde liegenden Daten ein Interesse haben. Welche Auskünfte geben Sie auf die folgenden Fragen:

2.10 Gibt es individuelle Merkmale, die nicht direkt erfasst oder gemessen, sondern durch andere Merkmale approximiert werden?

- Nein
- Ja, nämlich: _____

2.11 Welche Schätzmethoden liegen dem Scoring zugrunde? Mehrfachnennungen sind möglich.

- Logistische Regression
- Clusterbildende Methoden (z. B. kMeans)
- Entscheidungsbäume
- Ensemblemethoden (z. B. Boosting, Random Forest)
- Neuronale Netzwerke/Deep Learning
- Sonstiges: _____

2.12 Wird für die Schätzmethode ein Zielmerkmal (abhängige Variable) verwendet?

Falls ja, benennen Sie bitte das Zielmerkmal.

2.13 Wie wird die Güte des verwendeten Schätzmodells gemessen?

- Area Under the Curve (AUC)
- Gini-Koeffizient
- Quadratischer Fehler/Brier Score
- KS-Statistik (= Kolmogorov-Smirnov Test)
- F-Measure
- Lift
- Sonstiges: _____

2.14 Wie hoch ist der Wert des unter 2.13 genannten Gütemaßes?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

Verbraucher und die allgemeine Öffentlichkeit können an der Score-Berechnung und der Erhebung der ihr zugrunde liegenden Daten ein Interesse haben. Welche Auskünfte geben Sie auf die folgenden Fragen:

2.15 Welche Maßnahmen werden ergriffen, um die Güte des Modells aufrechtzuerhalten? Aus welchem Anlass wird das Schätzmodell verändert?

2.16 Welche Maßnahmen werden ergriffen, um die Güte des Modells zu verbessern?

2.17 Welche Daten, die nicht dem Zweck der Score-Berechnung dienen, werden zusätzlich über Verbraucher gespeichert?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

Verbraucher und die allgemeine Öffentlichkeit können an der Score-Berechnung und der Erhebung der ihr zugrunde liegenden Daten ein Interesse haben. Welche Auskünfte geben Sie auf die folgenden Fragen:

2.18 Nehmen Sie bei der Entwicklung und Evaluierung Ihres Scoring-Verfahrens externe wissenschaftliche Beratung in Anspruch? Falls ja, in welcher Form?

3. FRAGEN ZU AUFSICHT UND AUSKUNFTSRECHTEN

3.1 Wie erfahren Verbraucher, welche Daten über sie in die Berechnung ihres Bonitäts-Scores eingeflossen sind?

3.2 Welche Maßnahmen werden ergriffen, um die Aktualität und Richtigkeit der von Ihrem Unternehmen zur Berechnung der Bonitäts-Scores verwendeten Daten zu gewährleisten?

3. FRAGEN ZU AUFSICHT UND AUSKUNFTSRECHTEN

3.3 Welche Verfahren zur Fehlerkorrektur der Daten, die in die Score-Berechnung einfließen, stehen Verbrauchern zur Verfügung?

3.4 Bezüglich welcher Merkmale weisen Verbraucher Ihr Unternehmen am häufigsten auf fehlerhafte Daten hin?

3.5 Halten Sie es für möglich, dass die Ermittlung des Bonitäts-Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten kann? Falls ja, wie begegnen Sie diesem Problem?

3.6 Welche Aufsichtsbehörden sind für die Prüfung Ihres Scoring-Verfahrens zuständig? Wie oft und aus welchen Anlässen wird dieses überprüft?

4. WEITERFÜHRENDE EINSCHÄTZUNGEN

- 4.1 Halten Sie die Erweiterung der Datenbasis Ihres Scoring-Verfahrens um Merkmale aus anderen Bereichen (z. B. riskantes Fahrverhalten oder Gesundheitsverhalten) für sinnvoll, um die Prognosequalität zu verbessern?

- 4.2 Sehen Sie Bedarf, den Gesetzes- und Regulierungsrahmen in Bezug auf Bonitäts-Scoring anzupassen (z. B. im Bereich des Datenschutzrechts)? Falls ja, in welcher Hinsicht?

- 4.3 Wie schätzen Sie die Zukunft von Bonitäts-Scoring-Verfahren ein? Gibt es Trends z. B. aus dem Ausland, die Sie für die Weiterentwicklung Ihres Geschäftsmodells für relevant halten?

Weitere Anmerkungen

Kontaktperson für eventuelle Rückfragen

Name: _____

Telefonnummer: _____

Emailadresse: _____

Vielen Dank für Ihre Teilnahme!

Umfrage zu Telematik-Tarifen in der Kfz-Versicherung

Unsere Recherche hat ergeben, dass Ihr Unternehmen in der Kfz-Versicherung eine Telematik-Option anbietet. Wir würden uns freuen, wenn Sie dazu die nachfolgenden Fragen beantworten würden.

1. FRAGEN ZUM GESCHÄFTSMODELL

- 1.1 Was zeichnet den Telematik-Tarif in der Kfz-Versicherung Ihres Unternehmens besonders aus (z. B. in Hinblick auf eingesetzte Technik, Datengrundlage, Prämiengestaltung etc.)?

- 1.2 Gibt es Voraussetzungen, die Kunden erfüllen müssen, um bei Ihrem Unternehmen eine Kfz-Versicherung mit Telematik-Tarif abschließen zu können? Falls ja, welche? Mehrfachnennungen sind möglich.

- Altersvoraussetzungen, und zwar: _____
- Jahre seit Führerscheinwerb, und zwar: _____
- Technische Voraussetzungen, und zwar: _____
- Sonstiges: _____

- 1.3 Gibt es eine Zielgruppe, die mit dem Telematik-Tarif ihres Unternehmens besonders angesprochen werden soll? Falls ja, welche?

1. FRAGEN ZUM GESCHÄFTSMODELL

- 1.4 Wie groß ist der Anteil Ihrer Kunden in der Kfz-Versicherung, die an dem Telematik-Tarif teilnehmen?
Zeichnen sich die teilnehmenden Kunden im Vergleich zu nicht-teilnehmenden Kunden durch besondere Merkmale (z. B. Alter, Geschlecht) aus?

- 1.5 Mit welchen in Aussicht gestellten Effekten motivieren Sie Kunden zur Teilnahme an Ihrem Telematik-Tarif?

- 1.6 Auf welcher Skala (z. B. 1–100) bewegt sich der Score, der im Rahmen des Telematik-Tarifs für einen Kunden berechnet wird?

- 1.7 In welcher Form teilen Sie Kunden das Ergebnis Ihres Scorings mit? Mehrfachnennungen sind möglich.

- Punktestand
- Wahrscheinlichkeitswert
- Symbole (z. B. Ampel, Farbskalen)
- Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher
- Zielmarken (z. B. „Gold-Status“)
- Sonstiges: _____

- 1.8 Mit welcher Maßnahme können Kunden ihren Score am ehesten verbessern?

1. FRAGEN ZUM GESCHÄFTSMODELL

1.9 Welche potenziellen Vorteile werden den Kunden gewährt, wenn Sie an dem Telematik-Tarif Ihres Unternehmens teilnehmen? Mehrfachnennungen sind möglich.

- Rückzahlung von Versicherungsprämien
- Günstigere Versicherungsprämien im Folgejahr
- Geldleistungen
- Sachleistungen
- Gutscheine/Rabatte bei Partnerunternehmen
- Sonstiges: _____

1.10 Dient der Score als das einzige Entscheidungskriterium zur Gewährung der unter 1.9 genannten Vorteile?

1.11 Wie und wie oft wird der aktuelle Score dem Kunden mitgeteilt?

1.12 Bieten Sie Ihren Kunden die Möglichkeit, ihren Score digital, z. B. über eine App, zu erfahren?

- Ja
- Nein

1.13 Bieten Sie Ihren Kunden die Möglichkeit, ihren Score online zu veröffentlichen?

- Ja
- Nein

1. FRAGEN ZUM GESCHÄFTSMODELL

1.14 Bietet Ihr Unternehmen die Möglichkeit, dass Kunden ihre jeweiligen Scores in Form eines Rankings miteinander vergleichen können (z. B. über eine App oder ein Onlineportal)?

- Ja
- Nein

1.15 Ändert sich die Schadenshäufigkeit Ihrer Kunden durch die Teilnahme an Ihrem Telematik-Tarif?
Wie stellen Sie dies fest?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

2.1 Welche Technik wird verwendet, um die Daten zur Berechnung der Scores zu erheben?
Mehrfachnennungen sind möglich.

- Telematik-Stecker für den Zigarettenanzünder
- OBD-Telematik-Stecker
- Ins Auto fest eingebaute Telematik-Box
- App auf dem Smartphone
- Sonstiges: _____

2.2 Zieht Ihr Unternehmen Daten aus weiteren Quellen zur Berechnung der Scores heran
(z. B. digitales Kartenmaterial, Social Media Daten, Datenbanken wie GfK, Nielsen etc.)? Falls ja, welche?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

2.3 Bitte geben Sie an, ob die Berechnung des Scores in Ihrem Unternehmen selbst oder durch einen externen Dienstleister erfolgt:

- Im Unternehmen selbst
 Bei einem externen Dienstleister, und zwar: _____

Bitte geben Sie auch den Sitz des externen Dienstleisters an: _____

2.4 Wie viele Merkmale fließen immer in die Berechnung des Scores ein?

2.5 Welche Merkmale fließen immer in die Berechnung des Scores ein?

2.6 Welche Merkmale können darüber hinaus in die Berechnung des Scores einfließen?

2.7 Anhand welcher Kriterien wird entschieden, welche Merkmale für die Berechnung des Scores relevant und aussagekräftig sind?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

2.8 Anhand welcher Kriterien wird entschieden, mit welcher Gewichtung Merkmale in die Berechnung des Scores einfließen?

2.9 Gibt es individuelle Merkmale, die nicht direkt erfasst oder gemessen, sondern durch andere individuelle Merkmale approximiert werden?

- Nein
- Ja, nämlich: _____

Wird der Score mit statistischen Schätzmodellen berechnet, beantworten Sie uns bitte die folgenden Fragen. Ansonsten weiter mit Frage 3.1.

2.10 Welche Schätzmethoden liegen dem Scoring zugrunde? Mehrfachnennungen sind möglich.

- Logistische Regression
- Clusterbildende Methoden (z. B. kMeans)
- Entscheidungsbäume
- Ensemblemethoden (z. B. Boosting, Random Forest)
- Neuronale Netzwerke/Deep Learning
- Sonstiges: _____

2.11 Wird für die Schätzmethode ein Zielmerkmal (abhängige Variable) verwendet? Falls ja, benennen Sie bitte das Zielmerkmal.

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

2.12 Wie wird die Güte des verwendeten Schätzmodells gemessen?

- Area Under the Curve (AUC)
- Gini-Koeffizient
- Quadratischer Fehler/Brier Score
- KS-Statistik (= Kolmogorov-Smirnov Test)
- F-Measure
- Lift
- Sonstiges: _____

2.13 Wie hoch ist der Wert des unter 2.12 genannten Gütemaßes?

2.14 Welche Maßnahmen werden ergriffen, um die Güte des Modells aufrechtzuerhalten? Aus welchem Anlass wird das Schätzmodell verändert?

2.15 Welche Maßnahmen werden ergriffen, um die Güte des Modells zu verbessern?

2. FRAGEN ZU DATENERHEBUNG UND SCORE-BERECHNUNG

2.16 Nehmen Sie bei der Entwicklung und Evaluierung Ihres Scoring-Verfahrens externe wissenschaftliche Beratung in Anspruch? Falls ja, in welcher Form?

3. FRAGEN ZU AUFSICHT UND AUSKUNFTSRECHTEN

3.1 Wie erfahren Kunden Ihres Unternehmens, welche Daten über sie in die Berechnung ihres Scores einfließen?

3.2 Welche Maßnahmen werden ergriffen, um die Aktualität und Richtigkeit der von Ihrem Unternehmen zur Berechnung der Scores verwendeten Daten zu gewährleisten?

3. FRAGEN ZU AUFSICHT UND AUSKUNFTSRECHTEN

3.3 Welche Verfahren zur Fehlerkorrektur der Daten, die in die Score-Berechnung einfließen, stehen Kunden zur Verfügung?

3.4 Bezüglich welcher Merkmale weisen Verbraucher Ihr Unternehmen am häufigsten auf fehlerhafte Daten hin?

3.5 Halten Sie es für möglich, dass die Ermittlung des Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten kann? Falls ja, wie begegnen Sie diesem Problem?

3. FRAGEN ZU AUFSICHT UND AUSKUNFTSRECHTEN

3.6 Unterliegen Ihre Scoring-Verfahren staatlicher Aufsicht? Falls ja, welche Aufsichtsbehörden sind für die Prüfung zuständig? Wie oft und aus welchen Anlässen wird überprüft?

4. WEITERFÜHRENDE EINSCHÄTZUNGEN

4.1 Welche Vorteile sehen Sie im Einsatz von Telematik-Tarifen für Ihre Branche?

4.2 Welche Nachteile sehen Sie im Einsatz von Telematik-Tarifen für Ihre Branche?

4. WEITERFÜHRENDE EINSCHÄTZUNGEN

4.3 Wie würden Sie die Akzeptanz Ihrer Kunden einschätzen, für schlechtes Fahrverhalten höhere Prämien in der Kfz-Versicherung zahlen zu müssen?

4.4 Halten Sie die Erweiterung der Datenbasis Ihres Scoring-Verfahrens um Merkmale aus anderen Bereichen (z. B. Kreditwürdigkeit oder Gesundheitsverhalten) für sinnvoll, um die Prognosequalität zu verbessern?

4.5 Sehen Sie Bedarf, den Gesetzes- und Regulierungsrahmen in Bezug auf Telematik-Tarife anzupassen (z. B. im Bereich des Datenschutzrechts)? Falls ja, in welcher Hinsicht?

4.6 Wie beurteilen Sie den Einfluss von eCall-Systemen auf den Telematik-Bereich in Ihrer Branche?

4. WEITERFÜHRENDE EINSCHÄTZUNGEN

4.7 Wie schätzen Sie die Zukunft von Scoring-Verfahren in Ihrer Branche ein? Gibt es Trends z. B. aus dem Ausland, die Sie für die Weiterentwicklung Ihrer Kfz-Versicherung für relevant halten?

Weitere Anmerkungen

Kontaktperson für eventuelle Rückfragen

Name: _____

Telefonnummer: _____

Emailadresse: _____

Vielen Dank für Ihre Teilnahme!

Umfrage zu Bonussystemen in der gesetzlichen und privaten Krankenversicherung

Wir würden uns freuen, wenn Sie die nachfolgenden Fragen zu den Themen Bonusprogramme in der Krankenversicherung und verhaltensbasierte Bestandteile von Bonusprogrammen beantworten würden.

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.1 Bieten Sie im Rahmen Ihrer Krankenversicherung ein Bonusprogramm (bzw. ein Präventions- oder Vorteilsprogramm) an?

- Ja
- Nein (→ weiter mit Frage 1.21)

1.2 Gibt es eine Zielgruppe, die mit dem Bonusprogramm Ihrer Krankenversicherung besonders angesprochen werden soll? Falls ja, welche?

1.3 Wie groß ist der Anteil Ihrer Kunden in der Krankenversicherung, die am Bonusprogramm teilnehmen? Zeichnen sich die teilnehmenden Kunden im Vergleich zu nicht-teilnehmenden Kunden durch besondere Merkmale (z. B. Alter, Geschlecht) aus?

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.4 Welche potenziellen Vorteile haben Kunden, wenn Sie an Ihrem Bonusprogramm teilnehmen?
Mehrfachnennungen sind möglich.

- Rückzahlung von Versicherungsprämien/-beiträgen
- Niedrigere Versicherungsprämien/-beiträge im Folgejahr
- Geldleistungen
- Sachleistungen
- Gutscheine/Rabatte bei Partnerunternehmen
- Gutscheine/Rabatte für medizinische Behandlungen
- Sonstiges: _____

1.5 Mit welchen Aktivitäten können Kunden Bonuspunkte sammeln? Mehrfachnennungen sind möglich.

- Früherkennungs- bzw. Vorsorgeuntersuchungen
- Teilnahme an Gesundheitskursen, nämlich: _____
- Mitgliedschaft in einem Sportverein, Fitnessstudio oder im Betriebs-/Hochschulsport
- Teilnahme an Sportveranstaltungen, nämlich: _____
- Sportabzeichen
- Schutzimpfungen
- Erfassung von Daten mittels Fitness-Apps und /oder Wearables
- Blutspende
- Soziales Engagement, nämlich: _____
- Sonstiges: _____

1.6 Falls Kunden aufgrund gemeinnützigen Verhaltens (z. B. Blutspende, soziales Engagement) Bonuspunkte sammeln können: Wie begründen Sie die Einbeziehung solcher Aktivitäten in Ihr Bonusprogramm?

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.7 Anhand welcher Kriterien wird entschieden, mit welchen Aktivitäten Kunden Bonuspunkte sammeln können?

1.8 Anhand welcher Kriterien wird jeweils entschieden, wie viele Punkte Kunden mit einer Aktivität sammeln können?

1.9 Auf welcher Skala (z. B. 1–100) bewegt sich der Punktestand, der im Rahmen des Bonusprogrammes für einen Kunden berechnet wird?

1.10 In welcher Form teilen Sie Kunden ihren aktuellen Stand im Bonusprogramm mit?

Mehrfachnennungen sind möglich.

- Punktestand
- Wahrscheinlichkeitswert
- Symbole (z. B. Ampel, Farbskalen)
- Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher
- Zielmarken (z. B. „Gold-Status“)
- Sonstiges: _____

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.11 Mit welchen in Aussicht gestellten Effekten motivieren Sie Kunden zur Teilnahme am Bonusprogramm?

1.12 Mit welcher Maßnahme können Kunden ihren Punktestand am ehesten verbessern?

1.13 Bieten Sie Ihren Kunden die Möglichkeit, ihren Punktestand im Bonusprogramm digital, z. B. über eine App, einzusehen und absolvierte Aktivitäten zu erfassen?

- Ja
 Nein

1.14 Bieten Sie Ihren Kunden die Möglichkeit, ihren Punktestand online zu veröffentlichen?

- Ja
 Nein

1.15 Bieten Sie die Möglichkeit, dass Kunden ihre jeweiligen Punktestände in Form eines Rankings miteinander vergleichen können (z. B. über eine App oder ein Onlineportal)?

- Ja
 Nein

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.16 Ändert sich der Gesundheitszustand Ihrer Kunden durch die Teilnahme an Ihrem Bonusprogramm? Wie stellen Sie dies fest?

1.17 Nehmen Sie bei der Entwicklung und Evaluierung Ihres Bonusprogrammes externe wissenschaftliche Beratung in Anspruch? Falls ja, in welcher Form?

Sofern Versicherte im Rahmen Ihres Bonusprogramms durch Aufzeichnung und/oder Übermittlung von Bewegungsdaten mittels Fitness-/Gesundheits-Apps (z. B. Anzahl getätigter Schritte) Punkte sammeln können, beantworten Sie uns bitte die folgenden Fragen.

→ Falls nicht, weiter mit Frage 1.21.

1.18 Welche mittels Fitness-/Gesundheits-Apps erfassten Daten werden für das Bonusprogramm verwendet?

1. FRAGEN ZU BESTEHENDEN BONUSPROGRAMMEN

1.19 Wie detailliert erhält Ihre Versicherung Einblick in die Gesundheits-/Bewegungsdaten Ihrer Kunden (z. B. genaue Schrittzahl oder lediglich Erreichen einer Zielvorgabe)?

1.20 Arbeiten Sie bei der Erfassung von Daten mittels Fitness-/Gesundheits-Apps mit Drittanbietern zusammen? Falls ja, welche?

1.21 Bezuschusst Ihre Versicherung die Anschaffung von Geräten, die Aktivitätsdaten erfassen und weitergeben (wie z. B. Fitnessarmbänder)? Falls ja, unter welchen Voraussetzungen und in welchem Umfang?

2. WEITERFÜHRENDE EINSCHÄTZUNGEN

- 2.1 Sollte es aus Sicht Ihrer Krankenversicherung die Möglichkeit geben, bei der Prämien-gestaltung von Kranken-versicherungstarifen bzw. bei der Ausgestaltung von Bonusprogrammen neben gesundheitsförderlichen auch gesundheitsschädliche Verhaltensweisen von Versicherten zu berücksichtigen?

- 2.2 Sehen Sie Bedarf, den Gesetzes- und Regulierungsrahmen in Bezug auf Bonusprogramme oder verhaltensbasierte Tarife anzupassen (z. B. im Bereich des Datenschutz- oder Gesundheitsrechts)? Falls ja, in welcher Hinsicht?

- 2.3 Sollten elektronischen Patientenakten und die damit potenziell zur Verfügung stehenden Gesundheitsdaten von Versicherten für den Bereich Bonusprogramme und/ oder verhaltensbasierte Tarife in der Kranken-versicherung nutzbar sein?

- 2.4 Welche Vorteile sehen Sie im potenziellen Einsatz von verhaltensbasierten Tarifen für Ihre Branche?

2. WEITERFÜHRENDE EINSCHÄTZUNGEN

2.5 Welche Nachteile sehen Sie im potenziellen Einsatz von verhaltensbasierten Tarifen für Ihre Branche?

2.6 Wie schätzen Sie die Zukunft von verhaltensbasierten Tarifen in Ihrer Branche ein? Gibt es Trends (z. B. aus dem Ausland), die Sie für die Weiterentwicklung Ihrer Krankenversicherung für relevant halten?

Weitere Anmerkungen

Kontaktperson für eventuelle Rückfragen

Name: _____

Telefonnummer: _____

Emailadresse: _____

Vielen Dank für Ihre Teilnahme!

III. Darstellung der Befragungsergebnisse (Marktbefragung)

1. Auskunfteien

Geschäftsmodell und Ausgestaltung der Bonitätsauskunft

Die drei Auskunfteien, die an der Befragung teilgenommen haben, bieten Bonitätsauskünfte über Privatpersonen an. Diese können neben branchenübergreifenden Gesamt-Scores auch in Form von branchenspezifischen Scores wie z. B. einem Score für Banken oder einem Score für Telekommunikationsunternehmen mitgeteilt werden. Voraussetzung dafür, dass Geschäftskunden Bonitäts-Scores über Privatpersonen erhalten, ist bei allen ein bestehender Vertrag mit der Auskunftei und gemäß den gesetzlichen Vorgaben der Nachweis eines berechtigten Interesses. Alle drei Auskunfteien nennen bei den Voraussetzungen für die Auskunftserteilung explizit das Bundesdatenschutzgesetz (BDSG) bzw. die Datenschutz-Grundverordnung (DSGVO).

Die Skalen, auf denen sich die Bonitäts-Scores der an der Befragung teilnehmenden Auskunfteien bewegen, weisen einige Unterschiede auf. Bei einer Auskunftei gibt es je nach Score eine Skala von 0-1000 Punkten oder 100-600 Punkten. Die zweite Auskunftei bildet den Score auf einer Skala von ebenfalls 0-1000 Punkten ab, wobei der „tatsächlich genutzte Bereich“ zwischen 250 und 800 Punkten liege. Die dritte Auskunftei verortet die Scores auf einer Skala von 1-999 oder 1-9999 Punkten. Alle Skalen sind so zu interpretieren: bei Null liegt keinerlei Kreditwürdigkeit vor, beim Maximalwert völlige Kreditwürdigkeit (wird in der Realität nie erreicht), dazwischenliegende Werte werden „linear“ in Richtung weniger oder mehr Kreditwürdigkeit interpretiert.

Die Bonitätseinschätzung bzw. die Ergebnisse des Scorings werden auf unterschiedliche Art dargestellt und mitgeteilt. Wenn Privatpersonen ihre eigene Bonität anfragen, bekommen sie diese bei allen drei Auskunfteien als Punktestand und als Wahrscheinlichkeitswert (für das Auftreten eines Negativmerkmals) mitgeteilt. Eine der Auskunfteien visualisiert das Ergebnis zusätzlich noch mit Symbolen und spricht eine schriftliche Empfehlung an den Verbraucher aus. Wie genau eine solche Empfehlung aussieht, wird aber nicht beschrieben.

Datengrundlage

Alle drei an der Befragung teilnehmenden Auskunfteien geben an, Daten aus öffentlich zugänglichen Registern und Verzeichnissen sowie amtliche Bekanntmachungen zu nutzen. Inkassomeldungen, Insolvenzinformationen und Schuldnerverzeichnisse liefern die für das Bonitäts-Scoring notwendigen Negativinformationen zum Zahlungsverhalten. Zwei der drei Auskunfteien geben zudem an, Daten von Vertragspartnern zu erhalten. Um was für Daten es sich konkret handelt, wird nicht erläutert. In einem Fall handelt es sich um eine Haushaltsdatenbank eines Kooperationspartners (wahrscheinlich ein Adress- und/oder Datenhändler), im anderen Fall um verschiedene Unternehmen, wie u. a. Banken, Sparkassen, Warenhändler, Telekommunikationsgesellschaften oder Energieversorger. Es zeigt sich also, dass die Quellenlage der Daten für die einzelnen Auskunfteien unterschiedlich ist.

Bei der Frage nach der Anzahl der verwendeten Merkmale können die Auskunfteien keine konkreten Angaben machen. Je nach Vertragspartner variiere die Anzahl. Die Anbieter sprechen in diesem Zusammenhang von „Scorekarten“, die sie anbieterspezifisch erstellen. Eine Auskunftei nennt als Minimum der verwendeten Merkmale die Zahl Acht, eine andere Auskunftei die Zahl 25. Die dritte Auskunftei gibt an, die Frage nicht beantworten zu können, da sie je nach Vertragspartner variiert.

Auch bei der Frage, welche Merkmale immer bzw. welche Merkmale darüber hinaus manchmal in die Score-Berechnung einfließen, geben die Unternehmen wenig Auskunft. Zwei Auskunfteien verweisen darauf, dass grundsätzlich alle im Datenbestand vorhandenen Daten zur Score-Wert-Bildung herangezogen werden können. Eine Auskunftei führt einige Merkmale auf: „Informationen zum Zahlungsverhalten, Zahlungserfahrungen im unmittelbaren Wohnumfeld, Alter, Geschlecht und Umzugsverhalten...“. Bei dieser Auskunftei würden alle branchenspezifischen Scores auf den gleichen Merkmalen gebildet, diese aber unterschiedlich gewichtet.

Beim Geosoring wird die Kreditwürdigkeit eines Verbrauchers anhand seines Wohnorts beurteilt. Es wird das Zahlungsverhalten in der unmittelbaren Wohngegend analysiert, um Rückschlüsse auf das individuelle Zahlungsverhalten eines einzelnen Verbrauchers zu ziehen. Die Frage, ob für das Scoring Geodaten ein-

fließen, beantwortet nur eine Auskunft mit „Ja“. Dies sei allerdings nur sehr selten der Fall, nämlich nur dann, wenn keine kreditrelevanten Informationen zur Person im Datenbestand zum Zeitpunkt der Anfrage vorliegen. Die anderen beiden Auskunfteien verneinen den Einsatz von Daten zur Wohngegend, obgleich eine von ihnen an anderer Stelle angegeben hatte, Zahlungsverhalten im unmittelbaren Wohnumfeld würde immer berücksichtigt (s.o.), was Geoscoring gleichgesetzt werden kann.

Kein Unternehmen gibt an, Daten aus dem Onlineverhalten von Verbrauchern (z. B. aus sozialen Netzwerken) einzusetzen.

Schließlich wurde noch erfragt, welche Daten, die nicht dem Zweck der Score-Berechnung dienen, zusätzlich über Verbraucher gespeichert werden. Alle Auskunfteien nennen die nach Datenschutzgesetz vorgeschriebene Protokollierung der Anfragen und der Zahl der erteilten Auskünfte. Eine Auskunft speichert Namen und Wohnorte von Verbrauchern, ohne diese Daten zum Scoring zu verwenden.

Methodik der Scoring-Verfahren

Es folgten einige Fragen zu den Details der verwendeten statistischen Verfahren. Die Auskunfteien in der Befragung verwenden alle das Verfahren der logistischen Regression. Eine der drei gab aber zusätzlich auch noch die von uns zur Auswahl vorgegebenen Schätzmethoden Entscheidungsbäume, Ensemblemethoden und Neuronale Netzwerke/Deep Learning an. Anhand der Gespräche, die geführt wurden, ist zu vermuten, dass diese Verfahren gegenwärtig getestet werden.

Darüber hinaus wurde das Zielmerkmal (die abhängige Variable) im Schätzmodell erfragt. Zwei Unternehmen nennen das Auftreten eines Negativmerkmals innerhalb eines bestimmten Zeitraums. Konkret: „Zielmerkmal ist die sog. externe Performance. D. h., es wird geprüft, ob innerhalb von neun Monaten nach der Anfrage ein (neues) Negativ-Merkmal zum Zahlungsverhalten der Person bekannt wird“. Die dritte Auskunft beantwortete die Frage nicht.

Die Güte des verwendeten Schätzmodells messen alle drei Auskunfteien in der Befragung mit dem Gini-Koeffizienten und eine zusätzlich noch mit dem Kolmogorov-Smirnov Test⁹⁸. Einen konkreten Wert für den minimal notwendigen Wert des Gütemaßes nennt eine Auskunft nicht, da dieser je nach Vertragspartner unterschiedlich sei. Die anderen beiden werden etwas konkreter: ein Unternehmen gibt an, dass je nach Score-Produkt und Branche eine Güte zwischen 60 bis 75 % im Gini-Koeffizient erreicht würde. Beim anderen erziele der Gini-Koeffizient je nach Portfolio eine Güte zwischen 60 und 80 %. Diese Kennziffern machen freilich keine Aussagen zu gruppenspezifischen Unterschieden der Prognosegüte, d. h. dem Ausmaß der in verschiedenen Dimensionen erreichten Fairness (siehe B.VI).

Weiterhin wurde erfragt, anhand welcher Kriterien entschieden wird, welche Merkmale für die Berechnung des Bonitäts-Scores relevant und aussagekräftig sind sowie mit welcher Gewichtung Merkmale in die Berechnung einfließen. Alle drei Unternehmen machen keine konkreten Angaben, sondern verweisen auf die grundsätzliche Verwendung mathematisch-statistischer Verfahren. Zwei Auskunfteien spezifizieren, dass die statistische Relevanz eines Merkmals, abgelesen an den Ergebnissen von Signifikanztests, ausschlagend sei. Eine dieser Auskunfteien führt noch zwei weitere Punkte an, die darüber entscheiden, ob ein Merkmal für das Scoring verwendet wird: zum einen eine „datenschutzrechtliche Beurteilung“, zum anderen die „Verfügbarkeit des Merkmals im operativen Betrieb“.

Bei der Frage nach der Gewichtung verweisen wiederum alle auf die Verwendung statistischer Verfahren, speziell die logistische Regression. Eine Auskunft führt weiterhin an, die Modellbildung erfolge „auf Basis von Erfahrungswerten bzgl. der Stabilität der Parameter“.

Als nächstes wurde erfragt, welche Maßnahmen ergriffen werden, um die Güte der Modelle aufrechtzuerhalten und aus welchem Anlass ein Schätzmodell verändert wird. Es ist nicht überraschend, dass alle Anbieter eine regelmäßige Validierung durchführen. Eine Auskunft „führt ein regelmäßiges Reporting (mindestens quar-

⁹⁸ Der Gini-Koeffizient ist ein statistisches Maß zur Messung der Ungleichheit einer Verteilung. Der Kolmogorov-Smirnov Test prüft zwei Wahrscheinlichkeitsverteilungen auf Übereinstimmung. Beide Maße eignen sich auszudrücken, wie sehr ein statistisches Verfahren geeignet ist, Verbraucher hinsichtlich ihres Risikos (Kreditausfallrisiko, Unfallwahrscheinlichkeit, Krankheitswahrscheinlichkeit usw.) zu klassifizieren (siehe B.IV. für eine ausführliche Beschreibung der Gütemaße).

talsmäßig) durch, das die vorhandenen Scorekarten hinsichtlich Stabilität und Trennschärfe überprüft.“ Bei einer anderen „erfolgt eine periodische Validierung der Scorekarten im 3-Jahres-Rhythmus. Außerdem werden in Kundenprojekten die Scores laufend verprobt und ihre Trennschärfe überwacht.“ Auf Grundlage der Ergebnisse dieser Evaluationen werden die Modelle dann, wenn notwendig, verändert.

Schließlich interessierte, ob die Auskunftsteien bei der Entwicklung und Evaluierung ihres Scoring-Verfahrens externe wissenschaftliche Beratung in Anspruch nehmen. Zwei Auskunftsteien lassen jedes ihrer Scoring-Verfahren von externen Gutachtern überprüfen. Die dritte gibt an, über „eigene kompetente Scoring-Experten“ zu verfügen und deshalb keine externen Gutachten in Anspruch zu nehmen.

Pflege des Datenbestandes

Ein zentraler Aspekt ist die Frage, wie die Auskunftsteien die Richtigkeit und Aktualität ihrer Daten gewährleisten, denn die Aussagekraft des Scores ist auch davon abhängig, ob die verwendeten Daten korrekt sind. Alle Unternehmen führen gleich mehrere Maßnahmen an, zu denen kontinuierliche Datenüberwachung und Einhaltung von Lösfristen gehören.

„Laufende Datenpflege der eigenen Negativ-Daten, Monitoring der Änderungsmeldungen, Löschung der Daten in Abhängigkeit der gesetzlichen Speicherfristen. Beim Bezug von externen Daten werden routinemäßig auf Stichprobenbasis Vollständigkeit und Korrektheit überprüft, und Bestandsveränderungen statistisch überwacht.“ – so die Aussage eines der antwortenden Unternehmen. Eine andere Auskunftstei schreibt: „Wenn Verbraucher auf fehlerhafte Daten hinweisen, werden diese umgehend korrigiert. Zu unseren qualitätssichernden Maßnahmen zählen unter anderem Schulungen bei unseren Kunden, regelmäßige Datenanalysen, automatisierte Plausibilitäts-Checks, sowie regelmäßige Bestandsabgleiche und Dateninventuren, die wir gemeinsam mit unseren Kunden durchführen. Unabhängig davon unterliegen die Vertragspartner einer Aktualisierungspflicht, d. h. sie müssen Änderungen im Datenbestand melden.“

Auskunftsrechte und Korrekturmöglichkeiten für Verbraucher

Alle drei an der Befragung teilnehmenden Auskunftsteien verweisen auf das datenschutzrechtliche Selbstauskunftsrecht. Zwei Auskunftsteien verweisen zusätzlich auf die Möglichkeit, sich über ein persönliches Gespräch vor Ort bzw. über einen Verbraucherservice und die Unternehmenswebseite zu informieren.

Welche Möglichkeit haben Verbraucher, Fehler in den über sie gesammelten Daten korrigieren zu lassen? Die Möglichkeit zur Beschwerde oder Korrekturanfrage (z. B. telefonisch oder schriftlich) bieten alle Auskunftsteien. Eine von ihnen berichtet, dass bei Beschwerden die Daten geprüft und (falls möglich) korrigiert werden, ansonsten erfolge „eine Score-Sperre, welche die Generierung und Übermittlung von Scores zu dem Betroffenen verhindert“. Was genau eine solche Sperre für den Verbraucher bedeutet, wird nicht erläutert. Eine andere Auskunftstei verweist neben ihrem Verbraucherservice auch auf eine kostenlose Schlichtungsstelle, ohne sie konkret zu benennen.

Eine Auskunftstei gibt an, am häufigsten gehe es bei Reklamationen um fehlerhafte Schreibweisen von Namen oder Adressen. Eine andere führt an, die meisten Beschwerden bezögen sich auf Negativ-Informationen zum Zahlungsverhalten. Die dritte Auskunftstei nennt gelegentliche Unstimmigkeiten sowohl bei Adressdaten als auch bei Inkassomerkmalen. Alle drei Auskunftsteien betonen, dass Hinweise auf fehlerhafte Daten äußerst selten vorkämen und meistens unbegründet seien.

Diskriminierung

Gefragt wurden die Auskunftsteien, ob sie es für möglich halten, dass die Ermittlung des Bonitäts-Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten könne. Eine Auskunftstei schreibt: „Nein, mathematisch, statistische Verfahren diskriminieren nicht.“ Eine zweite Auskunftstei verfolgt einen anderen Ansatz: Im Kontext der Bonitätsprüfungen sei es nicht möglich, das Geschäft (die Prüfung) „ohne Ansehen der Person“ durchzuführen. Ausgeschlossen werden aber diskriminierende Parameter wie Nationalität, Beruf, Einkommen oder Familienstand. Auch das Merkmal „Geschlecht“ sei nicht Teil der Scores. Eine andere Auskunftstei schrieb, sie würde geschlechts-spezifische

Variablen verwenden, sofern sie trennscharf sind, also der Prognosegenauigkeit dienen.

Aufsicht

Die Landesdatenschutzbeauftragten werden als zuständige Aufsichtsbehörde für die Prüfung der Scoring-Verfahren genannt. Eine Auskunft gibt an, diesem eine ausführliche Dokumentation des Verfahrens vorgelegt zu haben und ihm die Ergebnisse der regelmäßig stattfindenden Validierungen der verschiedenen Scores mitzuteilen. In Beschwerdefällen würden darüber hinaus einzelne Scoreblätter übermittelt, aus welchen sich das Zustandekommen des Scores ersehen lasse. Eine andere Auskunft schreibt, sie habe ihre Scorekarten sämtlichen Datenschutzbeauftragten des Bundes und der Länder offengelegt und auch neue Scorekarten würden der Aufsichtsbehörde vor Inbetriebnahme gezeigt. Die dritte Auskunft berichtet dagegen ausschließlich von regelmäßigen Treffen mit dem zuständigen Datenschutzbeauftragten und, dass Beschwerden vor Ort überprüft würden.

Erweiterung der Datenbasis

Die Auskunftsteilnehmer wurden gefragt, ob sie eine Erweiterung der Datenbasis um Merkmale aus anderen Bereichen (z. B. Fahrverhalten oder Gesundheitsverhalten) für sinnvoll halten, um die Prognosequalität Ihrer Scoring-Verfahren zu verbessern. Die Auskunftsteilnehmer sehen keine Notwendigkeit, den Datenbestand auf diese Art und Weise zu erweitern. Ein Unternehmen ist der Ansicht: „Da unsere Scores eine Zahlungs(ausfall)-Prognose zum Gegenstand haben, halten wir Daten aus anderen Bereichen diesbezüglich grundsätzlich für nicht relevant“. Ein anderes Unternehmen argumentiert, es habe bereits eine umfangreiche Datenbasis zur Verfügung und biete eine sehr hohe Prognosequalität, weshalb eine Erweiterung um Daten aus anderen Bereichen nicht erforderlich sei.

Anpassung des Gesetzes- und Regulierungsrahmens

Auf die Frage, ob die Auskunftsteilnehmer Bedarf sehen, den Gesetzes- und Regulierungsrahmen in Bezug auf Bonitäts-Scoring anzupassen, geben die Unternehmen in Bezug auf das BDSG n.F. Klärungsbedarf an.

Eine Auskunft schreibt: „Fraglich ist nur, ob der § 31 BDSG-NEU nicht gegen die DSGVO verstößt, was von Datenschutzexperten und -Aufsichtsbehörden teilweise bereits so gesehen wird. Die Gefahr besteht, dass durch eine eventuelle Ausweitung der Informationspflichten bzgl. Scoring die Geschäftsgeheimnisse der Auskunft verletzt werden (s. BGH-Urteil)“. Und eine andere bemerkt: „Mit der DS-GVO haben wir erstmalig einen neuen EU-weit einheitlichen gesetzlichen Rahmen, der nun ab dem 25.5.18 umgesetzt wird. Weitergehende nationale Regelungen sind vor diesem Hintergrund zunächst einmal nicht zielführend und müssten dann auch DSGVO-konform sein.“

Ausblick

Zum Schluss des Fragebogens wurden die Auskunftsteilnehmer gebeten einzuschätzen, wie die Zukunft des Bonitäts-Scorings aussehen könnte und ob es Trends, z. B. aus dem Ausland gebe, die sie für die Weiterentwicklung ihres Geschäftsmodells für relevant hielten.

Alle drei Unternehmen bemerken, dass im Ausland bereits Daten aus sozialen Netzwerken für das Bonitäts-Scoring genutzt werden – und das, wie zwei von ihnen hinzufügen, erfolgreich (wobei keine Quellen angeführt werden). Eine Auskunft sieht darin erhebliches Entwicklungspotenzial. Eine andere sieht sich auf solche Daten nicht angewiesen, fordert aber, dass, falls sich Social Mediadaten bzw. Big Data unter der DSGVO als zulässig für das Bonitäts-Scoring erweisen, der deutsche Gesetzgeber keine davon abweichenden Sonderregelungen für Deutschland schaffen solle.

Zwei Auskunftsteilnehmer betonen mit Blick auf die Zukunft den wichtiger werdenden Bereich der Betrugsprävention. So schreibt eine von ihnen: „Allerdings beruhen nur ca. 50 % der Zahlungsausfälle auf Zahlungsunfähigkeit; die anderen 50 % hingegen auf Zahlungsunwilligkeit, d. h. Betrug (insbes. Identitätsbetrug). Damit kommt im Rahmen des Scoring der Betrugserkennung eine immer gewichtigere Bedeutung zu. Die Betrugserkennung muss mit ganz anderen Methoden und Daten arbeiten als die Bonitätsprüfung, da die Betrüger es (z. B. durch Verwendung fremder oder auch nicht existierender Identitäten) geradezu darauf anlegen, bei der Bonitätsprüfung mit guter bzw. mit zumindest nicht schlechter Bonität zu erscheinen.“

2. Kfz-Versicherungen mit Telematikoption

Geschäftsmodell

Alle befragten Versicherer bieten Telematik-Optionen als einen Baustein an, der zu einem bestehenden Kfz-Versicherungsvertrag dazu gebucht werden kann. Es gibt also keine reinen Telematik-Tarife, bei denen die Tarifgestaltung ausschließlich auf der Auswertung der Fahrweise beruht. Bei allen Angeboten werden verschiedene Fahrdaten erhoben und ein Score berechnet, der die Fahrweise bewertet. Je besser der Score, desto höher die möglichen Vergünstigungen für den Versicherten, welche entweder aus günstigeren Prämien im Folgejahr oder Rückzahlungen von bereits gezahlten Beiträgen bestehen. Bei keinem der Telematik-Tarife ist es möglich, durch schlechtes Fahren bzw. einen schlechten Score einen expliziten Malus zu erhalten. Lediglich ein Geschäftsmodell unterscheidet sich etwas von den anderen: die HDI Versicherung ermöglicht ihren Privatkunden einen Gutschein zur kostenlosen Teilnahme an einem Telematik-basierten Bonusprogramm namens „ryd“ (ehemals „TankTaler“), welches vom Kooperationspartner ThinxNet durchgeführt wird. Für gute Fahrweise können hier Bonuspunkte gesammelt werden, um Gutscheine bei Partnerunternehmen zu erhalten.

Sechs der in der Befragung genannten Angebote stehen nur jungen Fahrern bis zu einem bestimmten Alter (22, 25 oder 28 Jahre) offen. Bei zwei dieser Tarife mit Altersbeschränkungen können auch ältere Fahrer die Telematik-Option nutzen, wenn sie den Führerschein bei Beginn des Programms erst bis zu fünf Jahre lang besitzen, also aus Sicht der Versicherer Fahranfänger sind.

Zielgruppe und besondere Kundenmerkmale

Die meisten Telematik-Angebote richten sich vorrangig an junge Fahrer (6 Angebote) und Fahranfänger (2 Angebote), weil diese aufgrund einer höheren Risikoeinschätzung höhere Prämien zahlen müssen. „Sachlicher Grund für diesen Produktzuschnitt ist, dass junge Erwachsene bis 25 Jahre im Straßenverkehr häufiger und schwerer verunglücken als Fahrer anderer Altersgrup-

pen“, erklärt ein Anbieter. Lediglich drei Befragte geben an, keine bestimmte Zielgruppe zu haben.

Auf die Frage, wie groß der Anteil der Kunden ist, die eine Telematik-Option in Anspruch nehmen, und ob sich diese durch besondere Merkmale auszeichnen, wird deutlich, dass Telematik von nur sehr wenigen Versicherten genutzt wird. Acht Anbieter sprechen von einem geringen Anteil ihrer Versicherten, sechs davon nannten Größenordnungen von weniger als 1% der von ihnen versicherten Fahrer.

Aufgrund der Altersvoraussetzungen ist das Durchschnittsalter der Teilnehmenden im Verhältnis zur Altersstruktur des jeweiligen Versichertenkollektivs geringer. Ein Versicherer gibt an, dass von den Nutzern der Telematik-Option ca. drei Viertel männlich und ein Viertel weiblich sind, es wird jedoch nicht erläutert, wie die Geschlechterverteilung bei den Tarifen ohne Telematik aussieht. Ein Unternehmen mit zwei Telematik-Angeboten behauptet, die Telematik-Teilnehmer zeichnen sich durch ein besseres Fahrverhalten aus, ohne diese Erkenntnis an dieser Stelle näher zu erläutern. Drei Anbieter können keine besonderen Merkmale bei ihren Telematik-Kunden feststellen.

Auf die Frage, mit welchen in Aussicht gestellten Effekten die Kunden zur Teilnahme am Telematik-Tarif motiviert werden, standen die mögliche Preisersparnis bzw. möglichen Boni klar an erster Stelle (8 Anbieter). Außerdem wird den Versicherten in Aussicht gestellt, dass sie Informationen über ihre Fahrweise erhalten und so darin unterstützt werden, diese zu verbessern, bzw. dass die Sicherheit erhöht werde (7 Anbieter). Vier Anbieter sehen als Anreizfaktoren einen verbesserten Service und andere Zusatzleistungen, die mit der Telematik-Option verbunden sind, wie z. B. Notruffunktionen, Tankstellenfinder oder Diebstahlwarnungen. Ein Versicherer hebt die Motivation hervor, sich über Bestenlisten in der App mit anderen Nutzern in Bezug auf das Fahrverhalten vergleichen zu können.

Vorteile und Prämien durch hohe Scores

Dem Versicherten kann die Bewertung des Fahrverhaltens auf verschiedenste Weise mitgeteilt werden. Alle an der Befragung teilnehmenden Anbieter geben zunächst einen Punktestand aus. Fünf von ihnen teilen

dem Kunden das Ergebnis ihres Scorings zusätzlich mit Symbolen wie z. B. Farbskalen, Ampeln oder Medaillen mit. Zwei Anbieter sprechen dem Kunden ausformulierte Empfehlungen zum Fahrverhalten aus.

Bei guter Fahrweise bzw. ab einem bestimmten Score gewähren die Unternehmen ihren Kunden verschiedene Vorteile. Üblicherweise geschieht dies in Form einer günstigeren Versicherungsprämie im Folgejahr (6 Anbieter), aber auch eine Rückzahlung von bereits geleisteten Versicherungsbeiträgen ist möglich (1 Anbieter). Ein Anbieter gewährt seinen Kunden bei entsprechender Fahrweise Gutscheine oder Rabatte bei Partnerunternehmen.

Als einziges Entscheidungskriterium für die Gewährung der genannten Vorteile gilt bei allen Anbietern der Score des Fahrers. Dieser wird üblicherweise nach jeder Fahrt aktualisiert (5 Anbieter) und ist für den Kunden jederzeit in der App oder über eine Webseite einsehbar (6 Anbieter). Nur ein Versicherer gibt an, den Score des Kunden lediglich einmal im Monat mitzuteilen. Das Einsehen des Scores über eine App auf dem Mobiltelefon ist hierbei Standard. Die Möglichkeit, den Score darüber hinaus auch online zu veröffentlichen, bieten drei Anbieter an. Bei ihnen ist es Kunden ebenfalls möglich, ihren Score im Ranking mit anderen Fahrern zu vergleichen.

Datengrundlage

Anbieter setzen die Telematik-Option mithilfe unterschiedlicher technischer Möglichkeiten um, um Daten über die Fahrweise der Versicherten zu erheben (siehe Abbildung III.1).

Eine Möglichkeit stellt die sogenannte Telematik-Box dar, die im Motorraum des Autos eingebaut wird (3 Anbieter). Die Box ist an die Stromversorgung des Fahrzeugs angeschlossen und zeichnet alle Daten mit eigenen Sensoren auf.

Eine weitere Möglichkeit besteht darin, dem Kunden einen Telematik-Stecker auszugeben, den dieser selbst in den Zigarettenanzünder seines Fahrzeugs steckt (2 Anbieter). Dieser hat üblicherweise einen Beschleunigungssensor und ermittelt mittels GPS-Sensor Positionsdaten und Geschwindigkeit des Fahrzeugs.

Schließlich gibt es noch die Optionen eines sogenannten OBD-Telematik-Steckers (On-Board-Diagnose), der an die Fahrzeugdiagnose-Schnittstelle, die mittlerweile in allen Fahrzeugen vorhanden und standardisiert ist, angeschlossen wird (1 Anbieter). Über diese Schnittstelle können alle relevanten Fahrzeugdaten ausgelesen werden.

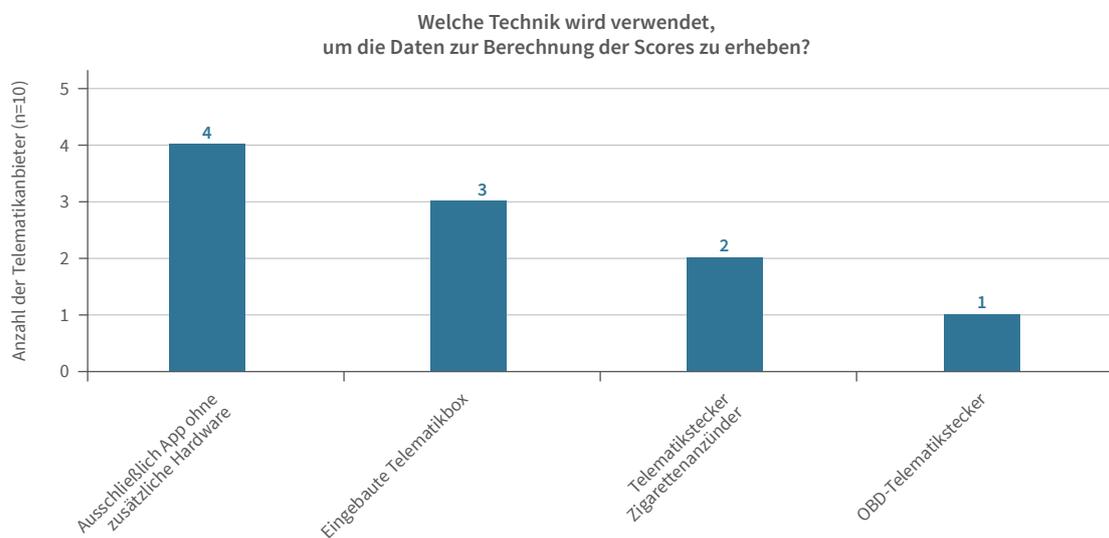


Abbildung III.1

Vier Versicherer geben an, Daten ausschließlich über eine auf dem Smartphone des Fahrers installierte App zu sammeln und somit auf zusätzliche Hardware zu verzichten (nur in einem Fall bedarf es zusätzlich einer Verbindung mit einer Bluetoothquelle im Fahrzeug). Über die üblicherweise in Smartphones eingebauten Sensoren (z. B. Gyro-Sensoren) in Kombination mit GPS-Signalen werden die Daten zum Fahrverhalten gemessen.

Zur Berechnungsgrundlage für den Score gehören üblicherweise drei bis neun Merkmale, wie z. B. Fahrweise (zusammengesetzt aus z. B. Geschwindigkeit, Bremsverhalten, Lenk- bzw. Kurvenverhalten, Beschleunigungsverhalten), Tag/Uhrzeit, Ort, Länge bzw. Dauer einer Fahrt, Straßentyp, Bevölkerungsdichte (siehe Abbildung III.2). Neben den Daten, die durch die Telematik-Technik bzw. die Apps selbst erfasst werden, wird als weitere Datenquelle häufig digitales Verkehrs- und Straßenkartenmaterial verwendet.

Die von den Anbietern genannten Merkmale fließen immer in die Berechnung des Scores ein, es gibt keine Fälle, in denen einzelne Faktoren nur fallweise berücksichtigt werden.

Methodik der Scoring-Verfahren

Alle Versicherungsanbieter, die an unserer Befragung teilgenommen haben, lassen die Daten von externen Dienstleistern auswerten und von diesen auch den Score berechnen. Von den externen Dienstleistern haben fünf ihren Hauptsitz in Deutschland, zwei in England und einer in Kanada.

Die Auswahl der zu verwendenden Kriterien und die Festlegung ihrer Gewichtung erfolgt auf unterschiedliche Art und Weise. Häufig liegen der Modellbildung Auswertungen von Statistiken zu Fahrweisen und Unfallursachen, z. B. vom Statistischen Bundesamt oder der Royal Society for the Prevention of Accidents, zugrunde (7 Nennungen sowohl bei Merkmalsauswahl als auch bei Merkmalsgewichtung). Ein Anbieter erläutert: „Bei der Auswahl der Score-Parameter im Hinblick auf die Schadenseintrittswahrscheinlichkeit orientieren wir uns auf die Ergebnisse der Testphasen mit internen Usern in Anlehnung an den vom Statistischen Bundesamt veröffentlichten Statistiken über Unfallursachen, über die Ortslage von Verkehrsunfällen und über die Tageszeit von Unfällen. Erheblich ist zudem die Statistik der Bundesanstalt zur Verkehrsentwicklung auf Bundesfernstraßen.“

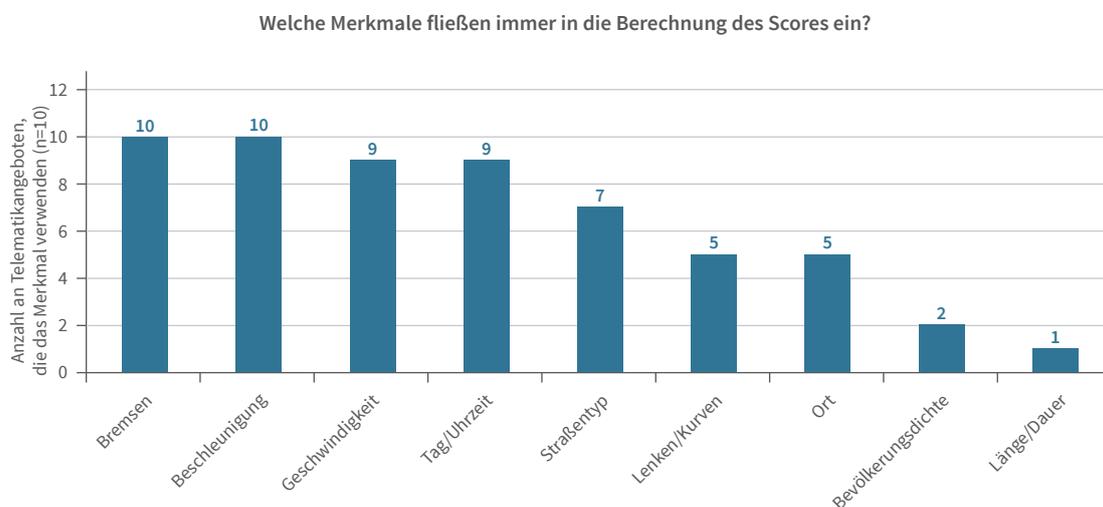


Abbildung III.2

Die Untersuchung der Signifikanz bzw. Aussagekraft einzelner Merkmale mit statistischen Methoden und Tests wird ebenfalls durchgeführt – sowohl mit Blick auf die Merkmalsauswahl (4 Nennungen) als auch deren Gewichtung (5 Nennungen). Eine eher heuristische Herangehensweise („Faustregel“) liegt vor, wenn Erfahrungen aus anderen Ländern oder Unternehmen oder eigene Erfahrungswerte zu Rate gezogen werden, um Kriterien auszuwählen und Gewichtungen festzulegen (4 Nennungen bei Merkmalsauswahl, 1 Nennung bei deren Gewichtung). „Es gibt Merkmale, die signifikant sind, die sich aber dem menschlichen Verständnis erstmal intuitiv nicht so einfach erschließen. Bei solchen Merkmalen wird man entscheiden müssen, ob man diese in ein Modell aufnimmt oder nicht“, begründet ein Anbieter das Vorgehen.

Weiterhin wurde gefragt, welche statistischen Schätzmethoden dem Scoring zugrunde liegen. Die logistische Regression kommt bei vier Angeboten zum Einsatz. Ein Konzern kreuzte für seine beiden Telematik-Tarife weiterhin sämtliche Verfahren an, die zur Auswahl gestellt wurden: Clusterbildende Methoden, Entscheidungsbäume, Ensemblemethoden und Neuronale Netzwerke würden allesamt verwendet, um Scoring Modelle zu entwickeln. Die restlichen Anbieter beantworteten die Frage überhaupt nicht. Auch, was die Angabe eines Zielmerkmals (abhängige Variable) im verwendeten Modell angeht, blieben die Anbieter wenig auskunftsbereit. Zweimal wird als Zielmerkmal die Schadenseintrittswahrscheinlichkeit genannt. Bei zwei anderen Angeboten ist es die Prognose des „Schadengeschehens“ mittels der Zielvariablen „Schadenhäufigkeit und Schadenbedarfe“.

Auf unsere Fragen zur Güte der verwendeten Modelle (Wie wird diese gemessen, wie hoch ist sie?) gaben uns die Unternehmen nur wenige Antworten. Gütemaße werden überhaupt nicht genannt und auch wie die Güte der Scoring-Modelle aufrechterhalten und verbessert wird, konnte uns nicht im Detail beschrieben werden. „Die Ergebnisse des operativ laufenden Algorithmus werden laufend überwacht und mit den Schadendaten abgeglichen. Auf Basis neuer statistischer Erkenntnisse können Anpassungen am Modell vorgenommen werden“, heißt es etwa oder es wird auf frühere Aussagen zur Merkmalsauswahl und Gewichtung verwiesen.

Bei der Entwicklung und Evaluierung der Scoring-Verfahren gab lediglich ein Versicherer (mit zwei Angeboten) an, externe wissenschaftliche Beratung in Anspruch zu nehmen. Das Unternehmen verweist hier jedoch lediglich auf eine Untersuchung zu dem Thema im Rahmen einer Masterarbeit. Der Rest verneinte die Frage nach externer wissenschaftlicher Beratung oder beantwortete die Frage überhaupt nicht. Ein Unternehmen (mit zwei Angeboten) gibt darüber hinaus an, intern genügend eigene, wissenschaftlich ausgebildete Experten zu beschäftigen, die Forschung im Bereich der Versicherungstelematik betreiben.

Pflege des Datenbestandes

Verschiedene Maßnahmen werden ergriffen, um die Aktualität und Richtigkeit der verwendeten Daten zu gewährleisten. Es gibt Qualitätschecks und regelmäßige Aktualisierungen (z. B. des verwendeten Kartenmaterials und der Verkehrsdaten) (6 Anbieter). Die Fahrdaten werden auf Plausibilität geprüft (2 Anbieter) und die ordentliche Funktionalität der eingesetzten Technik z. B. mit Betriebsleuchten angezeigt (2 Anbieter). Bei einem Unternehmen (mit zwei Angeboten) wird außerdem explizit gesagt, dass Messfehler nicht auszuschließen seien: „Wichtig ist daher zum einen, das Möglichste zu tun, um solche Messfehler zu erkennen und auszuschließen. Zum anderen aber auch dafür zu sorgen, dass es „Toleranzen“ gibt, so dass nicht der eine fehlerhafte Datenpunkt den Score eines Fahrers gleich dauerhaft zerstört. Und zu guter Letzt klar und deutlich zu sagen, dass jeder Fahrer ein gewisses Maß an fehlerhaften Daten in seinen Aufzeichnungen haben wird. Dass somit aber auch jeder Fahrer davon betroffen ist und dies in der Kalibrierung der Algorithmen also schon mit enthalten ist.“

Auskunftsrechte und Korrekturmöglichkeiten für Verbraucher

Auf die Frage, wie Kunden von Telematik-Tarifen erfahren, welche Daten über sie in die Berechnung des Scores einfließen, führen alle Versicherungsunternehmen, mit einer Ausnahme, einzelne Komponenten der Vertragsunterlagen an (Allgemeine Bedingungen für die Kfz-Versicherung, Nutzungsvereinbarungen, Datenschutzerklärung etc.). Bei sieben Angeboten finden Kunden zusätzlich Informationen und Veröffentlichungen auf der Homepage oder in der Telematik-App bzw. erhal-

ten Handbücher oder Informationsblätter. Ein Anbieter schreibt außerdem, die Kunden würden vor Abschluss des Vertrages durch die Beratung des Versicherungsmittlers über die Datenverwendung aufgeklärt.

Die häufigsten Fehler, auf die von den Kunden hingewiesen wird, gibt es laut Anbieter aufgrund von veraltetem Kartenmaterial (7 Nennungen). Fälschlicherweise werden z.B. Überschreitungen der erlaubten Höchstgeschwindigkeit berechnet, wenn temporäre Tempolimits (z.B. wegen Baustellen) im hinterlegten Kartenmaterial falsch verzeichnet oder die herangezogenen Verkehrsdaten nicht aktuell sind. Zwei Anbieter erwähnen, dass Kunden mitunter reklamieren, dass die automatische Fahrtaufzeichnung nicht begonnen oder unterbrochen wurde, weil es Empfangsprobleme (z.B. wegen schlechtem GPS-Signal) gab.

Bis auf einen Anbieter ermöglichen alle ihren Kunden, Fehler zu reklamieren und Korrekturen zu veranlassen. Allerdings betonen einige Unternehmen, dies geschehe nur selten oder in Ausnahmefällen aus Kulanz. Sechs von ihnen verweisen bei der Beantwortung der Frage explizit auf ihren telefonischen Supportservice oder Kontaktmöglichkeiten per Email. Nur ein Anbieter schließt Fehlerkorrekturen komplett aus – nicht das Versicherungsunternehmen, sondern nur der externe Dienstleister habe Zugriff auf die Daten.

Diskriminierung

Die bei der Befragung antwortenden Unternehmen sind sich einig: die Möglichkeit, dass die Ermittlung des Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten könne, schließen sie aus. „Persönlichkeitsmerkmale“ würden nicht verwendet (3 Nennungen) und gesetzliche Vorgaben wie durch das AGG bei der Datenverwendung eingehalten (2 Nennungen).

Aufsicht

Die Frage, ob die verwendeten Scoring-Verfahren der Telematik-Tarife staatlicher Aufsicht unterliegen, und wenn ja, welche Aufsichtsbehörden für eine Prüfung zuständig seien, wurde unterschiedlich beantwortet. Fünf Anbieter geben an, es gäbe keine staatliche Aufsicht, vier Anbieter benennen die BaFin als zuständig. Bei zwei Angeboten wird eine Landesdatenschutzbehörde

zusätzlich zur BaFin aufgeführt, diese sei vor Produkteinführung konsultiert worden.

Erweiterung der Datenbasis

Auf die Frage, ob sich die Versicherer auch vorstellen können Merkmale aus anderen Bereichen heranzuziehen (zum Beispiel Bonitätsinformationen oder Gesundheitsdaten), um die Prognosequalität Ihrer Scoring-Modelle zu verbessern, lautet die Antwort einstimmig „Nein“; nur solche Daten mit Bezug zum Fahrverhalten seien sinnvoll.

Zusammenhang zwischen Telematik und Fahrweise

Auf die Frage, ob die Unternehmen Erkenntnisse darüber haben, dass sich die Schadenhäufigkeit ihrer Kunden durch die Teilnahme an einem Telematik-Tarif verringert, kann keines der Unternehmen valide Aussagen machen. Ein Anbieter gibt lediglich an, es zeigten sich erste Anzeichen, dass Kunden, die an Telematik-Optionen teilnehmen, seltener Schäden verursachen, als Kunden die nicht teilnehmen: „Erste Analysen bestätigen die Hypothese, dass Kunden, die den Telematik-Baustein gewählt haben, im Vergleich zum durchschnittlichen Kunden ohne Telematik-Baustein vorausschauender fahren“.

Gleichzeitig wird von einem Unternehmen (mit drei Angeboten) eingeräumt, dass Vorher-Nachher-Vergleiche, die einen konkreten Einfluss von Telematik auf die Fahrweise zeigen könnten, nicht durchgeführt werden bzw. nicht durchgeführt werden können. Es sei nicht auszuschließen, dass ohnehin vermehrt defensive Fahrer mit geringerer Schadenhäufigkeit Telematik-Verträge in Erwägung ziehen und risikoreiche, schadensanfällige Kunden Telematik-Optionen nicht nutzen: „Kunden, die wissen, dass sie umsichtig und verantwortungsvoll Auto fahren, sind auch eher bereit, an einem Telematik-Tarif teilzunehmen.“

Anpassung des Gesetzes- und Regulierungsrahmens

Auf die Frage an die Unternehmen, ob sie im Zusammenhang mit Telematik-Tarifen Bedarf sehen, den Gesetzes- und Regulierungsrahmen anzupassen, gehen die Meinungen auseinander. Die Hälfte der Antwortenden

sieht keinen Anpassungsbedarf. Die andere Hälfte führt verschiedene Wünsche an: Datensouveränität – es müsse deutlich sein, dass die Daten Eigentum des Kunden sind und dem Versicherer nur deren Nutzung gestattet wird (4 Nennungen), und Vermeidung von Überregulierung (3 Nennungen).

Weitere Forderungen umfassen das Verhindern von Datenmonopolen, eine eindeutige Regelung von Informationspflichten, eine eindeutige Regelung des Datenschutzes, die Ermöglichung von Forschung sowie vermehrte Diskurse zu Digitalisierung und Verbraucherschutz.

Akzeptanz von Malussystemen

Alle Telematik-Tarife auf dem deutschen Markt ermöglichen lediglich eine Belohnung guten Fahrens. Eine Bestrafung fahrlässigen und risikoreichen Fahrens wird bislang ausgeschlossen. Deswegen wurden die Unternehmen gefragt, wie sie die Akzeptanz ihrer Kunden einschätzen würden, für schlechtes Fahrverhalten höhere Prämien zahlen zu müssen.

Die Hälfte der Anbieter schätzt die Akzeptanz eines Malussystems als sehr niedrig bzw. überhaupt nicht vorhanden ein. Zweimal wurde die Aussage getätigt, es fehlten dazu Erkenntnisse und eine Einschätzung sei deshalb nicht möglich. Ein Unternehmen (mit zwei Telematik-Angeboten) betont mit Verweis auf die Schadenfreiheitsklasse aber, dass höhere Prämien bei schlechtem Fahren in traditionellen Kalkulationsmodellen bereits üblich seien.

Vor- und Nachteile von Telematik-Tarifen

Als Vorteil von Telematik-Tarifen sehen fast alle Anbieter den Rückgang von Schadensfällen und die Erhöhung von Sicherheit im Straßenverkehr (9 Nennungen). In diesem Zusammenhang wird auch der Aspekt der Kostenersparnis genannt (5 Nennungen). Durch weniger Unfälle würden die Kosten sowohl für Unternehmen als auch für die Gesellschaft sinken. Auch eine durch die Auswertung von Fahrdaten ermöglichte risikogerechte Prämienkalkulation wird von den meisten Unternehmen als Nutzen genannt (9 Nennungen).

Die Hälfte der Antwortenden führt außerdem einen Vorteil für Geschäftsmodelle und Wettbewerb durch Tele-

matik-Angebote an. Einige Versicherer erkennen in Telematik das Potenzial für weitere Serviceleistungen, wie Notfallhilfe und einfachere Kommunikation zwischen Kunden und Unternehmen (4 Nennungen). Schließlich spielt bei zwei Unternehmen für die positive Bewertung von Telematik auch eine Rolle, dass Kunden in Zeiten der Digitalisierung entsprechende Weiterentwicklungen in der Kfz-Versicherung erwarten würden.

Nachteile in dem Einsatz von Telematik in der Kfz-Versicherung sehen vier Anbieter überhaupt nicht. Die anderen benennen hohe Kosten, die durch die eingesetzte Technik entstehen würden (4 Anbieter) sowie, dass mit der Einführung solcher Neuerungen gewisse Risiken für das Unternehmen einhergehen (2 Anbieter). Konkreter benannt wurden die Risiken allerdings nicht.

Zwei Anbieter sehen die Gefahr, dass Fahrzeughersteller ein Datenmonopol schaffen könnten, wenn nur Sie über den Zugriff auf die durch die eingebaute Telematik-Technik gesammelten Fahrdaten bestimmen könnten.

Ein Unternehmen stellt den Nutzen von Telematik für die Risikokalkulation in Frage: „Das Produkt Kfz-Versicherung ist mit den heute schon vorhandenen Tarifierungsmerkmalen sehr risikogerecht kalkuliert. Es bleibt daher abzuwarten, ob Telematik-Tarife tatsächlich bereits vorhandene Tarifierungsmerkmale ablösen oder zumindest ergänzen können.“

eCall

Mit einer spezifischen Frage wurde erhoben, ob und welche Bedeutung das von der Europäischen Union vorgeschriebene automatische Notrufsystem für Kraftfahrzeuge namens eCall (Kurzform für emergency call) für Telematik-Tarife hat bzw. haben könnte. Hersteller müssen eCall seit dem 31. März 2018 in alle neuen Modelle von Pkw und leichten Nutzfahrzeugen einbauen. Im Fahrzeug montierte Geräte sollen einen Verkehrsunfall automatisch an die einheitliche europäische Notrufnummer 112 melden und durch die rascher initiierten Rettungsmaßnahmen die Zahl der Verkehrstoten senken.

Die Einführung von eCall bedingt unter anderem die Ausstattung von Fahrzeugen mit einem Airbag-Sensor, der die Information bereit stellt, ob ein Unfall mit Verletzungsgefahr passiert ist, einem Satelliten-Empfänger,

damit das Auto via GPS und/oder Galileo seinen Standort zum Zeitpunkt des Unfalles ermitteln kann (wird vielfach zusammen mit dem Navigationssystem genutzt), eine Mobilfunkantenne zum Senden des Notrufes an die Notrufzentrale und eine eCall-Steuerung (teilweise im Infotainment integriert), die die für den eCall benötigten Daten sammelt, nach einem Unfall übermittelt und eine Sprachverbindung zur Notrufzentrale öffnet.

Fünf Anbieter geben an, die eCall-Technik fördere in der Tat Dienstleistungen wie Telematik.

Ebenso viele betonen jedoch auch den Aspekt der Datensouveränität, der Kunde müsse Herr seiner Daten sein. Augenscheinlich besteht bei den Kfz-Versicherern die Befürchtung, es könnten Datenmonopole (z. B. durch Original Equipment Manufacturer, sogenannte OEMs) entstehen und die Versicherer keinen Zugang zu den Daten aus eCall Systemen erlangen.

Sechsmal wird darauf hingewiesen, dass nur herstellerunabhängige Plattformen einen fairen Wettbewerb ermöglichen würden. Bereits bei den Fragen zu den Nachteilen von Telematik und zum Regulierungs- und Gesetzesrahmen war auf diesen Aspekt hingewiesen worden (s.o.).

Ausblick

Den Unternehmen wurde die Möglichkeit gegeben, eine Einschätzung darüber abzugeben, wie die Zukunft von Scoring-Verfahren in der Kfz-Versicherung aussehen könnte und ob es ihrer Meinung nach interessante Trends, z. B. im Ausland gibt, die in diesem Kontext relevant sein könnten.

Die Meinungen gehen hier durchaus erkennbar auseinander. Einige Anbieter prognostizieren eher eine nur langsame Verbreitung (3 Nennungen). So argumentiert ein Unternehmen (mit drei Angeboten): „Anders als in den Ländern, in denen Telematik-Tarife bereits sehr verbreitet sind (z. B. in Italien), führt die sehr differenzierte Tarifstruktur in Deutschland allerdings dazu, dass sich Telematik-Tarife in der KFZ-Versicherung nicht mit derselben zeitlichen Dynamik durchsetzen werden wie in unseren europäischen Nachbarländern“. Ein anderer Anbieter ist u. a. der Meinung, der zusätzliche Erkenntnisgewinn durch Telematik sei als sehr gering einzuschätzen und die zu erwartenden Kosten zu hoch.

Doch es gibt auch Stimmen, die eine andere Einschätzung ausdrücken. Ein Unternehmen (mit zwei Angeboten) sieht zukünftig eine eher schnelle Verbreitung von Telematik-Tarifen: „Mit Blick auf die eCall Systeme, sowie aufgrund der vielfach gebotenen Mehrwerte der Telematik-Optionen zum Thema Service und Sicherheit, ist der Trend aus unserer Sicht nicht aufzuhalten.“

Weiterhin findet sich die Meinung, Telematik würde es ermöglichen, die bisherigen Geschäftsmodelle und Versicherungsprodukte grundlegend zu verändern (4 Nennungen) und der Kfz-Versicherungsbranche wäre daran gelegen, neue Zielgruppen zu erschließen (1 Nennung). So schreibt ein Versicherer (mit drei Angeboten): „Wir glauben, dass der Kunde heute von seinem Versicherer mehr erwartet, als die Regulierung im Schadenfall.“

3. Krankenversicherungen

Von den 62 Krankenversicherungen, die an der Befragung teilgenommen haben, handelt es sich um 47 gesetzliche Krankenkassen und um 15 private Krankenversicherer. Unter den gesetzlichen Krankenkassen befinden sich landes- und bundesweite Versicherer, Ersatzkrankenkassen, Betriebskrankenkassen und Innungskrankenkassen.

Bis auf zwei Ausnahmen bieten alle gesetzlichen Krankenkassen, die an der Befragung teilgenommen haben, ein Bonusprogramm für ihre Versicherten an. Im Rahmen von Bonusprogrammen werden Maßnahmen der Prävention und ein gesunder Lebensstil mit Bar- oder Sachprämien belohnt. Rechtliche Grundlage für die Bonusprogramme bildet § 65a SGB V.

Vergleichbare Präventions- oder Vorteilsprogramme werden von den an der Befragung teilnehmenden privaten Krankenversicherern nicht angeboten. Üblich sind hier – in den entsprechenden Tarifen – wohl eher Boni in Form einer Beitragsrückerstattung, wenn in einem festgelegten Zeitraum keine Leistungen in Anspruch genommen wurden. Lediglich ein privater Krankenversicherer berichtet von einem Beitragsrabatt, der den Kunden gewährt wird, wenn bestimmte Gesundheitswerte im Normbereich liegen und eine Bestätigung des Nicht-Raucherstatus vorliegt.

Die hier beschriebenen Erkenntnisse zum Thema Bonusprogramme beziehen sich daher auf die Angebote der an der Befragung teilnehmenden 45 gesetzlichen Krankenkassen mit Bonusprogrammen. Bei den anschließend folgenden, allgemeineren Fragen zum Thema verhaltensbasierte Tarife in der Krankenversicherung werden die Antworten aller 62 Antwortenden ausgewertet.

Zielgruppe und besondere Merkmale der Teilnehmer

Bei den meisten Bonusprogrammen soll keine bestimmte Zielgruppe angesprochen werden (33 Versicherungen). Einige wenige Bonusprogramme richten sich gezielt an Familien (3 Versicherungen) oder Versicherte ab einem bestimmten Mindestalter wie z. B. 15, 16 oder 18 Jahre, schließen also Kinder aus (3 Versicherungen).

Fünf Versicherer geben an, gezielt gesundheitsbewusste Versicherte ansprechen zu wollen. Zweimal wurden kostenbewusste Kunden als Zielgruppe genannt.

Die Höhe des Anteils der Versicherten, die jeweils am Bonusprogramm teilnehmen, variiert:

- 0–5 % der Gesamtzahl der Versicherten:
10 Bonusprogramme
- 6–10 % der Gesamtzahl der Versicherten:
17 Bonusprogramme
- 11–20 % der Gesamtzahl der Versicherten:
13 Bonusprogramme
- 21–30 % der Gesamtzahl der Versicherten:
1 Bonusprogramm
- 31–40 % der Gesamtzahl der Versicherten:
1 Bonusprogramm

Einige Versicherer antworten, die am Bonusprogramm teilnehmenden Versicherten seien eher weiblich (13 Nennungen). Bei den Altersstrukturen lässt sich im Vergleich mit Nicht-Teilnehmern insgesamt keine Struktur erkennen: bei einigen sei die Altersstruktur eher jünger, bei einigen eher älter – es fehlt jedoch bei diesen Angaben der Vergleich mit den Merkmalen der Nicht-Teilnehmer bzw. der Gesamtzahl der Versicherten.

In Bezug auf die Teilnehmer an Bonusprogrammen war außerdem von Interesse, mit welchen in Aussicht gestellten Effekten sie von den Krankenkassen zur Teilnahme motiviert werden. Die Vergütung steht hier an erster Stelle, 33 Krankenkassen nennen Sie als Motivator. Die unmittelbaren und präventiven Gesundheitseffekte stellen dagegen 28 Versicherer ihren Versicherten in Aussicht, um sie zur Teilnahme zu bewegen. Häufig wurden die beschriebenen Antworten auch in Kombination genannt (17 Mal).

Vorteile und Prämien durch Bonuspunkte

Hat der Versicherte eine vom Versicherer vorgegebene Anzahl an Bonuspunkten gesammelt oder eine vorgegebene Anzahl an Aktivitäten absolviert, kann er verschiedene Vorteile erlangen.

Bei 41 der 45 Bonusprogramme in der Befragung erfolgt die Ausschüttung der Boni in Form von Geldleistungen. 14 Versicherer geben an, den Versicherten Sachleistungen zu gewähren. Bei sechs Bonusprogrammen können die Teilnehmer Gutscheine oder Rabatte bei Partnerunternehmen erhalten. Möglich sind weiterhin Rückzahlungen bereits gezahlter Beiträge (5 Bonusprogramme) sowie Gutscheine oder Rabatte für medizinische Behandlungen (4 Bonusprogramme). Es wurden weiterhin die Erstattung von oder ein Budget für weitere Gesundheits-/ Vorsorgeleistungen wie z. B. eine professionelle Zahnreinigung genannt. Die Krankenkassen können so gegenüber ihren Versicherten ein breiteres Leistungsspektrum über die Regelleistungen hinaus anbieten. Teilweise bieten Krankenkassen ihren Versicherten eine Auswahl zwischen mehreren Arten der Bonifizierung an.

Aufbau und Ablauf der Bonusprogramme

Mit zahlreichen unterschiedlichen Aktivitäten und Maßnahmen können die Versicherten Bonuspunkte sammeln (siehe Abbildung III.3). Bei der Befragung wurde eine Reihe von Antwortmöglichkeiten vorgegeben, es folgt eine Auflistung darüber, welche Aktivitäten bzw. Maßnahmen wie häufig von den Anbietern von Bonusprogrammen als bonifizierungsfähig genannt wurden:

- Früherkennungs- bzw. Vorsorgeuntersuchungen: 45 Nennungen
- Teilnahme an Gesundheitskursen: 41 Nennungen
Z. B.: Kurse zu den Themen Ernährung, Bewegung, Entspannung, Sucht, allgemeine Präventionskurse, Eltern-Kind-Kurse
- Mitgliedschaft in einem Sportverein, Fitnessstudio oder im Betriebs-/ Hochschulsport: 41 Nennungen
- Teilnahme an Sportveranstaltungen: 29 Nennungen
Z. B.: Laufveranstaltungen/ Volksläufe, Wanderungen oder Wanderabzeichen, Radtouren, Bundesjugendspiele, Sportkurse
- Sportabzeichen: 41 Nennungen
- Schutzimpfungen: 41 Nennungen
- Erfassung von Daten mittels Fitness-Apps und/ oder Wearables: 3 Nennungen

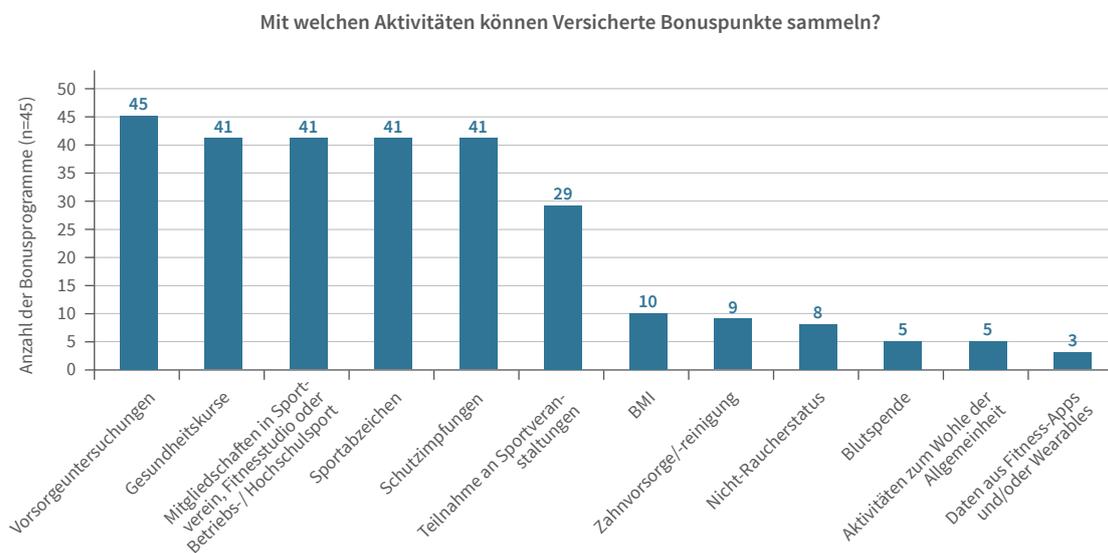


Abbildung III.3

- Blutspende: 5 Nennungen
- Soziales Engagement (Maßnahmen zum Wohle der Allgemeinheit): 5 Nennungen⁹⁹
Z. B.: Erste-Hilfe-Kurs, Knochenmarktypisierung, Organspendeausweis, Fahrsicherheitstraining, Patientenverfügung
- Sonstiges: normaler BMI (10 Nennungen), Zahnvorsorge/-reinigung (9 Nennungen), Nicht-Raucherstatus (8 Nennungen), Online-Coaches, sportmedizinische Untersuchungen, Mutterpass, Eltern-Kind-Kurse, Erwerb eines Wearables

Falls Bonusprogramm-Teilnehmer aufgrund gemeinnützigen Verhaltens Bonuspunkte sammeln können, wurden die Versicherer gefragt, wie die Berücksichtigung solcher Aktivitäten, die keinen unmittelbaren Gesundheits- bzw. Krankheitsbezug erkennen lassen, begründet wird. Insgesamt fünf der auf die Befragung antwortenden 45 Versicherer mit Bonusprogrammen geben an, Blutspende und/oder andere gemeinnützige Aktivitäten wie Erste-Hilfe-Kurse oder Fahrsicherheitstraining in ihr Bonusprogramm aufgenommen zu haben.

Als Begründung wurde von drei Versicherern angegeben, es handle sich hierbei um eine Belohnung altruistischen Verhaltens bzw. um Prävention für Andere. Die Bonifizierung erfolge, „(...) um Teilnehmer am Bonusprogramm zu motivieren, einen Beitrag zur Verwirklichung sozialer Gerechtigkeit und sozialer Sicherheit zu leisten (...)“, schreibt eine Krankenkasse und eine andere bemerkt: „Denn, wer nicht nur an sich selbst denkt, sondern sich auch aktiv für die Gesundheit anderer Menschen einsetzt, hat Bonuspunkte verdient.“

Ein Versicherer argumentiert, dass die angeführten Aktivitäten auch einen individuellen gesundheitlichen Nutzen für den Versicherten haben: „Bei einer Blutspende wird das eigene Blut ebenfalls gründlich untersucht, somit dient die Spende auch der eigenen Vorsorge.“

Einige Befragte kommentierten diese Frage, obwohl es in ihren Bonusprogrammen nicht möglich ist, mit

gemeinnützigen Aktivitäten Punkte zu sammeln. Neun Krankenkassen äußerten sich darüber, dass ihnen diese Möglichkeit durch das Bundesversicherungsamt (BVA) explizit verwehrt würde, auch wenn sie selbst es gerne einführen würden.

Weiterhin wurde erfragt, anhand welcher Kriterien die Versicherer entscheiden, mit welchen Aktivitäten Kunden Bonuspunkten sammeln können. 14 der Befragten geben an, eine Aktivität müsse, um in das Bonusprogramm aufgenommen zu werden, einen nachgewiesenen gesundheitlichen Nutzen haben bzw. gesundheitsbewusstes Verhalten fördern. Etwa gleich häufig (15 Nennungen) wurden die gesetzlichen Vorgaben gemäß §65a SGB V angeführt. Auch die Vorgaben des Bundesversicherungsamts (9 Nennungen) und des GKV-Spitzenverbandes (4 Nennungen) wurden als Grundlage für die Auswahl der Aktivitäten und Maßnahmen genannt. Eine weitere Rolle spielen Kriterien wie die Förderung von Angeboten zur Früherkennung und Prävention (6 Nennungen), Vorgaben in der Satzung (4 Nennungen), Markttrends (3 Nennungen), Nachhaltigkeit (3 Nennungen), Attraktivität für Versicherte (3 Nennungen), nicht näher benannte Qualitätskriterien (3 Nennungen), Nachweisbarkeit (3 Nennungen), wirtschaftlicher Nutzen (2 Nennungen), Eigenverantwortung (2 Nennungen), Kundenbefragungen (1 Nennungen), und Bezug zum Trägerunternehmen (1 Nennungen).

Auch die Grundlage für die Gewichtung der einzelnen Maßnahmen wurde erfragt. Eine Reihe von Bonusprogrammen gewichtet die möglichen Aktivitäten im Bonusprogramm überhaupt nicht (das heißt jede Aktivität oder Maßnahme zählt gleich viel) oder es gibt kein Punktesystem, es muss nur eine bestimmte Mindestanzahl von Aktivitäten absolviert werden (17 Bonusprogramme).

Am häufigsten erfolgt die Gewichtung der Aktivitäten auf Grundlage von Aufwand und Unkompliziertheit (8 Nennungen). Die Bepunktung richtet sich also danach, wie leicht oder schwer es für die Versicherten ist, eine Maßnahme durchzuführen. Bei anderen Kassen bestimmt der gesundheitliche Nutzen die Gewichtung der Aktivitäten (5 Nennungen). Ebenfalls genannt wurden

⁹⁹ Gezählt wurden hier auch Bonusprogramme, für die nicht explizit die Antwortmöglichkeit „Soziales Engagement“ angekreuzt wurde, bei denen aber unter „Sonstiges“ gemeinnützige Maßnahmen wie Erste-Hilfe-Kurse, Organspendeausweis, Fahrsicherheitstraining oder Patientenverfügung aufgeführt wurden.

Wirtschaftlichkeit und Effizienz des Bonusprogramms (6 Nennungen) sowie die Beliebtheit der einzelnen Maßnahmen, abgeleitet davon wie häufig Versicherte sie durchführen (2 Nennungen). Vier Krankenversicherer verweisen bei der Frage nach der Gewichtung auf ihre Satzung.

Die Teilnehmer an Bonusprogrammen bekommen ihren Stand am Häufigsten in Form eines Punktestandes mitgeteilt. Zwei Krankenkassen arbeiten zusätzlich mit Symbolen (z. B. Ampel oder Farbskalen), eine bescheinigt das Erreichen von Zielmarken (z. B. „Gold-Status“) und eine spricht ausformulierte Empfehlungen an den Versicherten aus. Üblich ist bei 10 Krankenkassen auch ein Bonusheft (auch Bonuskarte, Bonuspass), in dem die absolvierten Aktivitäten festgehalten werden. Bei ebenso vielen Bonusprogrammen erfolgt eine Mitteilung über das Erreichen eines bestimmten Standes erst bei Auszahlung der Prämie.

Am ehesten verbessern können die Versicherten ihren Punktestand bei den meisten Bonusprogrammen mit ärztlichen (Früherkennungs-)Untersuchungen (9 Bonusprogramme). Sportliche Aktivitäten wurden von sechs Versicherern als die Maßnahme genannt, mit welcher der Punktestand am ehesten verbessert werden kann. Gesundheitskurse wurden von zwei Befragten genannt.

Bei der Entwicklung und Evaluierung der Bonusprogramme nehmen laut gegebener Aussagen über 70% der Krankenkassen externe wissenschaftliche Beratungen in Anspruch. Die meisten Versicherer sprechen von einer regelmäßigen Evaluation durch externe Dienstleister, vor allem in Hinblick auf die Berichte für das BVA. Sehr häufig wurde beispielsweise „Team Gesundheit – Gesellschaft für Gesundheitsmanagement GmbH“ genannt. Beim „Team Gesundheit“ handelt es sich um eine Arbeitsgemeinschaft der Krankenkassen nach § 94 SGB X, Gesellschafter sind insgesamt 17 Betriebskrankenkassen¹⁰⁰.

Zusammenhang zwischen Bonusprogrammen und Gesundheit

Wenig aussagekräftig waren die Antworten auf die Frage, ob die Krankenkassen Erkenntnisse darüber haben, dass sich der Gesundheitszustand Ihrer Versicherten durch die Teilnahme am Bonusprogramm ändert. Die große Mehrheit (35 Versicherer) verweist in diesem Zusammenhang auf regelmäßige Evaluationen zum Nachweis der Wirtschaftlichkeit ihrer Programme (dazu sind sie nach §65aa Abs. 3 AGB V verpflichtet). Der Gesundheitseffekt werde demnach indirekt durch den Nachweis der Senkung der Leistungsausgaben/Gesundheitskosten von Teilnehmern gezeigt.

Vier Versicherer geben an, sie hätten direkt eine Veränderung des Gesundheitszustandes durch die Teilnahme am Bonusprogramm bei ihren Versicherten festgestellt. Als Methode der Evaluierung wurden von ihnen Kontrollgruppenvergleiche genannt, die nicht näher beschrieben werden. Ein Versicherer gibt eine Kundenbefragung an, die Gesundheitseffekte gezeigt hätte.

Drei Versicherer bemerkten ausdrücklich, dass keine Erkenntnisse zu einem veränderten Gesundheitszustand durch Bonusprogramme vorlägen bzw. dass für solche Aussagen Langzeitstudien notwendig wären.

Digitalisierung der Bonusprogramme und Einsatz von Fitness-/Gesundheitsapps

Sieben Versicherer bieten ihren Bonusprogramm-Teilnehmern die Möglichkeit, ihren Punktestand im Bonusprogramm digital, z. B. über eine App, einzusehen und absolvierte Aktivitäten zu erfassen. Eine Veröffentlichung des Punktestandes online oder ein Vergleich der Punktestände in Form von Rankings z. B. über eine App oder in einem Onlineportal, bietet keiner der an der Befragung teilnehmenden Versicherer.

Drei Krankenkassen ermöglichen es ihren Versicherten Bonuspunkte zu sammeln, indem mittels Fitnessapps erfasste Daten verwendet werden. Bei allen drei wird die Schrittzahl des Versicherten innerhalb eines festgelegten Zeitraums (z. B. pro Tag) erfasst. Bei zwei Krankenkassen werden darüber hinaus auch Herzfrequenz und

Kalorienverbrauch gemessen¹⁰¹. Ist ein vorgegebener Schwellenwert erreicht, erhält der Versicherte Punkte.

Um Schrittzahl, Herzfrequenz oder Kalorienverbrauch zu erfassen, benötigt der Nutzer einen Fitness-Tracker (auch Fitnessarmband) und eine auf dem Smartphone installierte App eines Drittanbieters (z. B. Google Fit, Apple Health oder Samsung Health). Die App des Bonusprogramms selbst erfasst und speichert laut zwei Versicherern keine Fitness- oder Vitaldaten: „Datum, Typ der Aktivität sowie die erzielten Prämienpunkte, die aus der entsprechenden Aktivität ermittelt werden, werden gespeichert, wenn eine Zielvorgabe erreicht ist. Es erfolgt keine Speicherung von Bewegungsdaten, z. B. der genauen Schrittzahl“, schreibt einer von ihnen. Der dritte Versicherer erklärt, die genaue Schrittzahl würde von der App des Bonusprogramms abgerufen und „temporär lokal ausgewertet“. Die Versicherung würde darüber keine Informationen erhalten.

Auch wenn bislang nur sehr wenige Versicherungen den Einsatz von Fitness-Trackern in ihren Bonusprogrammen ermöglichen, gewähren elf Krankenkassen in der Befragung ihren Versicherten eine Bezuschussung zur Anschaffung solcher Geräte. Gesammelte Bonuspunkte können gegen einen Zuschuss zum Erwerb eingetauscht werden oder die Fitnessarmbänder können direkt als Sachprämie ausgewählt werden.

Akzeptanz von Malussystemen

Alle Bonusprogramme ermöglichen Boni für gesundheitsförderliche Maßnahmen und Aktivitäten. Doch wie schätzen die Krankenkassen (sowohl gesetzlich als auch privat) die Möglichkeit ein, bei der Ausgestaltung von Bonusprogrammen oder aber auch bei der Prämiengestaltung von Krankenversicherungstarifen neben gesundheitsförderlichen auch gesundheitsschädliche Verhaltensweisen von Versicherten zu berücksichtigen? Fünf gesetzliche Krankenkassen sprechen sich für ein solches Vorgehen aus.

Malussysteme würden die Eigenverantwortung der Versicherten stärken (2 Nennungen) und ein besserer ge-

sundheitlicher Effekt könne erzielt werden (1 Nennung). „Bei der Ausgestaltung von Bonusprogrammen ist es aus unserer Sicht durchaus vorstellbar auch gesundheitsschädliche Verhaltensweisen bei der Ermittlung des Bonus entsprechend zu berücksichtigen. Durch ein Bonus-Malus-System bei einem freiwilligen Bonusprogramm kann sicherlich ein besserer Effekt als bei einem reinen Bonus-System erzielt werden, da der Ansatz dann ganzheitlicher ist“, erläutert ein Versicherer.

Die große Mehrheit (55 Versicherungen, das sind fast 90% der Antwortenden) sprach sich gegen ein Malussystem aus. Hauptgründe: das Solidaritätsprinzip würde gefährdet, Diskriminierung und Ausgrenzung drohten (17 Nennungen). Zudem wurden datenschutzrechtliche Bedenken genannt (10 Nennungen). Außerdem wurde angeführt, dass es schwierig sei Gesundheitsschädlichkeit zu definieren oder gesundheitsschädliche Aktivitäten zu erfassen und zu kontrollieren (7 Nennungen). Diese Schwierigkeiten sind sicherlich vorhanden, sie gelten für Bonus-Systeme freilich gleichermaßen. Darauf geht an dieser Stelle keine der antwortenden Versicherer ein.

Schließlich waren einige Versicherer der Meinung, Versicherte würden ein solches Malussystem nicht akzeptieren oder würden dadurch demotiviert (3 Nennungen). Auch erlaube der gesetzliche Rahmen ein Malus-System gar nicht (2 Nennungen). Elf Versicherungen sprechen im Zusammenhang mit Bonus-Malus-Systemen und Datenschutz von Überwachung: „Überwachung der Lebensführung von Versicherten kann aus datenschutzrechtlichen Aspekten nicht Aufgabe der gesetzlichen Krankenkassen sein.“

Vor- und Nachteile von verhaltensbasierten Tarifen in der Krankenversicherung

Auch wenn verhaltensbasierte Tarife auf dem deutschen Krankenversicherungsmarkt noch nicht existieren, ist die grundsätzliche Einstellung der Versicherer zu diesem Thema relevant. Gefragt wurde deswegen, welche Vorteile gesetzliche und private Krankenversicherer im potenziellen Einsatz von verhaltensbasierten Tarifen für ihre Branche grundsätzlich sehen könnten.

101

Ein Versicherer beschreibt: Innerhalb eines Zeitraums von mindestens 30 Minuten muss die durchschnittliche Herzfrequenz bei mindestens 111 Schlägen pro Minute liegen bzw. müssen mindesten 150 kcal verbraucht werden.

Rund 54% der Versicherer, die diese Frage beantworteten, sehen überhaupt keine Vorteile in solchen Tarifmodellen und lehnen verhaltensbasierte Tarife grundsätzlich ab. Von manchen Versicherern wurden jedoch einige Vorteile aufgezählt.

An erster Stelle in den sich für verhaltensbasierte Tarife aussprechenden Antworten steht das Argument, eine solche Differenzierung würde einen Anreiz zu gesundheitsförderlichem Verhalten stiften und das Gesundheitsbewusstsein der Versicherten steigern (17 Nennungen). Weiterhin genannt wurden von sieben Versicherern der Aspekt Kostenersparnis (Senkung der Leistungsausgaben/Krankheitskosten), von vier Versicherern der Aspekt Wettbewerbsvorteile und Kundenbindung. Ein privater Versicherer sieht in verhaltensbasierten Tarifen die Möglichkeit eines besseren Leistungsangebots: „Verhaltensbasierte Tarife würde eine Entwicklungsrichtung weg vom reinen Kostenerstatter und hin zu einem Gesundheitsmanager unterstützen.“

Und welche Nachteile sehen die befragten Versicherer im potenziellen Einsatz verhaltensbasierter Tarife? Immerhin fünf Krankenversicherer (alle aus der GKV) sehen gar keine Nachteile (das sind knapp 8% aller Befragten). Fast 50% (30 Versicherer) sehen in verhaltensbasierten Tarifen eine Gefährdung des Solidaritätsprinzips und eine mögliche Benachteiligung und Stigmatisierung bestimmter Versichertengruppen (z. B. genetisch bedingt Kranke). Viele sehen auch die Gefahr, dass es zu Auswirkungen auf die Risikoselektion (eine Risiko-Entmischung) kommt (12 Nennungen, davon überdurchschnittlich viele PKV). „Durch die möglicherweise eintretende Antiselektion kann eine gute Absicherung der Gesundheit für Personen, die sich nicht für verhaltensbasierte Tarife entscheiden, gegebenenfalls zu deutlich erhöhten Prämien führen. Hintergrund ist, dass sich gute Risiken eher für die günstigeren verhaltensbasierten Tarife entscheiden und die Beiträge für die Personen in den nicht verhaltensbasierten Tarifen damit ansteigen werden, da sich die Risikomischung verschlechtert“, beschreibt ein Versicherer.

Ebenfalls genannt wird der Aspekt hoher Kosten und Verwaltungsaufwände die mit der praktischen Umsetzung verhaltensbasierter Tarife eingehen würden (15 Nennungen).

Auch datenschutzrechtliche Bedenken werden geäußert (5 Nennungen).

Ebenfalls als negative Aspekte im Zusammenhang mit verhaltensbasierten Tarifen gesehen werden unter anderem eine befürchtete Monopolbildung, weil kleinere Kassen bei der Umsetzung abgehängt werden (2 Nennungen), Missbrauch der Tarife als Marketinginstrument (2 Nennungen), oder Fehlanreize, wenn Versicherte Arztbesuche vermeiden, um ihre Prämien nicht zu verlieren (2 Nennungen).

Anpassung des Gesetzes- und Regulierungsrahmen

Im Zusammenhang mit Bonusprogrammen und/oder verhaltensbasierten Tarifen wurden die Versicherer gefragt, ob sie Bedarf sehen, den Gesetzes- und Regulierungsrahmen anzupassen. Zwei Drittel (41 Versicherer) sehen hierfür keinen Änderungsbedarf. Anbieter, die für einen Anpassungsbedarf votieren (17), argumentieren unter anderem, dass der Gestaltungsspielraum bei der Ausgestaltung der Bonusprogramme erweitert werden sollte (13 Anbieter), beispielsweise in Hinblick auf gemeinnützige Aktivitäten oder den Einsatz von Fitness-Trackern gebe es unterschiedliche Zulassungspraxen zwischen den landesweiten Aufsichtsbehörden und dem für die bundesweiten Krankenkassen zuständigen Bundesversicherungsamt. „Das BVA sollte den Krankenkassen die Möglichkeit geben, gemeinnütziges Verhalten im Rahmen der Bonusmodelle zu belohnen“, wird z. B. gefordert.

Verschiedene Forderungen finden sich zum Thema Daten: Datensouveränität für die Versicherten (der Versicherte entscheidet selbst, wem er seine Daten zu welchem Zweck zur Verfügung stellt) fordern drei Versicherer. Eine Lockerung des Datenschutzes befürwortet ein Versicherer. Ein anderer Versicherer hegt den Wunsch, dass eine Verknüpfung von Datenpools möglich wird: „Erleichtert werden sollte aus unserer Sicht die anonymisierte Auswertung/ Verknüpfung großer Datenpools, um bessere Vorhersagen inkl. Präventivmaßnahmen zu treffen sowie die gezielte Ansprache von Versicherten auf die Teilnahme an Programmen/ Maßnahmen, auch wenn diese hierzu ihre Einwilligung noch nicht ausdrücklich gegeben haben.“

Von zwei privaten Versicherern wird angeführt, dass für eine Einführung verhaltensbasierter Tarife eine Änderung der Kalkulationsvorgaben notwendig sei. Jeweils ein Versicherer fordert den verstärkten Schutz der Solidargemeinschaft und sogar die Abschaffung der Bonusprogramme. „Die Prämienhöhe und -auswahl einzelner Krankenkassen erweckt den Eindruck, dass es sich um reine Marketing- und Kundengewinnungsprogramme handelt, bei denen der einzelne Teilnehmer einen unverhältnismäßig hohen Benefit aus Versichertengeldern der Solidargemeinschaft erhält. Hier sollte regulatorisch eingegriffen werden“, schreibt einer der Befragten – insgesamt sprechen sich aus diesem Grund zwei Versicherer für mehr Regulierung aus.

Daten aus der elektronischen Patientenakte

Die elektronische Patientenakte (auch elektronische Gesundheitsakte) ist eine Datenbank, in der Gesundheitsdaten und elektronische Dokumente wie Arztbrief, Medikationsplan, Notfalldatensatz, Impfausweis etc. gesammelt und gespeichert werden. Arztpraxen, Kliniken und andere Gesundheitseinrichtungen haben nach Zustimmung der Patienten bei Bedarf schnellen Zugriff. Ab 2019 soll sie laut dem sogenannten „E-Health-Gesetz“ als freiwillige Anwendung für alle gesetzlich Versicherten zur Verfügung stehen.

Insofern ist die Frage interessant, ob bei den Krankenversicherungen ein Interesse besteht, die elektronische Patientenakte und die damit potenziell zur Verfügung stehenden Gesundheitsdaten von Versicherten für den Bereich Bonusprogramme und/oder verhaltensbasierte Tarife in der Krankenversicherung zu nutzen. Die Meinungen gehen hier auseinander.¹⁰²

23 Versicherer würden die Patientenakte für Bonusprogramme bzw. verhaltensbasierte Tarife nutzen, 32 Versicherer lehnen dies in der Befragung ab. Viele von den Befürwortern betonen, dies solle nur auf Wunsch bzw. nach Zustimmung der Versicherten möglich sein (13 Nennungen). Sie versprechen sich durch die Nutzung der Daten aus der elektronischen Patientenakte eine bessere Versorgung (4 Nennungen) oder besseren

Service bzw. bessere Informationen für ihre Kunden (1 Nennung). Auch ein verringerter Verwaltungsaufwand wird von einem Versicherer als Argument angebracht: „Derzeit haben wir jede eingereichte Maßnahme (...) sehr verwaltungsaufwändig zu prüfen und freizugeben. Hier sollte eine nachträgliche Stichprobe per Abgleich mit der Patientenakte ermöglicht werden“. Ein Versicherer würde gerne die Daten nutzen, aber nur in anonymisierter Form als „valide Datenbasis für die Entwicklung und Ausgestaltung von Bonusprogrammen und ggf. verhaltensbasierten Tarifen“.

Die meisten der Befragten sprechen sich gegen die Verwendung der elektronischen Patientenakte für Bonusprogramme oder verhaltensbasierte Tarife aus. Begründet wurde die Haltung nur wenig. Einige führen datenschutzrechtliche Bedenken an (4 Nennungen). Ein Versicherer führt aus: „Nein – das würde das Schreckensszenario des „Gläsernen Patienten“ befeuern. Dies ist aus Sicht der EU-DSGVO und auch der BfDI unvorstellbar.“ Ein Versicherer befürchtet eine Benachteiligung kranker Versicherter.

Ausblick

Wie schon bei den Auskunfteien und bei den Telematik-Tarifen, wurden auch die gesetzlichen und privaten Krankenversicherern nach ihrer Einschätzung zukünftiger Entwicklungen gefragt. Gefragt wurde, wie sie die Zukunft von verhaltensbasierten Krankenversicherungstarifen bewerten und ob es Trends (z. B. aus dem Ausland) gibt, die sie für die Weiterentwicklung Ihrer Krankenversicherung für relevant halten.

Im Bereich der verhaltensbasierten Tarife sehen neun Versicherer eine eher positive Markttendenz. Beispielsweise schreibt einer der Befragten: „Auch der deutsche Gesundheitsmarkt wird sich dem nicht verschließen können, nur es dauert in Deutschland alles länger.“

Demgegenüber stehen 19 Versicherer, die eher keine Markttendenz wahrnehmen. Einige sagen aus, solche Entwicklungen wären wenn, dann höchstens in der PKV denkbar, in der gesetzlichen Krankenversicherung gebe es diese Möglichkeiten überhaupt nicht. Einige führen

eine starke gesetzliche Regulierung an, die eine Einführung verhaltensbasierter Tarife in Deutschland verhindern: „Aufgrund der sozialpolitischen Verankerung und dem entsprechenden engen aufsichtsrechtlichen Rahmen sehen wir das Thema (...) nicht als relevant an.“

Etwa ebenso viele Versicherer (20) äußern ausdrücklich eine Ablehnung gegenüber verhaltensbasierten Tarifen. Besonders in der gesetzlichen Krankenversicherung wird das Solidaritätsprinzip unterstrichen: „Verhaltensbasierte Tarife können dazu führen, dass einzelne Versicherungsträger dies auf Kosten von Menschen, deren Krankheiten nicht verhaltensbedingt sind, ausnutzen. Insofern sehen wir hier sehr kritisch drauf.“

Verschiedene Zukunftstrends werden weiterhin von den Krankenversicherern interessiert wahrgenommen: Zum einen wird der technische Fortschritt und die zunehmende Digitalisierung (z. B. Telemedizin) als wichtiger Trend genannt (4 Nennungen). Zum anderen würde der Präventionsgedanke in der Krankenversicherung stärker werden (3 Nennungen). „Die allgemeine Entwicklung der Branche wird sich immer mehr weg vom Versicherer als reinen Risikoträger und Kostenerstatter und hin zum Gesundheitsmanager/verlässlichen Partner für das eigene Wohlbefinden entwickeln“, wird an einer Stelle formuliert.

Auch wird argumentiert, dass Gesundheitsbewusstsein der Versicherten nehme zu und der Umgang mit Gesundheitsdaten würde sich verändern. Im Zuge dieser Entwicklung würden die Versicherten verhaltensbasierte Versicherungen und personalisierte Angebote erwarten (2 Nennungen): „Die Digitalisierung verbunden mit dem Trend der Selbst-Optimierung der Menschen wird auch Auswirkungen in der Gesundheitswirtschaft haben. Die Skepsis gegenüber der Datenerfassung und -weitergabe bei digitalen Anwendungen und Therapien wird abnehmen. Wenn es zum Alltag gehört, seine Gesundheitsdaten zu teilen werden Kunden auch erwarten, individuelle Präventionsangebote und Gesundheitsservices zu erhalten. Künftig muss es deshalb mehr Spielraum geben, um entsprechende Angebote und Tarife zu gestalten.“

Schließlich würden neue Geschäftsmodelle und Versicherungsprodukte entstehen und verhaltensbasierte Tarife zum Unterscheidungsmerkmal im Wettbewerb werden (2 Nennungen).

IV. Tabellenband (Marktbefragung)

1. Auskunfteien

Tabelle IV.1

Gibt es Voraussetzungen, die Geschäftskunden erfüllen müssen, um von Ihrem Unternehmen Auskunft über die Bonität eines Verbrauchers zu erhalten? Falls ja, welche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Vertrag	Geschäftskunden müssen einen Vertrag mit der Auskunftei abschließen	3
Berechtigtes Interesse	Geschäftskunden müssen ein berechtigtes Interesse nachweisen	3

Tabelle IV.2

Mit welcher Maßnahme können Verbraucher ihren Bonitäts-Score am ehesten verbessern?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Verbindlichkeiten ausgleichen	Rechnungen bezahlen; Kreditverpflichtungen erfüllen; Zahlungstörungen meiden	3
Inkassoverfahren/gerichtliche Mahnverfahren meiden		1

Tabelle IV.3

In welcher Form teilen Sie Privatpersonen das Ergebnis Ihres Bonitäts-Scorings mit?	
ANTWORT	ANZAHL (n=3)
Punktstand	3
Wahrscheinlichkeitswert	3
Symbole (z. B. Ampel, Farbskalen)	1
Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher	1

Tabelle IV.4

In welcher Form teilen Sie Geschäftskunden das Ergebnis Ihres Bonitäts-Scorings mit?	
ANTWORT	ANZAHL (n=3)
Punkttestand	3
Wahrscheinlichkeitswert	3
Symbole (z. B. Ampel, Farbskalen)	2
Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher	1

Tabelle IV.5

Auf welcher Skala (z. B. 1-100) bewegt sich der Bonitäts-Score, den ihr Unternehmen errechnet?	
ANTWORT	ANZAHL (n=3)
0-1000	2
100-600	1
1-999	1
1-9999	1

Tabelle IV.6

Aus welchen Quellen bezieht Ihr Unternehmen die Daten zur Berechnung des Bonitäts-Scores?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Öffentliche Register und Verzeichnisse/amtliche Bekanntmachungen	z. B. Inkassomeldungen, Insolvenzinformationen und Schuldnerverzeichnisse; öffentlich zugängliche Adressdaten	3
Vertragspartnern	Daten von Vertragspartnern/kooperationspartner; z. B. Haushaltsdatenbanken, Banken, Warenhändler, Telekommunikationsgesellschaften, Energieversorger	2

Tabelle IV.7

Bitte geben Sie an, ob die Berechnung des Bonitäts-Scores in Ihrem Unternehmen selbst oder durch einen externen Dienstleister erfolgt.	
ANTWORT	ANZAHL (n=3)
Im Unternehmen selbst	3
Bei einem externen Dienstleister	0

Tabelle IV.8

Wie viele Merkmale fließen immer in die Berechnung des Bonitäts-Scores ein?	
ANTWORT	ANZAHL (n=3)
> 25	1
> 8	1
Keine Aussage möglich	1

Tabelle IV.9

Welche Merkmale fließen immer in die Berechnung des Bonitäts-Scores ein?	
ANTWORT	ANZAHL (n=3)
Alle gespeicherten/erhobenen Daten	2
Zahlungsverhalten	1
Zahlungserfahrungen im unmittelbaren Wohnumfeld	1
Alter	1
Geschlecht	1
Umzugsverhalten	1

Tabelle IV.10

Welche Merkmale können darüber hinaus (z. B. für branchenspezifische Scores) in die Berechnung des Bonitäts-Scores einfließen?	
ANTWORT	ANZAHL (n=3)
Alle gespeicherten/erhobenen Daten	3

Tabelle IV.11

Fließen Daten zur Wohngegend (Geodaten) in die Berechnung des Bonitäts-Scores mit ein?	
ANTWORT	ANZAHL (n=3)
Ja, immer	1
Ja, unter Umständen (wenn sonst keine kreditrelevanten Daten vorliegen)	1
Nein, nie	1

Tabelle IV.12

Fließen Daten aus Onlineverhalten (z. B. aus sozialen Netzwerken) in die Berechnung des Bonitäts-Scores mit ein?	
ANTWORT	ANZAHL (n=3)
Ja, immer	0
Ja, unter Umständen	0
Nein, nie	3

Tabelle IV.13

Anhand welcher Kriterien wird entschieden, welche Merkmale für die Berechnung des Bonitäts-Scores relevant und aussagekräftig sind?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Statistische Verfahren	Die Merkmalsauswahl beruht auf mathematisch-statistischen Verfahren; Signifikanztests	3
Datenschutzrechtliche Beurteilung		1
Verfügbarkeit	„Verfügbarkeit des Merkmals im operativen Betrieb“	1

Tabelle IV.14

Anhand welcher Kriterien wird entschieden, mit welcher Gewichtung Merkmale in die Berechnung des Bonitäts-Scores einfließen?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Statistische Verfahren	Die Festlegung der Gewichtungen beruht auf mathematisch-statistischen Verfahren (logistische Regression)	3
Erfahrung	Erfahrungswerten bzgl. der Stabilität der Parameter	1

Tabelle IV.15

Gibt es individuelle Merkmale, die nicht direkt erfasst oder gemessen, sondern durch andere Merkmale approximiert werden?	
ANTWORT	ANZAHL (n=3)
Nein	3
Ja	0

Tabelle IV.16

Welche Schätzmethoden liegen dem Scoring zugrunde?	
ANTWORT	ANZAHL (n=3)
Logistische Regression	3
Clusterbildende Methoden (z. B. kMeans)	0
Entscheidungsbäume	1
Ensemblemethoden (z. B. Boosting, Random Forest)	1
Neuronale Netzwerke/Deep Learning	1

Tabelle IV.17

Wird für die Schätzmethode ein Zielmerkmal (abhängige Variable) verwendet? Falls ja, benennen Sie bitte das Zielmerkmal.	
ANTWORT	ANZAHL (n=3)
Auftreten eines Negativmerkmals innerhalb eines bestimmten Zeitraums	2
Keine Antwort	1

Tabelle IV.18

Wie wird die Güte des verwendeten Schätzmodells gemessen?	
ANTWORT	ANZAHL (n=3)
Area Under the Curve (AUC)	0
Gini-Koeffizient	3
Quadratischer Fehler/Brier Score	0
KS-Statistik (= Kolmogorov-Smirnov Test)	1
F-Measure	0
Lift	0

Tabelle IV.19

Wie hoch ist der Wert des genannten Gütemaßes?	
ANTWORT	ANZAHL (n=3)
60–75% im Gini-Koeffizient	1
60–80% im Gini-Koeffizient	1
Keine Antwort möglich	1

Tabelle IV.20

Welche Maßnahmen werden ergriffen, um die Güte des Modells aufrechtzuerhalten? Aus welchem Anlass wird das Schätzmodell verändert?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Regelmäßige Validierung	Überprüfung des Scorekarten hinsichtlich Stabilität und Trennschärfe	3

Tabelle IV.21

Welche Maßnahmen werden ergriffen, um die Güte des Modells zu verbessern?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Evaluation in Bezug auf Modellvalidierung	Regelmäßige Evaluation auf Basis der Ergebnisse der Validierung	3
Evaluation in Bezug auf datenschutzrechtlichen Rahmenbedingungen		1
Evaluation in Bezug auf Nutzbarkeit/Verfügbarkeit		1
Evaluation in Bezug auf statistische Relevanz und Zusatznutzen		1
Evaluation in Bezug auf Kosten-Nutzen Aspekte		1

Tabelle IV.22

Welche Daten, die nicht dem Zweck der Score-Berechnung dienen, werden zusätzlich über Verbraucher gespeichert?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Anfragen & Auskünfte	Protokollierung der Anfragen und erteilten Auskünfte	3
Name und Wohnort von Verbraucher	Wird nicht fürs Scoring verwendet	1

Tabelle IV.23

Nehmen Sie bei der Entwicklung und Evaluierung Ihres Scoring-Verfahrens externe wissenschaftliche Beratung in Anspruch?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Ja	Prüfung der Score-Verfahren durch externe Gutachter	2
Nein	Entwicklung und Evaluierung im Unternehmens selbst	1

Tabelle IV.24

Wie erfahren Verbraucher, welche Daten über sie in die Berechnung ihres Bonitäts-Scores eingeflossen sind?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Selbstauskunft	Schriftliche Datenübersicht	3
Vor Ort	Persönliches Gespräch vor Ort	1
Verbraucherservice	Schriftlich oder telefonisch	1
Informationen auf der Webseite	Informationsmaterial	1

Tabelle IV.25

Welche Maßnahmen werden ergriffen, um die Aktualität und Richtigkeit der von Ihrem Unternehmen zur Berechnung der Bonitäts-Scores verwendeten Daten zu gewährleisten?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Löschfristen	Nach Ablauf von Speicherfristen werden Daten gelöscht	3
Datenüberwachung	Kontinuierliche Überwachung des Datenbestandes; Datenpflege; Monitoring; statistische Überwachung	3
Schulungen	Schulungen bei Kunden zur Qualitätssicherung	1
Plausibilitätschecks		1
Verbraucherhinweise	Wenn Verbraucher auf fehlerhafte Daten hinweisen, werden diese korrigiert	1
Aktualisierungspflichten	Vertragspartner müssen Änderungen im Datenbestand melden	1

Tabelle IV.26

Welche Verfahren zur Fehlerkorrektur der Daten, die in die Score-Berechnung einfließen, stehen Verbrauchern zur Verfügung?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Reklamation	Bei Reklamationen/Beschwerden von Verbrauchern werden die Daten geprüft; Korrekturen möglich	3
Score-Sperre	Können Daten nach Beschwerde nicht korrigiert werden, erfolgt eine Score-Sperre welche die Generierung und Übermittlung von Scores zu dem Betroffenen verhindert	1
Verbraucherservice	Es gibt einen Verbraucherservice, der bei Problemen kontaktiert werden kann	2
Schlichtungsstelle	Hinweis auf eine Schlichtungsstelle	1

Tabelle IV.27

Bezüglich welcher Merkmale weisen Verbraucher Ihr Unternehmen am häufigsten auf fehlerhafte Daten hin?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Adressdaten	Adresse ist nicht aktuell; falsche Schreibweise der Adresse	2
Name	Falsche Schreibweise von Namen	2
Inkassomerkmale	Inkassodaten sind nicht aktuell	2
Negativ-Informationen	Fehlerhafte Negativ-Informationen zum Zahlungsverhalten	2
Nur wenig Reklamationen	Es gibt insgesamt nur wenig (berechtigte) Reklamationen	3

Tabelle IV.28

Halten Sie es für möglich, dass die Ermittlung des Bonitäts-Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten kann? Falls ja, wie begegnen Sie diesem Problem?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Nein		3
Statistische Verfahren diskriminieren nicht	„Mathematisch, statistische Verfahren diskriminieren nicht.“	1
Geschlecht nur bei Trennschärfe	„Geschlechtsspezifische Score-Variablen finden ausschließlich Verwendung sofern sie trennscharf sind.“	1
Ausschluss von diskriminierenden Parametern	diskriminierende Parameter wie Nationalität, Beruf, Einkommen oder Familienstand werden ausgeschlossen	1

Tabelle IV.29

Welche Aufsichtsbehörden sind für die Prüfung Ihres Scoring-Verfahrens zuständig? Wie oft und aus welchen Anlässen wird dieses geprüft?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Zuständige Landesdatenschutzbeauftragte		3
Dokumentation der Scoring-Verfahren	Datenschutzbeauftragtem wurde ausführliche Dokumentation des Verfahrens vorgelegt und die Ergebnisse der regelmäßig stattfindenden Validierungen der Scorekarten mitgeteilt; Bevor eine Scorekarte in den Betrieb aufgenommen wird, wird sie der Aufsichtsbehörde vorgestellt	2
Regelmäßige Treffen	regelmäßige Treffen mit dem Datenschutzbeauftragten und Überprüfung von Beschwerden vor Ort	1
Sämtlichen Datenschutzbeauftragten des Bundes und der Länder	Offenlegung der Scorekarten gegenüber sämtlichen Datenschutzbeauftragten des Bundes und der Länder	1

Tabelle IV.30

Halten Sie die Erweiterung der Datenbasis Ihres Scoring-Verfahrens um Merkmale aus anderen Bereichen (z. B. riskantes Fahrverhalten oder Gesundheitsverhalten) für sinnvoll, um die Prognosequalität zu verbessern?

ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Nein	Erweiterung der Datenbasis wird als nicht sinnvoll/notwendig erachtet	3

Tabelle IV.31

Sehen Sie Bedarf, den Gesetzes und Regulierungsrahmen in Bezug auf Bonitäts-Scoring anzupassen (z. B. im Bereich des Datenschutzrechts)? Falls ja, in welcher Hinsicht?

ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
BDSGneu & DSGVO	Es gibt Klärungsbedarf im Hinblick auf BDSG n. F. und DSGVO	3

Tabelle IV.32

Wie schätzen Sie die Zukunft von Bonitäts-Scoring-Verfahren ein? Gibt es Trends z. B. aus dem Ausland, die Sie für die Weiterentwicklung Ihres Geschäftsmodells für relevant halten?

ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Daten aus sozialen Netzwerken	Im Ausland werden bereits Daten aus sozialen Netzwerken für das Bonitäts-Scoring genutzt; Entwicklungspotenzial wird hier gesehen; Wenn nach DSGVO zulässig, sollte das so bleiben	3
Betrugsprävention	Der Bereich der Betrugsprävention wird in Zukunft wichtiger	2

2. Kfz-Versicherungen

Tabelle IV.33

Gibt es Voraussetzungen, die Kunden erfüllen müssen, um bei Ihrem Unternehmen eine Kfz-Versicherung mit Telematik-Tarif abschließen zu können? Falls ja, welche?	
ANTWORT	ANZAHL (n=10)
Altersvoraussetzungen	6
Jahre seit Führerscheinerwerb	2
Technische Voraussetzungen	9
Sonstiges	4

Tabelle IV.34

Gibt es eine Zielgruppe, die mit dem Telematiktarif ihres Unternehmens besonders angesprochen werden soll? Falls ja, welche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Keine	Keine besondere Zielgruppe	3
Junge Fahrer	Junge Fahrer, unter einem bestimmten Alter	6
Fahranfänger	Fahrer, die den Führerschein erst weniger als eine bestimmte Anzahl an Jahren haben	2
Teure Tarife	Fahrer, mit teuren Verträgen	1

Tabelle IV.35

Wie groß ist der Anteil Ihrer Kunden in der Kfz-Versicherung, die an dem Telematik-Tarif teilnehmen? Zeichnen sich die teilnehmenden Kunden im Vergleich zu nicht-teilnehmenden Kunden durch besondere Merkmale (z. B. Alter, Geschlecht) aus?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Wenige	Im Vergleich mit Zahl der Gesamtversicherten nur wenig Teilnehmer, z. B. <1%	8
Keine besonderen Merkmale	Keine besonderen Merkmale bei den Telematik-Versicherten; Vergleiche werden nicht durchgeführt	3
Jünger	Teilnehmer sind jünger; andere Altersstruktur auch dadurch, dass die Zielgruppe junge Fahrer sind	5
Mehr Männer	Angabe über die Geschlechterverteilung bei Telematik-Versicherten	1
Gutes Fahrverhalten	Die Teilnehmer am Telematik-Tarif fahren besser	2
Keine Antwort		1

Tabelle IV.36

Mit welchen in Aussicht gestellten Effekten motivieren Sie Kunden zur Teilnahme an Ihrem Telematik-Tarif?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Vergütung	Die Kunden sparen Geld oder erhalten Boni, Gutscheine etc.	8
Fahrweise & Sicherheit	Die Kunden erhalten Informationen zur Fahrweise, werden unterstützt; Kunden können ihr Fahren verbessern;	7
Service & Zusatzleistungen	Kunden bekommen Zusatzleistungen wie z. B. Notrufaktionen, Tankstellenfinder oder Diebstahlwarnungen	4
Vergleich mit anderen	Kunden können ihr Fahrverhalten mit anderen Fahrern vergleichen	1

Tabelle IV.37

Auf welcher Skala (z. B. 1–100) bewegt sich der Score, der im Rahmen des Telematik-Tarifs für einen Kunden berechnet wird?	
ANTWORT	ANZAHL (n=10)
1–100	9
Keine Skala	1

Tabelle IV.38

In welcher Form teilen Sie Kunden das Ergebnis Ihres Scorings mit?	
ANTWORT	ANZAHL (n=10)
Punkttestand	10
Wahrscheinlichkeitswert	0
Symbole (z. B. Ampel, Farbskalen)	5
Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher	2
Zielmarken (z. B. „Gold-Status“)	0
Sonstiges	0

Tabelle IV.39

Welche potenziellen Vorteile werden den Kunden gewährt, wenn Sie an dem Telematik-Tarif Ihres Unternehmens teilnehmen?	
ANTWORT	ANZAHL (n=10)
Rückzahlung von Versicherungsprämien	1
Günstigere Versicherungsprämien im Folgejahr	6
Geldleistungen	0
Sachleistungen	1
Gutscheine/Rabatte bei Partnerunternehmen	1

Tabelle IV.40

Dient der Score als das einzige Entscheidungskriterium zur Gewährung der oben genannten Vorteile?	
ANTWORT	ANZAHL (n=10)
Ja	10
Nein	0

Tabelle IV.41

Wie und wie oft wird der aktuelle Score dem Kunden mitgeteilt?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Nach jeder Fahrt	Der Score wird nach jeder Fahrt aktualisiert	5
Jederzeit	Der aktuelle Score ist jederzeit einsehbar	6
Monatlich	Der Score wird monatlich aktualisiert/mitgeteilt	1
App oder Online	Der Score wird über eine App oder ein Onlineportal mitgeteilt	7

Tabelle IV.42

Bieten Sie Ihren Kunden die Möglichkeit, ihren Score digital, z. B. über eine App, zu erfahren?	
ANTWORT	ANZAHL (n=10)
Ja	10
Nein	0

Tabelle IV.43

Bieten Sie Ihren Kunden die Möglichkeit, ihren Score online zu veröffentlichen?	
ANTWORT	ANZAHL (n=10)
Ja	3
Nein	7

Tabelle IV.44

Bietet Ihr Unternehmen die Möglichkeit, dass Kunden ihre jeweiligen Scores in Form eines Rankings miteinander vergleichen können (z. B. über eine App oder ein Onlineportal)?	
ANTWORT	ANZAHL (n=10)
Ja	3
Nein	7

Tabelle IV.45

Ändert sich die Schadenhäufigkeit Ihrer Kunden durch die Teilnahme an Ihrem Telematik-Tarif? Wie stellen Sie dies fest?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Keine Aussagen möglich	Es gibt darüber keine Erkenntnisse; Messung mit Vorher-Nachher-Vergleichen nicht durchgeführt/nicht möglich	9
Weniger Schäden	Telematik-Teilnehmer verursachen weniger Schäden	1
Risikoselektion	Es besteht die Möglichkeit, dass nur ohnehin sichere Fahrer Telematik-Tarife wählen	3

Tabelle IV.46

Welche Technik wird verwendet, um die Daten zur Berechnung der Scores zu erheben?	
ANTWORT	ANZAHL (n=10)
Telematik-Stecker für den Zigarettenanzünder	2
OBD-Telematik-Stecker	1
Ins Auto fest eingebaute Telematik-Box	3
App auf dem Smartphone	4

Tabelle IV.47

Zieht Ihr Unternehmen Daten aus weiteren Quellen zur Berechnung der Scores heran (z. B. digitales Kartenmaterial, Social Media Daten, Datenbanken wie GfK, Nielsen etc.)? Falls ja, welche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Kartenmaterial	Digitales Verkehrs- und Straßenkartenmaterial wird hinzugezogen	8
Nein	Keine Daten aus weiteren Quellen	2

Tabelle IV.48

Bitte geben Sie an, ob die Berechnung des Scores in Ihrem Unternehmen selbst oder durch einen externen Dienstleister erfolgt.	
ANTWORT	ANZAHL (n=10)
Im Unternehmen selbst	0
Bei einem externen Dienstleister	10

Tabelle IV.49

Welche Merkmale fließen immer in die Berechnung des Scores ein?	
ANTWORT	ANZAHL (n=10)
Geschwindigkeit/Geschwindigkeitsüberschreitungen	9
Bremsverhalten	10
Lenk- bzw. Kurvenverhalten	5
Beschleunigungsverhalten	10
Tag/Uhrzeit	9
Ort	5
Länge/Dauer	1
Straßentyp	7
Bevölkerungsdichte	2

Tabelle IV.50

Welche Merkmale können darüber hinaus in die Berechnung des Scores einfließen?	
ANTWORT	ANZAHL (n=10)
Keine	10

Tabelle IV.51

Anhand welcher Kriterien wird entschieden, welche Merkmale für die Berechnung des Scores relevant und aussagekräftig sind?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Auswertung von Unfallstatistiken	Auswertung von Statistiken zu Fahrweisen und Unfallursachen, z. B. vom Statistischen Bundesamt oder der Royal Society for the Prevention of Accidents	7
Erfahrungswerte anderer & Faustformeln	Orientierung an Erfahrungen in anderen Ländern oder anderen Unternehmen; Auswahl auf Basis menschlicher Intuition	4
Statistische Methoden & Tests	Die Signifikanz einzelner Merkmale wird mit statistischen Methoden/ Tests geprüft	4
Keine Antwort möglich	Unternehmen kann Frage nicht beantworten (z. B. weil ein anderes Unternehmen die Modelle entwickelt)	1

Tabelle IV.52

Anhand welcher Kriterien wird entschieden, mit welcher Gewichtung Merkmale in die Berechnung des Scores einfließen?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Auswertung von Unfallstatistiken	Auswertung von Statistiken zu Fahrweisen und Unfallursachen, z. B. vom Statistischen Bundesamt oder der Royal Society for the Prevention of Accidents	7
Erfahrungswerte anderer	Orientierung an Erfahrungen in anderen Ländern oder anderen Unternehmen	1
Statistische Methoden und Tests	Die Gewichtung der Merkmale ist Ergebnis statistischer Modelle und Analysen, die Aussagekraft der Merkmale ist für Gewichtung entscheidend; Ergebnisse unternehmensinterner Tests und Forschung	5
Keine Antwort möglich	Unternehmen kann Frage nicht beantworten (z. B. weil ein anderes Unternehmen die Modelle entwickelt)	1

Tabelle IV.53

Wird für die Schätzmethode ein Zielmerkmal (abhängige Variable) verwendet? Falls ja, benennen Sie bitte das Zielmerkmal.		
ANTWORT	BESCHREIBUNG	ANZAHL (n=3)
Schadeneintrittswahrscheinlichkeit	Die Wahrscheinlichkeit, dass ein Schaden eintreten wird	2
Schadenhäufigkeit und Schadenbedarfe	Wie häufig Schäden vorkommen werden	2
Keine Aussage möglich		1
Keine Antwort		5

Tabelle IV.54

Welche Maßnahmen werden ergriffen, um die Güte des Modells aufrechtzuerhalten? Aus welchem Anlass wird das Schätzmodell verändert?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Modellgüte	Anpassungsmaßnahmen richten sich nach dem Gütemaß; bei unzureichender Modellgüte wird angepasst	2
Neue Erkenntnisse zur Vorhersagegenauigkeit	Es erfolgt ein Abgleich mit tatsächlichen Schadendaten	2
Keine Aussage möglich		1
Keine Antwort		5

Tabelle IV.55

Nehmen Sie bei der Entwicklung und Evaluierung Ihres Scoring-Verfahrens externe wissenschaftliche Beratung in Anspruch? Falls ja, in welcher Form?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Ja	Inanspruchnahme externer wissenschaftlicher Entwicklung und Evaluierung	2
Nein	Keine, oder nur interne wissenschaftliche Beratung	3
Keine Antwort		5

Tabelle IV.56

Wie erfahren Kunden Ihres Unternehmens, welche Daten über sie in die Berechnung ihres Scores einfließen?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Vertragsunterlagen	Allgemeinen Bedingungen für die Kfz-Versicherung (AKB); Nutzungsbedingungen oder Nutzungsvereinbarungen; Datenschutzinformationen oder Datenschutzerklärung;	9
Informationsmaterial	Informationen und Veröffentlichung auf der Homepage/Webseite oder in der App; Handbuch oder Informationsblätter, die dem Kunden zur Verfügung gestellt werden	7
Beratung durch Vermittler	Persönliche Beratung, z. B. durch Versicherungsvermittler	1

Tabelle IV.57

Welche Maßnahmen werden ergriffen, um die Aktualität und Richtigkeit der von Ihrem Unternehmen zur Berechnung der Scores verwendeten Daten zu gewährleisten?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Aktualisierung & Qualitätschecks	Daten unterliegen Qualitätschecks; Verkehrsdaten und Kartenmaterial werden z. B. regelmäßig aktualisiert	6
Plausibilitätsprüfung	Fahrdaten werden auf Plausibilität geprüft	2
Funktionsprüfung	Die Technik wird auf Funktionalität geprüft, z. B. mit Betriebsleuchten	2
Echtzeitverarbeitung	Die Daten werden in Echtzeit verarbeitet	1
Toleranzen für Messfehler	Messfehler sind nicht auszuschließen und müssen zu einem gewissen Grad in Kauf genommen werden bzw. bei der Modellbildung von vornherein berücksichtigt werden	2
Keine Aussage möglich		1

Tabelle IV.58

Welche Verfahren zur Fehlerkorrektur der Daten, die in die Score-Berechnung einfließen, stehen Kunden zur Verfügung?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Korrekturmöglichkeiten vorhanden	Korrekturen können vom Kunden veranlasst werden und werden auch durchgeführt; Korrekturen werden aus Kulanz vorgenommen	9
Keine Korrekturmöglichkeiten vorhanden	Korrekturen können nicht vorgenommen werden	1
Telefon-/E-Mail-Support	Es gibt einen telefonischen Support oder Kontaktmöglichkeiten per Email	6

Tabelle IV.59

Bezüglich welcher Merkmale weisen Verbraucher Ihr Unternehmen am häufigsten auf fehlerhafte Daten hin?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Veraltetes Kartenmaterial	Fehlerhafte Daten oder falsche Datenbewertung zu Geschwindigkeitsüberschreitungen (z. B. aufgrund verzögerter Aktualisierung des Kartenmaterials, temporärer Baustellen)	7
Aufzeichnung von Fahrten	Aufzeichnung von Fahrten wurde nicht durchgeführt oder sind unvollständig (z. B. wegen fehlendem GPS-Signal, Empfangsproblemen)	2
Keine	Es gibt keine Hinweise auf Fehler	2

Tabelle IV.60

Halten Sie es für möglich, dass Ermittlung des Scores mit antidiskriminierungsrechtlichen Gleichbehandlungsgeboten in Konflikt geraten kann? Falls ja, wie begegnen Sie diesem Problem?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Nein	Ein Konflikt mit Gleichbehandlungsgeboten wird nicht für möglich gehalten	10
Keine Verwendung persönlicher Daten	Es werden keine persönlichen Daten/Persönlichkeitsmerkmale verwendet	3
Gesetzliche Vorgaben	Nur Daten, die laut Gesetz (z. B. AGG) zulässig sind, werden verwendet	2

Tabelle IV.61

Unterliegen Ihre Scoring-Verfahren staatlicher Aufsicht? Falls ja, welche Aufsichtsbehörden sind für die Prüfung zuständig? Wie oft und aus welchen Anlässen wird überprüft?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Nein	Es gibt keine staatliche Aufsicht	5
BaFin	Die BaFin hat die Aufsicht	4
Datenschutzbehörde	Eine Landesdatenschutzbehörde hat die Aufsicht	2
Keine Aussage möglich		1

Tabelle IV.62

Welche Vorteile sehen Sie im Einsatz von Telematik-Tarifen für Ihre Branche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Geschäftsmodell & Wettbewerb	Weiterentwicklung des Geschäftsmodells; Wettbewerbsvorteile durch Produktvielfalt	5
Rückgang von Schadensfällen & Erhöhung von Sicherheit	Minimierung oder Verhinderung von Schadensfällen; Verbessertes Fahrverhalten	9
Risikogerechte Prämienkalkulation	Prämienkalkulation basiert auf individuellem Risiko und ist genauer und gerechter	9
Kostensparnis	Die Kosten werden geringer, z. B. durch weniger Unfälle; Kosten für Unternehmen und Gesellschaft	5
Service	Telematik bietet neue Service-Leistungen, wie Notfallhilfe, bessere Kommunikation mit dem Unternehmen	4
Kundenerwartungen & Akzeptanz	Die Kunden erwarten Veränderungen in Zeiten der Digitalisierung	1

Tabelle IV.63

Welche Nachteile sehen Sie im Einsatz von Telematik-Tarifen für Ihre Branche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Keine	Es gibt keine Nachteile	4
Kosten	Die eingesetzte Technik verursacht Kosten	4
Datenmonopol	Datenmonopol bei den Herstellern ist eine Gefahr	2
Fraglicher Nutzen	Es ist unklar ob Telematik bessere Risikokalkulation ermöglicht als bisherige Verfahren; die heutige Kalkulation ist schon sehr risikogerecht	1
Unternehmerisches Risiko	Das Unternehmen trägt das alleinige Risiko bei Einführung von Neuerungen	2

Tabelle IV.64

Wie würden Sie die Akzeptanz Ihrer Kunden einschätzen, für schlechtes Fahrverhalten höhere Prämien in der Kfz-Versicherung zahlen zu müssen?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Kein Malussystem	Versicherer betont, es gibt derzeit kein Malussystem	7
Keine/Niedrige Akzeptanz	Die Akzeptanz eines Malussystems wird als niedrig eingeschätzt	5
Beurteilung nicht möglich	Es fehlen die Erkenntnisse für eine Antwort	2

Tabelle IV.65

Halten Sie die Erweiterung der Datenbasis Ihres Scoring-Verfahrens um Merkmale aus anderen Bereichen (z. B. Kreditwürdigkeit oder Gesundheitsverhalten) für sinnvoll, um die Prognosequalität zu verbessern?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Nein	Eine solche Erweiterung der Datenbasis wird nicht als sinnvoll erachtet und nicht praktiziert	10
Risikobezug	Nur Daten zum Fahrverhalten erscheinen zur Risikokalkulation sinnvoll	4

Tabelle IV.66

Sehen Sie Bedarf, den Gesetzes- und Regulierungsrahmen in Bezug auf Telematik-tarife anzupassen (z. B. im Bereich des Datenschutzrechts)? Falls ja, in welcher Hinsicht?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Nein	Kein Anpassungsbedarf	5
Ja	Anpassungsbedarf wird gesehen	5
DSGVO	DSGV genügt als Regulierungsrahmen	3
Datensouveränität	Es sollte gesetzlich geregelt werden, dass die Daten dem Kunden gehören	4
Datenschutz	Datenschutz sollte klar geregelt sein	1
Diskurs	Gespräche zu Digitalisierung und Verbraucherschutz sollen geführt werden	2
Informationspflichten	Informationspflichten eindeutig regeln	2
Forschung	Es sollte mehr Forschung ermöglicht werden	2
Vermeidung von Überregulierung	Überregulierung sollte vermieden werden	3
Datenmonopole verhindern	Datenmonopole sollten verhindert werden; Zugriff auf Daten soll allen Unternehmen ermöglicht werden	1

Tabelle IV.67

Wie beurteilen Sie den Einfluss von eCall-Systemen auf den Telematik-Bereich in Ihrer Branche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Telematik wird gefördert	ECall-Technik ermöglicht weitere Dienstleistungen wie Telematiktarife; Ecall wird begrüßt	5
Herstellerabhängigkeit & Datenmonopole	Herstellerunabhängige Plattformen stärken Wettbewerb; (Daten-) Monopole (z. B. durch OEMs) müssen verhindert werden	6
Datensouveränität	Kunden sollten Herr ihrer Daten sein	5
Kein Zusammenhang	Ein Zusammenhang zwischen eCall und Telematik wird nicht gesehen	1
Nachfrage	Nachfrage noch gering, kann durch eCall-Gesetz gesteigert werden	1

Tabelle IV.68

Wie schätzen Sie die Zukunft von Scoring-Verfahren in ihrer Branche ein? Gibt es Trends z. B. aus dem Ausland, die Sie für die Weiterentwicklung Ihrer Kfz-Versicherung für relevant halten?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=10)
Prävention	Der Präventionsgedanke bei Versicherung wird stärker werden	3
Langsame Verbreitung	Telematik wird sich in Deutschland nicht schnell verbreiten (wie in anderen Ländern)	3
Schnelle Verbreitung	Telematik wird sich stark verbreiten	2
Neue Zielgruppen	Neue Zielgruppen werden erschlossen	1
Geschäftsmodelle	Neue Geschäftsmodelle und Versicherungsprodukte mit Telematik werden entstehen	4
Keine Trends aus dem Ausland	Keine Trends (auch aus dem Ausland) werden gesehen	3

3. Krankenversicherungen

Tabelle IV.69

Bieten Sie im Rahmen Ihrer Krankenversicherung ein Bonusprogramm (bzw. ein Präventions- oder Vorteilsprogramm) an?	
ANTWORT	ANZAHL (n=62)
Ja	45
Nein	17

Tabelle IV.70

Gibt es eine Zielgruppe, die mit dem Bonusprogramm Ihrer Krankenversicherung besonders angesprochen werden soll? Falls ja, welche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Keine	Keine besondere Zielgruppe wird angesprochen	33
Familien	Das Bonusprogramm richtet sich an Familien	3
Gesundheitsbewusste Kunden	Es werden gezielt gesundheitsbewusste Kunden angesprochen	5
Mindestalter	Kunden ab einem bestimmten Mindestalter werden angesprochen; Kinder bis zu einem bestimmten Alter dürfen nicht teilnehmen	3
Kostenbewusste Kunden	Es werden Kunden angesprochen, die auf Kosten achten	2
Keine Antwort		3

Tabelle IV.71

Wie groß ist der Anteil Ihrer Kunden in der Krankenversicherung, die am Bonusprogramm teilnehmen? Zeichnen sich die teilnehmenden Kunden im Vergleich zu nicht-teilnehmenden Kunden durch besondere Merkmale (z. B. Alter, Geschlecht) aus?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
0–5% der Gesamtversicherten		10
6–10% der Gesamtversicherten		17
11–20% der Gesamtversicherten		13
21–30% der Gesamtversicherten		1
31–40% der Gesamtversicherten		1
Keine besonderen Merkmale	Teilnehmer sind nicht durch besondere Merkmale gekennzeichnet oder es sind keine bekannt	15
eher weiblich	Teilnehmer sind eher weiblich	13
jünger	Die Altersstruktur der Teilnehmer ist jünger; es finden sich weniger ältere Versicherte unter den Teilnehmern	4
älter	Die Altersstruktur der Teilnehmer ist älter; es finden sich weniger jüngere Versicherte unter den Teilnehmern	3
Weniger Arbeitslose		2
Geringeres Ausgabenniveau	Teilnehmer verursachen weniger Kosten	1

Tabelle IV.72

Welche potenziellen Vorteile haben Kunden in der Krankenversicherung, die am Bonusprogramm teilnehmen?	
ANTWORT	ANZAHL (n=45)
Rückzahlung von Versicherungsprämien/-beiträgen	5
Niedrigere Versicherungsprämien/-beiträge im Folgejahr	0
Geldleistungen	42
Sachleistungen	14
Gutscheine/Rabatte bei Partnerunternehmen	6
Gutscheine/Rabatte für medizinische Behandlungen	4
Keine Antwort	3

Tabelle IV.73

Mit welchen Aktivitäten können Kunden Bonuspunkte sammeln?	
ANTWORT	ANZAHL (n=45)
Früherkennungs- bzw. Vorsorgeuntersuchungen	45
Teilnahme an Gesundheitskursen	41
Mitgliedschaft in einem Sportverein, Fitnessstudio oder im Betriebs-/Hochschulsport	41
Teilnahme an Sportveranstaltungen	29
Sportabzeichen	41
Schutzimpfungen	41
Erfassung von Daten mittels Fitness-Apps und/oder Wearables	3
Blutspende	5
Soziales Engagement/gemeinnützige Aktivitäten	5
BMI im Normalbereich	10
Zahnvorsorge/-reinigung	9
Nicht-Raucherstatus	8

Tabelle IV.74

Falls Kunden aufgrund gemeinnützigen Verhaltens (z. B. Blutspende, soziales Engagement) Bonuspunkte sammeln können: Wie begründen Sie die Einbeziehung solcher Aktivitäten in Ihr Bonusprogramm?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Gesundheitlicher Eigennutz	Aktivitäten wie z. B. Blutspende haben auch einen individuellen gesundheitlichen Nutzen	1
Belohnung altruistischen Verhaltens	Soziale Gerechtigkeit und Engagement für die Gesellschaft soll gefördert und belohnt werden	2
Gesundheitlicher Nutzen für Andere	Gemeinnützige Aktivitäten haben präventiven Effekt auf andere	1
Nicht praktiziert	KV bonifiziert kein gemeinnütziges Verhalten	8
Nicht genehmigt	Wurde (von BVA) nicht genehmigt; KV würde aber gerne	9
Keine Antwort		24

Tabelle IV.75

Anhand welcher Kriterien wird entschieden, mit welchen Aktivitäten Kunden Bonuspunkte sammeln können?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Gesundheitseffekt	Aktivität muss einen nachgewiesenen (evidenzbasiert) gesundheitlichen Nutzen haben bzw. gesundheitsbewusstes Verhalten fördern	14
Gesetzliche Vorgaben	z. B. gemäß §65a SGB V	15
BVA	Vorgaben durch das Bundesversicherungsamt	9
GKV-Spitzenverband	Leitfaden Prävention des GKV-Spitzenverbandes	4
Markttrends/Angebote anderer Kassen	Der Markt und Bonusprogramme anderer Kassen werden beobachtet	3
Nachhaltigkeit	Aktivität muss nachhaltig sein	3
Attraktivität für Versicherte	Aktivität muss attraktiv sein	3
Förderung von Angeboten zur Früherkennung und Prävention	Die Inanspruchnahme von Vorsorge- bzw. Früherkennungs- und Präventionsangeboten soll gefördert werden	6
Wirtschaftlicher Nutzen	Das Bonusprogramm muss wirtschaftlich sein	2
Eigenverantwortung	Die Aktivitäten müssen die Eigenverantwortung der Versicherten stärken	2
Qualitätskriterien	Unternehmensinterne Qualitätskriterien	3
Nachweisbarkeit	Die Absolvierung der Maßnahmen muss nachweisbar sein	3
Kundenbefragungen	Die Auswahl erfolgt auf Basis von Erkenntnissen aus Kundenbefragungen	1
Satzung	Auswahl der Aktivitäten ist in der Satzung der Kasse festgeschrieben	4
Bezug zum Trägerunternehmen		1
Keine Antwort		7

Tabelle IV.76

Anhand welcher Kriterien wird jeweils entschieden, wie viele Punkte Kunden mit einer Aktivität sammeln können?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Keine Gewichtung	Die Aktivitäten im Bonusprogramm werden nicht gewichtet, jede Aktivität zählt gleich viel; Es gibt kein Punktesystem	17
Aufwand & Unkompliziertheit	Der Aufwand darf für den Kunden nicht zu hoch sein, der Zugang muss für alle Versicherten unkompliziert sein	8
Wirtschaftlichkeit	Wirtschaftlichkeit der Maßnahmen wird bewertet	6
Gesundheitlicher Nutzen	Der gesundheitliche Nutzen bestimmt die Gewichtung	5
Beliebtheit der Maßnahme	Die Gewichtung richtet sich danach, wie häufig Maßnahmen von Kunden in Anspruch genommen werden	2
Satzung	Die Gewichtung der Maßnahmen ist in der Satzung der Versicherung festgeschrieben	4
Keine Antwort		9

Tabelle IV.77

In welcher Form teilen Sie Kunden ihren aktuellen Stand im Bonusprogramm mit?	
ANTWORT	ANZAHL (n=45)
Punktstand	19
Wahrscheinlichkeitswert	0
Symbole (z. B. Ampel, Farbskalen)	2
Ausformulierte Empfehlungen an den Kunden bzw. den Verbraucher	1
Zielmarken	1
Sonstiges (z. B. Bonushefte)	20
Keine Antwort	6

Tabelle IV.78

Mit welchen in Aussicht gestellten Effekten motivieren Sie Kunden zur Teilnahme am Bonusprogramm?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Vergütung	Die Kunden sparen Geld oder erhalten Boni, Gutscheine etc.; Statusvorteile	33
Unmittelbare & präventive Gesundheitseffekte	Dem Kunden wird ein gesundheitlicher Nutzen in Aussicht gestellt bzw. Fitness, Wohlbefinden; Dem Kunden wird Früherkennung und Vorsorge in Aussicht gestellt	28
Keine Antwort		1

Tabelle IV.79

Mit welcher Maßnahme können Kunden ihren Punktestand am ehesten verbessern?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Sportliche Aktivitäten		6
Ärztliche (Früherkennungs-)Untersuchungen	Früherkennungs- und Vorsorgeuntersuchungen; Absolvierung einer Reihe von ärztlichen Untersuchungen	9
Gesundheitskurse		2
Keine Gewichtung	Versicherer weist darauf hin, dass alle Maßnahmen gleichgewichtet sind	11
Verschiedene Aktivitäten	Es wird nicht eine einzelne Maßnahme sondern eine Reihe von Aktivitäten aufgezählt bzw. auf Auflistungen in vorherigen Fragen verwiesen	8
Keine Antwort		7

Tabelle IV.80

Bieten Sie Ihren Kunden die Möglichkeit, ihren Punktestand im Bonusprogramm digital, z. B. über eine App, einzusehen und absolvierte Aktivitäten zu erfassen?	
ANTWORT	ANZAHL (n=45)
Ja	7
Nein	36
Keine Antwort	2

Tabelle IV.81

Bieten Sie Ihren Kunden die Möglichkeit, ihren Punktestand online zu veröffentlichen?	
ANTWORT	ANZAHL (n=45)
Ja	0
Nein	43
Keine Antwort	2

Tabelle IV.82

Bieten Sie die Möglichkeit, dass Kunden ihre jeweiligen Punktestände in Form eines Rankings miteinander vergleichen können (z. B. über eine App oder ein Onlineportal)?	
ANTWORT	ANZAHL (n=45)
Ja	0
Nein	43
Keine Antwort	2

Tabelle IV.83

Ändert sich der Gesundheitszustand Ihrer Kunden durch die Teilnahme an Ihrem Bonusprogramm? Wie stellen Sie dies fest?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Ja	Veränderung des Gesundheitszustands (nicht der Ausgaben); Kontrollgruppenvergleiche (Teilnehmer vs. Nicht-Teilnehmer) werden als Methode der Evaluierung erwähnt; Gesundheitseffekt wird durch Kundenbefragungen festgestellt	4
Gesundheitseffekt indirekt/abgeleitet feststellbar	Der Effekt wird ohne Begründung bestätigt; Die Methode der Evaluation wird nicht näher erläutert; Der Effekt wird indirekt durch Beweis der Wirtschaftlichkeit oder Senkung der Leistungsausgaben/ Gesundheitskosten gezeigt (Verweis auf regelmäßige Evaluation und die Nachweispflicht gegenüber dem BVA nach §65a Abs 3 AGB V); Teilnehmer nehmen häufiger an Präventionsangeboten teil	35
Nein	Keine Aussagen möglich; Eine konkrete Verbesserung des Gesundheitszustandes der Teilnehmer ist nicht messbar	3
Keine Antwort		3

Tabelle IV.84

Nehmen Sie bei der Entwicklung und Evaluierung Ihres Bonusprogrammes externe wissenschaftliche Beratung in Anspruch? Falls ja, in welcher Form?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=45)
Ja	Verweis auf externe Dienstleister (z. B. „Team Gesundheit“); Beratung in Hinblick auf Berichte fürs BVA	34
Nein	Keine externe wissenschaftliche Beratung	10
Keine Antwort		1

Tabelle IV.85

Welche mittels Fitness-/Gesundheits-Apps erfassten Daten werden für das Bonusprogramm verwendet?	
ANTWORT	ANZAHL (n=3)
Schrittzahl des Versicherten	3
Herzfrequenz	2
Kalorienverbrauch	2

Tabelle IV.86

Arbeiten Sie bei der Erfassung von Daten mittels Fitness-/ Gesundheits-Apps mit Drittanbietern zusammen? Falls ja, welche?	
ANTWORT	ANZAHL (n=3)
Google Fit	2
Apple Health	2
Samsung Health	2
Fitbit	1
Keine Antwort	1

Tabelle IV.87

Bezuschusst Ihre Versicherung die Anschaffung von Geräten, die Aktivitätsdaten erfassen und weitergeben (wie z. B. Fitnessarmbänder)?	
ANTWORT	ANZAHL (n=62)
Ja	11
Nein	41
Keine Antwort	10

Tabelle IV.88

Sollte es aus Sicht Ihrer Krankenversicherung die Möglichkeit geben, bei der Prämien-gestaltung von Krankenversicherungstarifen bzw. bei der Ausgestaltung von Bonusprogrammen neben gesundheitsförderlichen auch gesundheitsschädliche Verhaltensweisen von Versicherten zu berücksichtigen?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
Ja	Sollte möglich sein	5
Erhöhung Eigenverantwortung	Eigenverantwortung der Versicherten wird gestärkt	2
Erhöhung gesundheitlichen Nutzens	Ein besserer (gesundheitlicher) Effekt würde erzielt	1
Nein	Sollte nicht möglich sein; Verhalten zu bestrafen ist nicht Aufgabe einer KV	55
Gefährdung des Solidaritätsprinzips/ Diskriminierung	Das Solidaritätsprinzip würde gefährdet; Es besteht eine Gefahr der Diskriminierung oder Ausgrenzung	17
Datenschutzbedenken	Es gibt datenschutzrechtliche Bedenken	10
Definitionsproblem	Es ist schwierig „gesundheitsschädlich“ zu definieren	5
Überprüfbarkeit	Es ist schwierig gesundheitsschädliche Aktivitäten zu überprüfen/kontrollieren	2
Gesetzlich nicht möglich/ungeklärte Rechtsfragen	Der gesetzliche Rahmen erlaubt dies nicht	2
Demotivation/fehlende Akzeptanz	Versicherte würde demotiviert oder hätten negative Erfahrungen; Die Versicherten würden dies nicht akzeptieren, sie würden den Autonomieverlust nicht hinnehmen; Die Wirksamkeit wird angezweifelt	3
Überwachung	Es wird von einer „Überwachung“ der Lebensführung der Versicherten gesprochen	9
Keine Antwort		3

Tabelle IV.89

Sehen Sie Bedarf, den Gesetzes- und Regulierungsrahmen in Bezug auf Bonusprogramme oder verhaltensbasierte Tarife anzupassen (z. B. im Bereich des Datenschutz- oder Gesundheitsrechts)? Falls ja, in welcher Hinsicht?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
Nein	Kein Anpassungsbedarf	41
Ja	Anpassungsbedarf wird gesehen	17
Mehr Gestaltungsspielraum bei Ausgestaltung des Bonusprogramms	Gewünscht wird mehr Gestaltungsspielraum für Auswahl der Aktivitäten im Bonusprogramm, z. B. in Hinblick auf gemeinnütziges Verhalten oder Fitness-Tracker; unterschiedliche Aufsichtspraxis zw. Ländern und BVA wird erwähnt ; weniger Einfluss der Aufsichtsbehörde/Deregulierung gewünscht	13
Kalkulationsvorgaben	Änderung der Kalkulationsvorgaben gewünscht	2
Lockerung Datenschutz	Datenschutz sollte gelockert werden	1
Wunsch nach Verknüpfung von Daten	Die Verknüpfung von Datenpools sollte ermöglicht werden	1
Abschaffung von Bonusprogrammen	Bonusprogramme sollten abgeschafft werden	1
Schutz der Solidargemeinschaft	Die Solidargemeinschaft muss stärker geschützt werden; zu hohe Boni verhindern	1
Datensouveränität	Kunde sollte selbst über seine Daten entscheiden dürfen	3
Mehr Regulierung	Mehr Regulierung soll z. B. Missbrauch verhindern	2
Ausweitung datenbasierter Angebote	Neu datenbasierte Angebote, z. B. Beratungsangebote, sollten ermöglicht werden	1
Keine Antwort		4

Tabelle IV.90

Sollten elektronische Patientenakten und die damit potenziell zur Verfügung stehenden Gesundheitsdaten von Versicherten für den Bereich Bonusprogramme und /oder verhaltensbasierte Tarife in der Krankenversicherung nutzbar sein?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
Ja	Gesundheitsdaten aus der elektronischen Patientenakte sollten nutzbar sein	22
Datensouveränität	Betonung, dass dies nur auf Wunsch/nach Zustimmung der Versicherten geschehen sollte	13
Service und Information	KV versprechen sich durch die Nutzung von Daten aus der elektronischen Patientenakten besseren Service bzw. bessere Informationen für ihre Kunden	1
Bessere Versorgung	KV versprechen sich durch die Nutzung von Daten aus der elektronischen Patientenakten bessere Versorgung	4
Weniger Verwaltungsaufwand	KV versprechen sich durch die Nutzung von Daten aus der elektronischen Patientenakten weniger Verwaltungsaufwand	1
Höhere Akzeptanz	Die Akzeptanz der elektronische Patientenakte durch die Versicherten würd erhöht	1
Kundenwunsch	Es wäre im Interesse der Kunden	1
Nein	Gesundheitsdaten aus der elektronischen Patientenakte sollten nicht nutzbar sein	32
Benachteiligung verhindern	Kranke Versicherte dürfen nicht benachteiligt werden	1
Datenschutzrechtliche Bedenken	Bedenken in Bezug auf Datenschutz („gläserner Patient“)	4
Nur anonymisiert, für statistische Zwecke	Keine persönlichen Gesundheitsdaten, nur die anonymisierte Datenbasis sollte nutzbar sein, um Angebote weiterzuentwickeln	1
Keine Antwort		1

Tabelle IV.91

Welche Vorteile sehen Sie im potenziellen Einsatz von verhaltensbasierten Tarifen für Ihre Branche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
keine	Verhaltensbasierte Tarife werden von der Versicherung abgelehnt; es gibt keine Vorteile	28
Erhöhung von Gesundheitsverhalten und Gesundheitsbewusstsein	Anreiz zu gesundheitsförderlichem Verhalten/ Steigerung Gesundheitsbewusstsein	17
Kostensparnis	Senkung der Leistungsausgaben/Krankheitskosten	7
Geschäftsmodell & Wettbewerb	Vorteile im Wettbewerb	2
Kundenbindung		2
Besseres Leistungsangebot	Mehr Leistungen/bessere Portfolios werden geboten; Versicherer werden vom Kostenerstatter zum Gesundheitsmanager	1
Boni für gesundheitsbewusste Kunden	Die Versicherten können Boni erhalten	1
Keine Antwort		11

Tabelle IV.92

Welche Nachteile sehen Sie im potenziellen Einsatz von verhaltensbasierten Tarifen für Ihre Branche?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
Solidarprinzip/Diskriminierung	Verletzung/Gefährdung des Solidaritätsprinzips bestimmte, z. B. (genetisch bedingt) kranke Kunden oder Kunden in bestimmten Lebenssituationen werden benachteiligt und stigmatisiert	30
Datenschutz	Datenschutzrechtliche Bedenken	5
Risikoselektion	Auswirkung auf Risikoselektion/Risikoentmischung	12
Kosten/Verwaltungsaufwand	Höherer Verwaltungsaufwand/höhere Kosten Verhaltensweisen sind schwer nachzuweisen	15
Marketinginstrument	Missbrauch als Marketinginstrument	2
Vergleichsmöglichkeiten & Kassenwechsel	Erschwerter Vergleich und Kassenwechsel für die Versicherten	2
Monopolbildung	Kleinere Kassen können nicht mithalten, große setzen sich durch	2
Mitnahmeeffekte		1
Fehlsteuerung	Vermeidung von Arztbesuchen für Prämien führt zu Risiken	1
Negativerlebnisse	Manche Kunden könnten durch verhaltensbasierte Tarifierung negative Erlebnisse haben	2
keine	Es gibt keine Nachteile	5
Skandalanfälligkeit	z. B. Datenskandale; der Versicherte verliert so das Vertrauen	2
Keine Antwort		12

Tabelle IV.93

Wie schätzen Sie die Zukunft von verhaltensbasierten Tarifen in ihrer Branche ein? Gibt es Trends (z. B. aus dem Ausland), die Sie für die Weiterentwicklung Ihrer Krankenversicherung für relevant halten?		
ANTWORT	BESCHREIBUNG	ANZAHL (n=62)
Zukunftseinschätzungen		
Tendenziell positive Markttendenz	Verhaltensbasierte Tarife werden in den Gesundheitsmarkt Einzug halten, sich weiter verbreiten; Es gibt bereits erste Formen verhaltensbasierter Tarifierung, z. B. in andere Versicherungsarten oder durch Wahltarife; Versicherung bewertet verhaltensbasierte Tarife positiv und würde deren Verbreitung begrüßen; Verhaltensbasierte Tarife kann es wenn dann nur in der PKV geben	9
Keine Markttendenz	Verhaltensbasierte Tarife wird es nicht geben; Verhaltensbasierte Tarife nicht möglich, auch weil die gesetzliche Regulierung ist in Deutschland sehr stark ist	19
Ablehnung	Versicherung lehnt verhaltensbasierte Tarife ab; Entsolidarisierung/ Benachteiligungen werden dadurch befürchtet	20
Beobachtete/ interessante Trends		
Kundenerwartungen	Das Gesundheitsbewusstsein der Versicherten nimmt zu; Der Umgang mit Gesundheitsdaten wird sich ändern, Skepsis abnehmen; Kunden werden verhaltensbasierte Versicherungen und personalisierte Angebote erwarten	2
Technischer Fortschritt (z. B. Telemedizin)	Telemedizin wird als wichtiger Trend gesehen; Die Digitalisierung und damit verbundene Datennutzungsmöglichkeiten werden als wichtiger Trend gesehen	4
Geschäftsmodelle & Wettbewerb	Neue Geschäftsmodelle, neue Versicherungsprodukte werden entstehen und verhaltensbasierte Tarife zum Unterscheidungsmerkmal werden	2
Ausbau Prävention	Der Präventionsgedanke wird stärker werden; KV nicht nur als Kostenerstatter, sondern als Gesundheitsmanager	3
Keine Trends	Es sind keine Trends erkennbar	4
Keine Antwort		11

V. Detaillierte Darstellung der Methode und Ergebnisse (Vorstudie und Bevölkerungsbefragung)

1. Dokumentation der Vorstudie 2017

Im November 2017 wurde eine Vorstudie in Berlin durchgeführt, um Teile des für die repräsentative Befragung entwickelten Fragebogens zu testen (Rebitschek et al. 2018). Die Studie wurde in einem Berliner Kinosaal mit 91 Teilnehmern (52 % Frauen), im Alter zwischen 17 und 75 Jahren durchgeführt (Rebitschek et al. 2018). Teilnehmer wurden aus einem Probandenpanel des Max-Planck-Instituts für Bildungsforschung (MPIB) sowie über die Geschäftsstelle des Sachverständigenrats für Verbraucherfragen, über das DIW Berlin, das SIBIS Institut für Sozialforschung, die Humboldt Law Clinic, das Alexander von Humboldt Institut für Internet und Gesellschaft sowie in der Oberstufe eines Gymnasiums (Katholische Schule Salvator) mit einer themenspezifischen Einladung rekrutiert. 68 % der Teilnehmer hatten mindestens ein Hochschulstudium abgeschlossen, 25 % hatten eine Hochschulzugangsberechtigung (Fachhochschulabschluss oder Abitur), nur 7 % hatten keine (Real-, Haupt- oder Volksschulabschluss).

Eingesetzt wurden Paper-and-Pencil-Fragebögen (Rebitschek et al. 2018) mit Fragen u. a. zu Wissen (Merkmale, die in etablierte Bonitätsscorings eingehen) und der Akzeptanz von Scoring-Szenarien zu kombinierter Kfz-Krankenversicherung (kombiniertes Versicherungsscoring), Telematikmeldung an die Polizei (Strafzettel-Scoring) sowie einem umfassenden Bürgerscoring in verschiedenen Lebensbereichen.

Als Intervention, die das Erleben eines fiktiven Social Credit Systems darstellt, wurde die Episode „Nosedive“ der Netflix-TV-Serie Black Mirror gezeigt. In diesem Film werden von den Menschen ständig jede Dienstleistung sowie nahezu jede soziale Interaktion bewertet, jeder hat einen persönlichen Score und kann zugleich den je-

des anderen in Echtzeit erfahren. Denken und Handeln drehen sich nahezu ausschließlich um den Score. Die junge Protagonistin des Films braucht einen höheren Score, um eine bessere Wohnung mieten zu können. Sie hofft durch den hoch gerankten Freundeskreis einer alten Freundin ihren eigenen Score zu steigern, denn der Umgang mit Personen, die einen höheren Score haben, wirkt sich positiv auf den eigenen Score aus und ein Absinken gilt es um jeden Preis und mit allen Mitteln zu vermeiden, da sozialer Abstieg und Ausgrenzung drohen.

Im Anschluss an die Gruppenvorführung im Kinosaal wurden die Einstellungsfragen ein zweites Mal vorgelegt und mit offenen Fragen nach Hintergründen gefragt. Die Akzeptanz von neuartigen Scoring-Verfahren, welche Merkmale aus verschiedenen Scoringbereichen kombinieren, war vor und nach der Intervention durch die Filmvorführung für alle Szenarien auf eine Minderheit beschränkt (Tabelle V.1), wobei das Szenario „Bürgerscoring“ von 92 % (vorher) bzw. 96 % (nachher) der Teilnehmer abgelehnt wurde, das „Strafzettel-Scoring“ hingegen von 31 % bzw. 25 % befürwortet wurde. Knapp 28 % der Teilnehmer befürworteten mindestens eines der drei Szenarien. Dieser Anteil lag vor der Filmvorführung noch etwas höher; dies war zu erwarten, da die meisten TeilnehmerInnen vor der Filmvorführung über ein solches Szenario nicht vertieft nachgedacht haben werden.

Die Intervention, die den TeilnehmerInnen zusätzliches Wissen brachte, hat die erwarteten Ergebnisse gezeigt, nämlich mehr Skepsis bei mehr Wissen. Freilich sind die Effekte so klein, dass man davon ausgehen kann, dass die Ergebnisse einer Repräsentativ-Befragung ohne Intervention aussagekräftiger sind.

Tabelle V.1

Anteil der Personen, die Scoring-Szenarien vor bzw. nach der Medienpräsentation befürworten (in %).			
SZENARIO	ANTEIL DER PERSONEN, DIE	PRÄ	POST
Kombiniertes Versicherungsscoring	nutzen würden	15	11
Strafzettel-Scoring	einführen würden	31	25
Bürgerscoring	einführen würden	8	4
Selektives Scoring	mindestens eins der Szenarien befürworten	39	28

2. Dokumentation der Repräsentativbefragung

2.1 Methodik

Ziel der repräsentativen Bevölkerungsbefragung war es, einen Einblick bezüglich der Einstellungen zu und der Akzeptanz von etablierten und (potenziellen) neuartigen Scoring-Verfahren in den drei Bereichen Bonität, Kfz-Versicherung und Gesundheit der in Deutschland lebenden Bevölkerung zu erlangen. Zu diesem Zweck wurde eine Telefonbefragung (so genanntes computer-assisted telephone interview, CATI) im Zeitraum von Anfang Februar bis Ende April 2018 mit einer durchschnittlichen Dauer von 22,5 min durchgeführt. An der repräsentativen Bevölkerungsbefragung nahmen n = 2.215 Personen im Alter zwischen 16 und 94 Jahren teil (1.123 Frauen, 1.092 Männer). Die Struktur und der Inhalt des standardisierten, strukturierten Fragebogens, anhand dessen die Telefonbefragung durchgeführt wurde, ist nachfolgend beschrieben.

2.1.1 Aufbau und Inhalt der Telefonbefragung

Im Anschluss an einen Einstiegsblock, in dem kurz das Ziel der Befragung erläutert, die Teilnahmebereitschaft abgeklärt sowie Einstiegsfragen u. a. zu den Themen Fortbewegung und Gesundheit (siehe 2.1.5) gestellt wurden, erfolgten Fragen zu den drei Themenkomplexen Bonität (siehe 2.1.2), Kfz-Versicherung und Gesundheit (siehe 2.1.3). Im Kontext von potenziellem neuartigem Scoring wurde zudem die Akzeptanz eines umfassenden sozialen Bürger-Scorings (siehe 2.1.4) erfasst. Abschließend wurden soziodemographische Angaben erhoben (siehe 2.1.6).

2.1.2 Wissen zu etabliertem Bonitäts-Scoring

Obgleich das Wissen über Scoring mehr als die Merkmale, die dafür genutzt werden, umfasst, wurde aus Praktikabilitätsgründen scoringbezogenes Wissen ausschließlich über die Kenntnis der Merkmale, die zur Beurteilung der Bonität von Privatpersonen von in Deutschland etablierten Auskunfteien herangezogen werden, erfasst.

Zunächst sollten die Befragten für die Merkmale (1) *eigenes Zahlungsverhalten*, (2) *Alter*, (3) *laufender Kredit*, (4) *Vorhandensein von Vermögen*, (5) *Verhalten in sozialen Netzwerken*, (6) *Beruf* und (7) *Nationalität* jeweils entscheiden, ob dieses Merkmal bei der Berechnung des Bonitäts-Scores Berücksichtigung findet oder nicht. Um einem Reihenfolgeeffekt entgegenzuwirken, wurde die Abfolge der Präsentation über alle Teilnehmer randomisiert. Darüber hinaus wurde die Zustimmung bzw. Ablehnung bezüglich einer Veröffentlichung des eigenen Bonitäts-Scores (im Internet) bzw. des Bonitäts-Scores aller Personen in Deutschland erfragt. Der Themenkomplex Bonität endete mit der Erfragung, ob der eigene Bonitäts-Score in den letzten fünf Jahren erfahren und ob dieser für korrekt gehalten wurde.

2.1.3 Akzeptanz potenzieller neuartiger Scorings

Die Einstellung bzw. Akzeptanz bezüglich (potenzieller) neuartiger Scoring-Verfahren wurde für die Bereiche Gesundheit und Kfz-Versicherung erfasst. Hierbei erhielt die Hälfte (n = 1.111) der Befragten Fragen zum Thema Gesundheit (Split Gesundheit) und die andere Hälfte (n = 1.104) zum Thema Kfz-Versicherung (Split Kfz). Zunächst ging es darum, in Erfahrung zu bringen, inwiefern bestimmte verhaltens- bzw. situationsabhängige Merkmale bei der Ausgestaltung des Versicherungstarifs nach Meinung der in Deutschland lebenden Bevölkerung eine Rolle spielen sollten. Dafür erhielt je die Hälfte der Befragten im Split Gesundheit (n = 570) und je die Hälfte im Split Kfz-Versicherung (n = 536) Fragen mit verhaltens- bzw. situationsabhängiger positiver Konsequenz im Sinne einer Ersparnis bei einem Versicherungstarif (Split Bonus), während die andere Hälfte im Split Gesundheit (n = 541) und im Split Kfz-Versicherung (n = 568) die Fragen mit negativer Konsequenz im Sinne einer Erhöhung bei einem Versicherungstarif präsentiert bekam (Split Malus). Mit Hilfe der Antwortkategorien, die die Abstufungen „voll und ganz gerechtfertigt“, „eher gerechtfertigt“, „eher nicht gerechtfertigt“ und „überhaupt nicht gerechtfertigt“ umfassten, sollte in den Splits Gesundheit bzw. Kfz-Versicherung angegeben werden, für wie gerechtfertigt die Berücksichtigung des jeweiligen Merkmals für die Ausgestaltung des Versicherungstarifs gehalten wird. Um einem Reihenfolgeeffekt entgegenzuwirken, wurden die Merkmale über alle Befragten randomisiert. Die Auswahl von Merkmalen, die für die Ausgestaltung eines Versicherungstarifs herangezogen

werden können, orientierte sich daran, welche Merkmale in verhaltensbasierte Kfz-Versicherungstarifen bereits gegenwärtig einfließen können (z. B. Brems- und Beschleunigungsverhalten, Stadt- vs. Landfahrt, Tag- vs. Nachtfahrt, Geschwindigkeit) bzw. welche Merkmale bei Gesundheits-Apps und Fitnesstrackern und auch (teils) bei Bonusprogrammen von gesetzlichen Krankenversicherungen (z. B. Schrittzahl (AOK Nordost, Techniker Krankenkasse), Krebsfrüherkennungsuntersuchungen) berücksichtigt werden.

Kfz

Für den Split Kfz-Versicherung sollte jeweils für die Merkmale (1) *Geschwindigkeit*, (2) *Nachrichtenversendung bzw. Lesen von Nachrichten auf dem Handy während der Fahrt*, (3) *Tag- vs. Nachtfahrt*, (4) *Stadt- vs. Landfahrt* und (5) *Beschleunigungs- und Bremsverhalten* angegeben werden, für wie gerechtfertigt die Befragten die Berücksichtigung dieser Merkmale für die Höhe des Kfz-Versicherungstarifs halten. Beim Split Bonus, den jeweils die Hälfte der Befragten im Split Kfz erhielt, wurde für bestimmtes Fahrverhalten bzw. Umstände der Fahrt eine positive Konsequenz im Sinne einer Verringerung des Tarifs der Kfz-Versicherung in Aussicht gestellt. Im Einzelnen sollte jeweils in Erfahrung gebracht werden, wie die Befragten es bewerteten, dass die Einhaltung der zulässigen Höchstgeschwindigkeit, kein Lesen oder Schreiben von Nachrichten auf dem Handy, Fahrten bei Tag und auf dem Land sowie ein vorsichtiges Beschleunigen und Bremsen mit einer Preisersparnis assoziiert sind. Umgekehrt sollte beim Split Malus, den die andere Hälfte der Befragten im Split Kfz präsentiert bekam, jeweils bewertet werden, wie gerechtfertigt es von den Befragten empfunden wird, wenn das Überschreiten der zugelassenen Höchstgeschwindigkeit, das Lesen und Schreiben von Nachrichten auf dem Handy, Fahrten bei Nacht und in der Stadt sowie unvorsichtiges Beschleunigen und Bremsen zu einer Preiserhöhung der Kfz-Versicherung führt. Im Anschluss an den Bonus- und Malus-Split erhielten alle Personen im Split Kfz die nachfolgend beschriebenen Fragen.

Anknüpfend an die Fragen, bei denen die Berücksichtigung von verhaltens- bzw. situationsabhängigen Merkmalen für den Kfz-Tarif bewertet werden sollten, wurden alle Befragten im Split Kfz danach gefragt, ob sie es sich vorstellen könnten, einen derartigen Kfz-Versicherungs-

tarif selbst zu nutzen und wurden zudem gebeten, ihre Entscheidung kurz zu begründen. Darüber hinaus wurde erfragt wie häufig die Befragten bei Nutzung eines solchen verhaltensbasierten Tarifs Rückmeldung zu ihrem Fahrverhalten haben wollen würden, wobei zwischen den Antworten „täglich“, „wöchentlich“, „monatlich“, „bei starken Änderungen meines Fahrverhaltens“ und „ich würde keine Rückmeldung dazu haben wollen“ gewählt werden konnte. Zudem wurde nach der Zustimmung bzw. Ablehnung in Bezug auf die Einführung eines persönlichen Punktwerts für das Fahrverhalten und dessen Veröffentlichung im Internet gefragt.

Um einem bereits im Ausland zu beobachtenden Trend Rechnung zu tragen (z. B. Social Credit Score in China), wurde im Weiteren nach der Akzeptanz bezüglich der Verknüpfung von Scores aus verschiedenen Bereichen gefragt. Hierbei sollten die Befragten angeben, ob sie einen Kombinationstarif nutzen würden, der zu einer Kostenersparnis bei der Kfz-Versicherung führt, sofern sie dem Unternehmen auch Daten über das Gesundheitsverhalten zur Verfügung stellten. Mit Hilfe der Antworten „ja, mit Sicherheit“, „ja, möglicherweise“, „nein, eher nicht“ und „nein, sicher nicht“ sollte die Bereitschaft bzw. Ablehnung bezüglich der Nutzung eines Kombinationstarifs signalisiert werden und diese zudem kurz begründet werden. Der Fragenkomplex zum Thema Kfz endete mit der Präsentation eines fiktiven Szenarios, bei dem eine allgemeine Aufzeichnungspflicht der Fahrgeschwindigkeit in allen Autos einschließlich einer obligatorischen Übermittlung von Geschwindigkeitsübertretungen direkt an die Polizei postuliert wurde. Zunächst wurde die Zustimmung bzw. Ablehnung bezüglich der Einführung eines solchen Systems erfragt. Im Anschluss daran sollten die potenziellen Konsequenzen eingeschätzt werden. Dabei sollte unter Zuhilfenahme der Antwortvorgaben „stimme voll und ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ angegeben werden, ob durch so ein System (1) die Sicherheit auf den Straßen erhöht werden könnte, (2) die Gerechtigkeit erhöht werden könnte und (3) die persönliche Freiheit eingeschränkt werden würde. Die Reihenfolge der Präsentation der drei potenziellen Konsequenzen eines solchen Systems wurde über alle Befragten des Kfz-Splits randomisiert.

Gesundheit

Für den Split Gesundheit sollte jeweils bewertet werden, wie gerechtfertigt es die Befragten finden, wenn die Höhe des Krankenversicherungstarifs vom eigenen gesundheitsbezogenen Verhalten abhängt. Dafür sollte jeweils für die Merkmale (1) *Schrittzahl pro Tag*, (2) *Schlafdauer*, (3) *Alkoholkonsum*, (4) *Körpergewicht*, (5) *Teilnahme an Krebsfrüherkennungsuntersuchungen* und (6) *Zigarettenkonsum* angegeben werden, für wie gerechtfertigt die Befragten die Berücksichtigung des Merkmals für die Höhe des Krankenversicherungstarifs halten. Hierbei wurde beim Split Bonus jeweils als positive Konsequenz in Aussicht gestellt, weniger für den Krankenversicherungstarif zu zahlen, wenn ein bestimmtes gesundheitsbezogenes Verhalten wie (1) eine vorgegebene Schrittzahl pro Tag zu gehen, (2) eine bestimmte Schlafdauer pro Nacht einzuhalten, (3) nur geringe Mengen an Alkohol zu trinken, (4) normalgewichtigt zu sein, (5) Teilnahme an Krebsfrüherkennungsuntersuchungen und (6) Nichtraucherstatus nachgewiesen wird. Umgekehrt wurde für den Split Malus als negative Konsequenz bei Nichteinhaltung der gesundheitsbezogenen Verhaltensweisen in Aussicht gestellt, mehr für die Krankenversicherung zu zahlen. Im Anschluss an die Fragen im Bonus- und Malus Split, in denen jeweils nach der Akzeptanz bezüglich der Berücksichtigung von verhaltensabhängigen Merkmalen für den Krankenversicherungstarif gefragt wurden, erhielten alle Befragten die nachfolgend beschriebenen Fragen im Split Gesundheit.

Anknüpfend an die Fragen, bei denen die Berücksichtigung von verhaltensabhängigen Merkmalen für den Krankenversicherungstarif bewertet werden sollten, wurden alle Befragten im Split Gesundheit danach gefragt, ob sie sich die Nutzung so eines Krankenversicherungstarifs vorstellen könnten und wurden zudem gebeten, ihre Entscheidung kurz zu begründen. In diesem Zusammenhang wurde weiter erfragt, wie häufig die Befragten bei Nutzung eines solchen Tarifs Rückmeldung zu ihrem Gesundheitsverhalten haben wollen würden. Hierfür sollte zwischen den Antworten „täglich“, „wöchentlich“, „monatlich“, „bei starken Änderungen meines Gesundheitsverhaltens“ und „ich würde keine Rückmeldung dazu haben wollen“ gewählt werden. Darüber hinaus wurde nach der Zustimmung bzw. Ablehnung in Bezug auf die Einführung eines persönlichen Punkt-

werts für das Gesundheitsverhalten und dessen Veröffentlichung im Internet gefragt.

Auch im Split Gesundheit wurde die Akzeptanz der Verknüpfung von Scores aus unterschiedlichen Bereichen in Erfahrung gebracht. Unter Verwendung der Antwortkategorien „ja, mit Sicherheit“, „ja, möglicherweise“, „nein, eher nicht“ und „nein, sicher nicht“ sollten die Befragten angeben, ob sie sich die Nutzung eines Kombinationstarifs vorstellen könnten, der zu einer Kostenersparnis bei der Krankenversicherung führt, sofern dem Unternehmen auch Daten über das Fahrverhalten zur Verfügung gestellt werden würden. Diese Entscheidung sollte zudem kurz begründet werden. Der Fragenkomplex zum Thema Gesundheit endete mit der Präsentation eines fiktiven Szenarios, bei dem eine allgemeine Aufzeichnungspflicht der Schrittzahl pro Tag postuliert wurde. Bei zu wenigen Schritten würde eine stärkere Beteiligung an Arztkosten drohen. Zunächst wurde erfragt, ob die Befragten für oder gegen die Einführung eines solchen Systems wären. Im Anschluss daran sollten die potenziellen Implikationen eines solchen Systems eingeschätzt werden. Dabei sollte unter Zuhilfenahme der Antwortvorgaben „stimme voll und ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ angegeben werden, ob durch so ein System (1) die Gesundheit der Bevölkerung verbessert werden könnte, (2) die Gerechtigkeit erhöht werden könnte und (3) die persönliche Freiheit eingeschränkt werden würde. Die Reihenfolge der Präsentation der drei potenziellen Konsequenzen eines solchen Systems wurde über alle Befragten des Gesundheits-Splits randomisiert.

2.1.4 Soziales Bürgerscoring

In Anlehnung an das chinesische soziale Kreditsystem wurde ein fiktives Szenario präsentiert, bei dem alle in Deutschland lebenden Bürger einen sozialen Punktwert erhalten, dessen Höhe vom Verhalten (z. B. Fahrverhalten, Gesundheitsverhalten, Zahlungsverhalten, Sozialverhalten) abhängig ist. Bei allen Teilnehmern wurde zunächst erfragt, ob sie die Einführung eines solchen Systems befürworten oder ablehnen würden. Daran anknüpfend sollte das Ausmaß der Zustimmung bzw. Ablehnung zu den möglichen Implikationen eines sozialen Punktwertes, nämlich (1) Erhöhung der Sicherheit, (2) Erhöhung der Gerechtigkeit und (3) Einschränkung der

persönlichen Freiheit unter Verwendung der Antwortkategorien „stimme voll und ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ indiziert werden.

2.1.5 Einstellungen

Um zu erfassen, inwiefern die individuelle Betroffenheit sowie Verhaltensweisen und bestimmte Einstellungen im Zusammenhang mit der Einstellung und Akzeptanz zu etabliertem und potenziell neuartigem Scoring stehen, erhielten alle Befragten zum Einstieg in das Telefoninterview Fragen zu den Themen Fortbewegung und Gesundheit inklusive der Erfahrung mit der Selbstvermessung von gesundheitsbezogenem Verhalten. Zudem wurde die Risikobereitschaft, Kontrollüberzeugungen, sportliche Betätigung und Einstellungen zu den Themen Technik, Datenschutz, Preisbewusstsein bei Produkten und Dienstleistungen und Solidarität exploriert.

2.1.6 Soziodemographie

Zum Abschluss der Telefonbefragung wurden soziodemographische Informationen erfasst. Diese umfassten Geschlecht, Alter, Staatsbürgerschaft, Schul- bzw. Hochschulabschluss, Erwerbstätigkeit, Beziehungsstatus, Kfz- und Krankenversicherung, Fitness, Anzahl der im Haushalt lebenden Personen, Bundesland sowie Einwohnerzahl des Wohnorts.

2.1.7 Codierung offener Antworten

Die offenen Fragen wurden innerhalb des Kfz- und Gesundheits-Splits nur zufällig jedem dritten Befragten gestellt. Zur Auswertung der Antworten auf die offenen Fragen, ob sich die Befragten die Nutzung eines telematikabasierten Kfz- bzw. Gesundheitstarifs vorstellen können sowie, ob sie einen Kombinationstarif nutzen würden, wurde ein Kategoriensystem erstellt. Für das Vorgehen zur Entwicklung des Kategoriensystems siehe infas (2018). Nach Erstellung des Kategoriensystems wurden die offenen Antworten von zwei unabhängigen Beurteilern den einzelnen Kategorien zugeordnet. In Fällen nicht zufriedenstellender Übereinstimmung zwischen den beiden Beurteilern, wurde darüber hinaus ein dritter unabhängiger Beurteiler hinzugezogen.

2.2 Dokumentation der Ergebnisse

2.2.1 Wissen zu etabliertem Bonitäts-Scoring

Den Befragten wurden in randomisierter Reihenfolge Merkmale genannt, die in der Regel von etablierten Auskunfteien zur Beurteilung der Bonität von Privatpersonen herangezogen werden und Merkmale, die dafür nicht verwendet werden. Bei jedem der genannten Merkmale sollte jeweils von den Befragten angegeben werden, ob das Merkmal zur Beurteilung der Bonität Verwendung findet oder nicht. Aufgrund von Rundungen addieren sich die Ergebnisse in einigen Fällen nicht exakt immer zu 100 % auf.

Eigenes Zahlungsverhalten in der Vergangenheit

Die überwiegende Mehrheit der Befragten bejahte die Frage (78%), ob das *eigene Zahlungsverhalten* in der Vergangenheit für die Beurteilung der Bonität herangezogen wird und beantwortete die Frage somit richtig (16% verneinten die Frage, 6% wählten die Kategorie „weiß nicht“ und 0,2% verweigerten die Antwort).

Alter

Ob das Alter einer Person für die Bonitäts-Beurteilung eine Rolle spielt, bejahte die überwiegende Mehrheit korrekterweise (71%), knapp ein Viertel (23%) verneinte dieses und 7% gaben an, dieses nicht zu wissen (0,3% verweigerten die Antwort).

Kredit

Dass ein laufender Kredit Berücksichtigung bei der Beurteilung der Bonität findet, hat die überwiegende Mehrheit mit „ja“ (82%) richtig beantwortet. 12% verneinten diese Frage und 6% wählten die Antwort „weiß nicht“ (0,4% der Befragten verweigerten die Antwort).

Privates Vermögen

Knapp zwei Drittel (64%) der Befragten bejahten fälschlicherweise, dass das private Vermögen für die Beurteilung der Bonität berücksichtigt wird, gut ein Viertel (27%) verneinte dieses korrekterweise und 8% gaben an, dass sie das nicht wüssten (0,4% der Befragten verweigerten die Antwort).



Abbildung V.1

Dargestellt sind die Antworten der Befragten (n = 2.215) auf die Frage, ob das genannte Merkmal von etablierten Auskunfteien für die Beurteilung der Bonität genutzt wird oder nicht. Bei den dargestellten Merkmalen handelt es sich ausschließlich um Merkmale, die u. a. von Auskunfteien für die Bonitäts-Beurteilung genutzt werden. Korrekt ist also stets die Antwort „ja“.

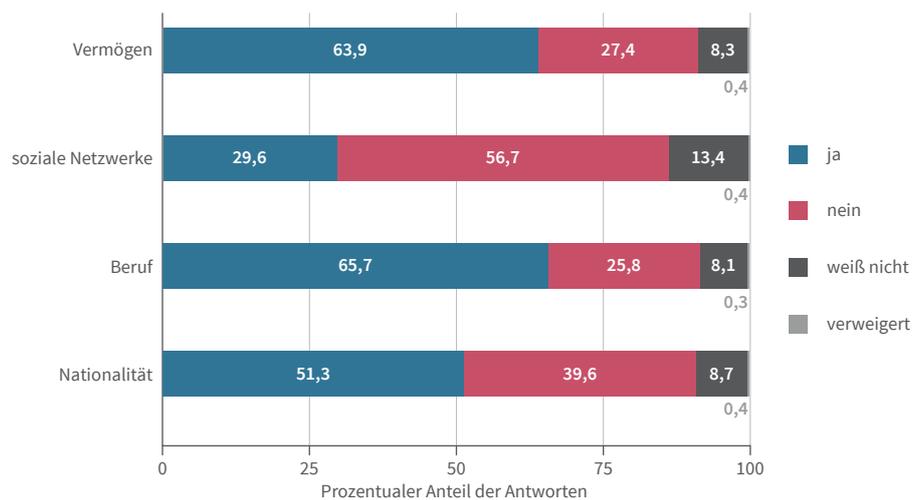


Abbildung V.2

Dargestellt sind die Antworten der Befragten (n = 2.215) auf die Frage, ob das genannte Merkmal von etablierten Auskunfteien für die Beurteilung der Bonität genutzt wird oder nicht. Bei den dargestellten Merkmalen handelt es sich ausschließlich um Merkmale, die nicht von etablierten Auskunfteien für die Bonitäts-Beurteilung genutzt werden. Korrekt ist also stets die Antwort „nein“.

Verhalten in sozialen Netzwerken

Ob das Verhalten in sozialen Netzwerken eine Rolle bei der Bonitäts-Beurteilung spielt, verneinte über die Hälfte der Befragten korrekterweise (57%), knapp ein Drittel bejahte dieses (30%) fälschlicherweise und 13% gaben an, dieses nicht zu wissen (0,4% der Befragten verweigerten die Antwort).

Beruf

Zwei Drittel der Befragten waren fälschlicherweise der Meinung, dass der Beruf eine Rolle bei der Beurteilung der Bonität spielen würde (66%), etwa ein Viertel (26%) war korrekterweise der Meinung, dass der Beruf dafür keine Rolle spielen würde und 8% gaben an, das nicht zu wissen (0,3% verweigerten die Antwort).

Nationalität

Etwa die Hälfte der Befragten (51%) gab fälschlicherweise an, dass die Nationalität bei der Beurteilung der Bonität berücksichtigt würde. 40% waren korrekterweise der Meinung, dass dieses nicht der Fall ist und 9% gaben an, dass sie das nicht wüssten.

Die Antworten der Befragten sind in Abbildung V.1 (Merkmal wird zur Beurteilung der Bonität von etablierten Auskunfteien genutzt) und Abbildung V.2 (Merkmal wird zur Beurteilung der Bonität von etablierten Auskunfteien nicht genutzt) zusammenfassend dargestellt.

2.2.2 Einstellung zur Veröffentlichung von Bonitäts-Scores

Lediglich etwas über ein Zehntel der Befragten wären bereit, den eigenen Bonitäts-Score im Internet zu veröffentlichen (12%). Die überwiegende Mehrheit (87%) lehnen das ab und 1% wählte auf diese Frage die Kategorie „weiß nicht“ (0,1% verweigerten die Aussage). Für die Veröffentlichung des Bonitäts-Scores aller Personen in Deutschland waren lediglich 7% der Befragten, die deutliche Mehrheit lehnt das ab (92%); knapp 1% wählte die Kategorie „weiß nicht“ und 0,1% verweigerten die Aussage.

Etwa ein Fünftel der Befragten hat in den letzten fünf Jahren eine Bonitäts-Selbstauskunft eingeholt (21%, n = 461). Von den Personen, die innerhalb der letzten fünf Jahre eine Bonitäts-Selbstauskunft eingeholt hatten, haben 83% den Bonitäts-Score für korrekt gehalten.

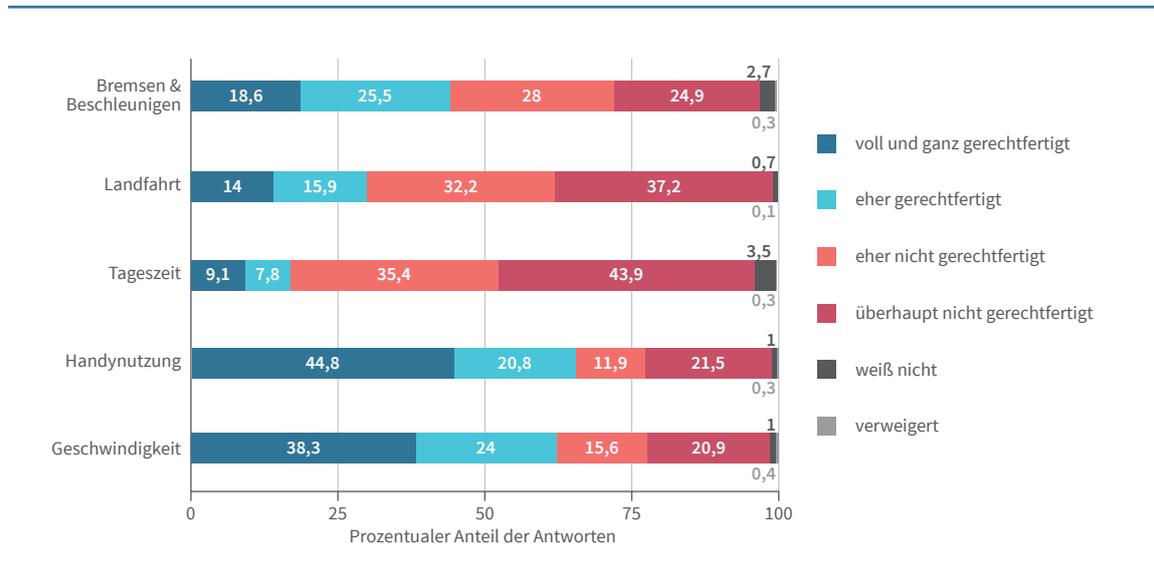


Abbildung V.3

Prozentualer Anteil der Akzeptanz (blau) bzw. Ablehnung (rot) der Befragten (n = 536) bezüglich der Erfassung und Nutzung von Merkmalen für einen Kfz-Telematiktarif mit Bonus-Framing.

ten, wohingegen 14% den Score für inkorrekt gehalten haben (3% wählten die Kategorie „weiß nicht“). Folgende Gründe wurden für die Annahme, dass der Bonitäts-Score nicht korrekt war genannt: (1) Angaben sind nicht korrekt eingetragen (13%), (2) gezahlte Kredite wurden nicht gelöscht (19%), (3) niedriger Wert wegen zu vieler Kredit- bzw. Kontoanfragen (14%), (4) nicht verständlich bzw. zu ungenau (22%), (5) Sonstiges (27%) (4% verweigerten die Angabe von Gründen und 2% antworteten mit „weiß nicht“).

2.2.3 Einstellung zu potenziellen neuartigen Scorings

Neuartige Scorings in der Kfz-Haftpflichtversicherung: Telematiktarife

Nachfolgend werden die deskriptiven Ergebnisse für den *Kfz-Bonus-Split* berichtet (n = 536). Die Berücksichtigung der *Geschwindigkeit* für die Tarifierung bei der Kfz-Haftpflichtversicherung hielten 38% der Befragten für „voll und ganz gerechtfertigt“, etwa ein Viertel (24%) für „eher gerechtfertigt“, 16% für „eher nicht gerechtfertigt“ und etwa ein Fünftel (21%) für „überhaupt nicht gerechtfertigt“, 1% wählte hier die Kategorie „weiß nicht“ (0,4% der Befragten verweigerten die Aussage).

Des Weiteren findet die überwiegende Mehrheit die Berücksichtigung des Merkmals *Handynutzung zum Nachrichtenschreiben oder -lesen* für die Tarifierung als „voll und ganz“ (45%) bzw. „eher gerechtfertigt“ (21%). Etwa ein Drittel hält zuvor genanntes Merkmal zur Berücksichtigung für die Tarifierung „eher nicht“ (12%) bzw. „überhaupt nicht gerechtfertigt“ (22%); 1% der Befragten wählte die Kategorie „weiß nicht“ (0,3% der Befragten verweigerten die Aussage). Auf sehr niedrige Akzeptanzraten stößt die Berücksichtigung der *Tageszeit* als Merkmal für die Tarifierung (tagsüber bei Bonus). Die überwiegende Mehrheit hält die Berücksichtigung dieses Merkmals für „eher nicht“ (35%) bzw. „überhaupt nicht gerechtfertigt“ (44%). Nur 9% bzw. 8% fänden die Berücksichtigung dieses Merkmals „voll und ganz“ bzw. „eher gerechtfertigt“. 4% wählt hier die Kategorie „weiß nicht“ (0,3% der Befragten verweigerten die Aussage). Ebenso stößt die Berücksichtigung des Merkmals *Fahren auf dem Land* für die Tarifierung bei der überwiegenden Mehrheit der Befragten auf Ablehnung (32% bzw. 37% halten dieses für „eher nicht“ bzw. „überhaupt nicht gerechtfertigt“, 14% bzw. 16% halten dieses Merkmal für „voll und ganz“ bzw. „eher gerechtfertigt“, 1% antworteten mit „weiß nicht“ und 0,1% verweigerten die Antwort). Das Merkmal *Brems-*

und *Beschleunigungsverhalten* zur Berücksichtigung für die Tarifierung findet etwas über die Hälfte „eher nicht“ bzw. „überhaupt nicht gerechtfertigt“ (28 % bzw. 25 %). Etwas weniger als die Hälfte hält die Berücksichtigung dieses Merkmals für „voll und ganz“ (19 %) bzw. „eher gerechtfertigt“ (26 %). Der restliche prozentuale Anteil der Antworten entfällt auf die Kategorie „weiß nicht“ bzw. die Verweigerung der Aussage. Der prozentuale Anteil der Akzeptanz bzw. Ablehnung in Bezug auf die zuvor genannten Merkmale für die Tarifierung sind für den Kfz-Bonus-Split zusammenfassend in Abbildung V.3 dargestellt.

Im Folgenden werden die deskriptiven Ergebnisse für den *Kfz-Malus-Split* berichtet (n = 568). Die Berücksichtigung des Merkmals *Geschwindigkeit* für die Tarifierung stößt überwiegend auf Ablehnung (36 % wählten die Kategorie „eher nicht gerechtfertigt“ und 32 % die Kategorie „überhaupt nicht gerechtfertigt“). Zusammengekommen knapp ein Drittel der Befragten ordneten dieses Merkmal für die Tarifierung als „voll und ganz gerechtfertigt“ (14 %) bzw. als „eher gerechtfertigt“ (17 %) ein (0,2 % antworteten mit „weiß nicht“ und weitere 0,2 % verweigerten die Aussage). Das Merkmal *Lesen und Schreiben von Nachrichten auf dem Handy*

während der Fahrt als Merkmal zur Berücksichtigung für die Tarifierung findet die überwiegende Mehrheit „voll und ganz gerechtfertigt“ (62 %) bzw. „eher gerechtfertigt“ (16 %). Zusammengekommen hält lediglich ca. ein Fünftel der Befragten dieses Merkmal für „eher nicht“ (9 %) bzw. „überhaupt nicht gerechtfertigt“ (12 %). Zusammengekommen 1 % der Befragten fiel in die Kategorien „weiß nicht“ bzw. „Verweigerung der Aussage“. Bei einer sehr großen Mehrheit der Befragten stößt die Berücksichtigung der *Tageszeit* (Nacht bei Malus) für die Tarifierung auf Ablehnung (34 % halten das für „eher nicht“ und 62 % für „überhaupt nicht gerechtfertigt“). Lediglich 0,3 % der Befragten ordnet die Berücksichtigung der Tageszeit als „voll und ganz gerechtfertigt“ ein und 3 % als „eher gerechtfertigt“. Zusammengekommen 1 % der Antworten fiel in die Kategorien „weiß nicht“ bzw. „Verweigerung der Aussage“. Ebenso steht der Großteil der Befragten der Berücksichtigung des Merkmals *Stadtfahrt* für die Tarifierung ablehnend gegenüber (38 % halten dieses Merkmal für „eher nicht“ und 45 % für „überhaupt nicht gerechtfertigt“; lediglich 12 % finden das Merkmal „eher“ und 4 % „voll und ganz gerechtfertigt“; 0,4 % wählte die Kategorie „weiß nicht“ und 1 % verweigerte die Aussage). Die Berücksichtigung von *Beschleunigungs- und Bremsverhalten* als Merkmal

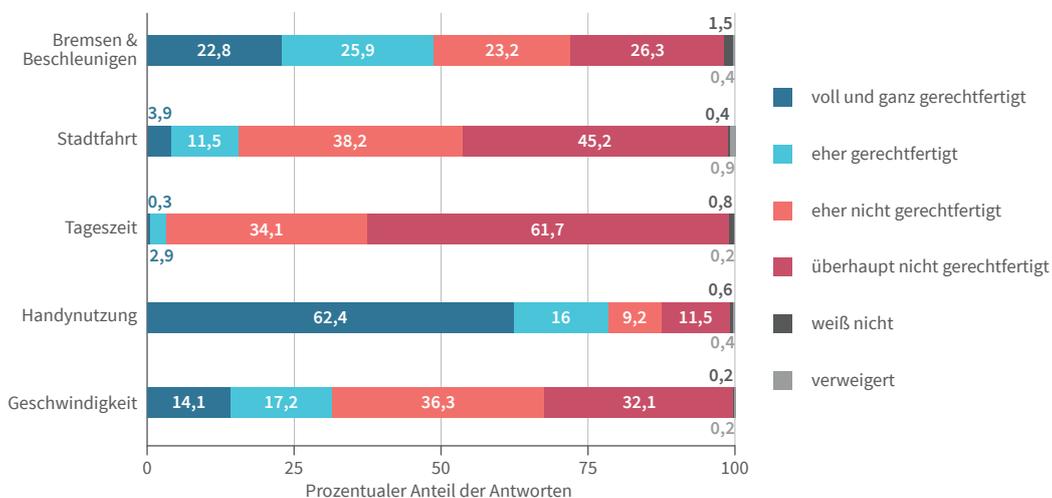


Abbildung V.4

Prozentualer Anteil der Akzeptanz (blau) bzw. Ablehnung (rot) der Befragten (n = 568) bezüglich der Erfassung und Nutzung von Merkmalen für einen Kfz-Telematiktarif mit Malus-Framing.

für die Tarifierung wird jeweils nahezu von der Hälfte befürwortet (23% halten dieses Merkmal „für voll und ganz“ bzw. 26% für „eher gerechtfertigt“) bzw. abgelehnt (23% wählen die Kategorie „eher nicht gerechtfertigt“ bzw. 26% „überhaupt nicht gerechtfertigt“; 2% wählte die Kategorie „weiß nicht“ und 0,4% verweigerte die Aussage). Der prozentuale Anteil der Akzeptanz bzw. Ablehnung in Bezug auf die zuvor genannten Merkmale für die Tarifierung sind für den Kfz-Malus-Split zusammenfassend in Abbildung V.4 dargestellt.

Potenzielle Nutzung des Telematiktarifs

Gut ein Drittel (36%) der Befragten im Kfz-Split (n = 1.160) würden einen Telematiktarif nutzen. Interessanterweise zeigt sich hier auch ein Effekt des Bonus- bzw. Malus-Framings: Während sich 43% (n = 563) der Befragten im Bonus-Split vorstellen können einen Telematiktarif zu nutzen, können das im Malus-Split nur 30% (n = 598) Die Kategorisierung der Antworten für die Zustimmung bzw. Ablehnung in Bezug auf die potenzielle Nutzung eines Kfz-Telematiktarifs sind in Tabelle V.2 dargestellt.

Die Analyse der erstgenannten Gründe für die Zustimmung zeigt, dass diese vor allem auf einem antizipierten Zugewinn an Sicherheit beim Fahren oder in Form reduzierter Unfallrisiken, beruht. Ebenso häufig wurde es als gerechter angesehen, dass sicheres Fahren belohnt wird. Diese Belohnung wurde jedoch nur von einer Minderheit, einem von vier Teilnehmern, die eine Zustimmung begründen, vorgebracht. Von einem kleineren Teil wurde die Möglichkeit persönlicher finanzieller Vorteile gesehen. Eine Vielzahl Gründe wurde nur von Einzelnen

genannt, darunter Umweltvorteile und allgemeine persönliche Vorteile von verschiedenen Fahrertypen. Die Analyse der erstgenannten Gründe für die Ablehnung zeigt, dass zwar allgemeine Datenschutzbedenken am zweithäufigsten vorgebracht wurden, die darüber hinaus konkretisierte Ablehnung des Eingriffs in die Privatsphäre und Überwachung mit Kontrolle jedoch mit 38% der Teilnehmer, die eine Ablehnung begründen, am häufigsten. Somit begründeten mehr als die Hälfte dieser Teilnehmer ihre Ablehnung des Telematiktarifs primär mit der Abgrenzung ihrer privaten bzw. persönlichen Daten und eigenen Verhaltensweisen. Ein kleinerer Teil führte eine Gerechtigkeitsproblematik als ersten Grund an, zum Beispiel, weil es Situationen gibt, die Abweichungen von den Regeln verlangen. Nur vereinzelt wurde die Ablehnung mit dem gewünschten Beibehalten des gegenwärtigen Verhaltens oder mit der Erwartung persönlicher Nachteile begründet. Ebenso sahen nur 3% der Teilnehmer, welche ihre Ablehnung begründeten, Validitäts- und Reliabilitätsprobleme einer solchen Tarifierungsmethode als maßgeblich an. Eine Vielzahl von Gründen wurde abermals nur von Einzelnen genannt, wie die Ablehnung digitaler Technologien oder der Unwille, Tarife zu wechseln. Die häufigsten gemeinsam genannten Worte auf die Frage, ob die Befragten einen Kfz-Telematiktarif nutzen würden sind in Abbildung V.5 illustriert.

Tabelle V.2

Kategorisierte erstgenannte Gründe der Zustimmung bzw. Ablehnung zu einer Kfz-Versicherung mit Telematiktarif.			
BEWERTUNG	KATEGORIE	N	
Zustimmung	es ist gerechter bzw. sicheres Fahrverhalten wird belohnt	25	
	mehr Sicherheit auf den Straßen bzw. weniger Unfälle bzw. sicherer fahren	28	
	Geld sparen durch eigenes Verhalten	18	
	dient der Selbstkontrolle	5	
	unspezifischer, persönlicher Vorteil antizipiert	16	
	generelle Zustimmung, sonstige Nennungen	19	
	TOTAL	111	
Ablehnung	zu viel Überwachung und Kontrolle, Eingriff in die Privatsphäre	87	
	allgemeine Datenschutzgründe, es werden zu viele Daten gesammelt	37	
	ungerecht, denn Fahrverhalten ist z. B. auch situationsabhängig	33	
	ändert Fahrverhalten nicht bzw. möchte Fahrverhalten nicht ändern	6	
	nicht umsetzbar, zu ungenau, fehlerhaft, zu undurchsichtig	7	
	brauche ich nicht, fahre schon vorsichtig bzw. sicher genug	5	
	unspezifischer, persönlicher Nachteil antizipiert	11	
	generelle Ablehnung, sonstige Nennungen	41	
	TOTAL	227	
	Keine Angabe bzw. kann ich nicht sagen, fahre kein Auto bzw. weiß nicht	21	
	TOTAL	359	
	Frage wurde nicht gestellt	1.856	
	TOTAL	2.215	

	ANTEIL PRO BEWERTUNG	BEISPIEL
	23 %	„wer vernünftig fährt, soll auch weniger zahlen“
	25 %	„dadurch passieren weniger Unfälle“
	16 %	„weil, wenn ich vernünftig fahre, habe ich finanzielle Vorteile“
	5 %	„Vorsichtigkeit unterstützt“
	14 %	„weil ich kaum gegen diese Sachen verstoße und eher aufpasse.“
	17 %	„noch zusätzlich einen Vorteil für den Umweltschutz sieht“
	100,0 %	
	38 %	„sich nicht in seiner Freiheit einschränken lassen möchte“
	16 %	„Sollte keine Verbindung zu den Merkmalen haben. Ist zu sehr Datenschutzfrage und sind zu private Daten.“
	15 %	„Ich habe kein Verständnis dafür, es kommt ja immer auch auf die Situation an.“
	3 %	„Wenn Menschen Nachtschicht haben – denke ich nicht, dass sie dann das Fahrverhalten der Menschen beeinflussen.“
	3 %	„weil mir das zu vage ist“
	2 %	„kein Bedarf“
	5 %	„weil er auch ab und zu mal schneller fährt etc. gehört alles dazu würde ihn ein Vermögen kosten“
	18 %	„ich bin für diese ganzen Erneuerungen nicht; vielleicht bin ich zu alt dazu; ich finde diese ganzen Sachen nicht gut“
	100,0 %	



Abbildung V.5

Häufigste gemeinsam genannte Worte auf die Frage „Würden Sie persönlich einen solchen Kfz-Versicherungstarif nutzen, der Merkmale wie etwa Geschwindigkeit, Handynutzung, Tag- oder Nachtfahrt, Land- oder Stadtfahrt und Brems- und Beschleunigungsverhalten berücksichtigt?“ Quellen: ifas, 2018; Martin Brümmer.

Die Mehrheit der Befragten im Kfz-Split möchte grundsätzlich keine Rückmeldung zum Fahrverhalten haben (55%), 17% würden bei starken Änderungen des Fahrverhaltens (Punktwerts) eine Rückmeldung haben wollen, gut jeweils ein Zehntel würde gerne monatlich (12%), bzw. wöchentlich (12%) eine Rückmeldung haben wollen und nur 4% würde täglich eine Rückmeldung begrüßen. Die verbleibenden Prozente entfielen auf die Antwort „weiß nicht“ (1%) bzw. auf die Verweigerung der Aussage (0,3%).

Veröffentlichung des Punktwerts des Telematiktarifs

Die überwiegende Mehrheit würde den eigenen Punktwert aus einem Telematiktarif nicht veröffentlichen (87%, n = 1.104), gut ein Zehntel wäre dazu bereit (12%)

und 1% wählte hier die Kategorie „weiß nicht“. Für die Veröffentlichung des Punktwertes aus dem Telematiktarif aller Autofahrer in Deutschland wäre nur ein Zehntel (10%) der Befragten, 90% lehnten dieses ab.

Einstellung zur verpflichtenden Aufzeichnung der Geschwindigkeit

Gut ein Viertel (28%) der Befragten im Kfz-Split (n = 1.104) befürworteten eine grundsätzliche Pflicht zur Aufzeichnung der Fahrgeschwindigkeit, während dieses bei der bedeutenden Mehrheit auf Ablehnung stieß (72%). Zudem stimmten knapp drei Viertel der Befragten der Aussage, dass durch die verpflichtende Aufzeichnung der Fahrgeschwindigkeit die persönliche Freiheit eingeschränkt werden könnte „voll und ganz“

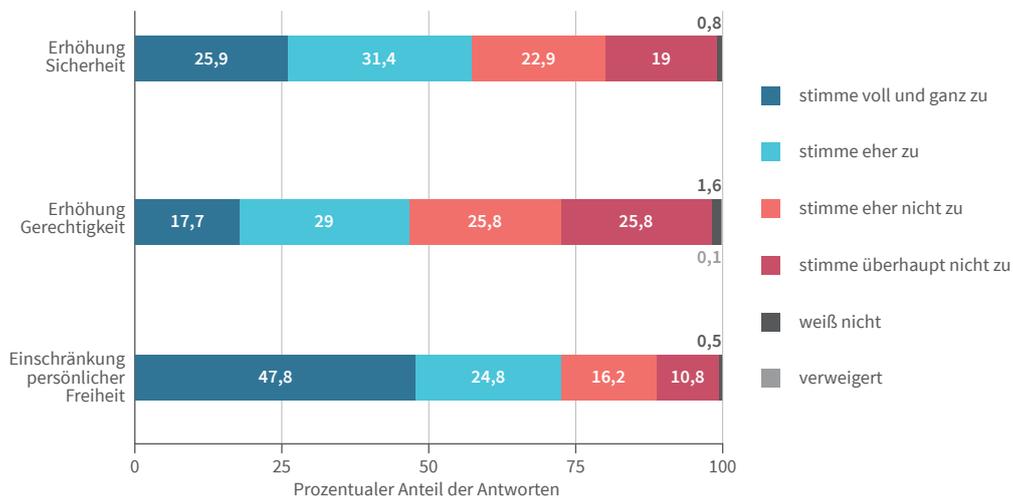


Abbildung V.6

Prozentualer Anteil der Zustimmung (blau) bzw. Ablehnung (rot) der Befragten (n = 1.104) zu einer verpflichtenden Aufzeichnung der Fahrgeschwindigkeit und daraus resultierender potenzieller Konsequenzen.

(48%) bzw. „eher“ zu (25%). Erhöht werden kann die Akzeptanz der verpflichtenden Aufzeichnung der Fahrgeschwindigkeit, wenn diese potenzielle Maßnahme mit positiven Konsequenzen assoziiert präsentiert wird. So stimmte die Mehrheit der Befragten der Aussage, dass durch die Aufzeichnung der Fahrgeschwindigkeit die Sicherheit auf den Straßen erhöht werden könnte „voll und ganz“ (26%) bzw. „eher“ (31%) zu, knapp ein Viertel (23%) stimmte dem „eher nicht“ bzw. knapp ein Fünftel (19%) „überhaupt nicht“ zu. 1% der Antworten entfiel auf die Kategorie „weiß nicht“. Zudem fand die Aussage, dass durch die verpflichtende Aufzeichnung der Fahrgeschwindigkeit die Gerechtigkeit erhöht werden könnte, bei knapp einem Fünftel der Befragten (18%) voll und ganz und bei 29% eher Zustimmung. Gut die Hälfte der Befragten stimmte dieser Aussage „eher nicht“ (26%) bzw. „überhaupt nicht“ (26%) zu. 1% entfiel zusammengefasst auf „weiß nicht“ bzw. die Verweigerung der Aussage. Der prozentuale Anteil der Zustimmung bzw. Ablehnung der Befragten zu einer verpflichtenden Aufzeichnung der Fahrgeschwindigkeit und daraus resultierender potenzieller Konsequenzen ist zusammenfassend in Abbildung V.6 dargestellt.

Neuartige Scorings im Gesundheitsbereich: Verhaltensbasierte Tarife

Im *Gesundheits-Bonus-Split* (n = 570) stieß die Berücksichtigung des Merkmals *tägliche Schrittzahl* für die Tarifierung überwiegend auf Ablehnung (37% wählten die Kategorie „eher nicht gerechtfertigt“ und 36% die Kategorie „überhaupt nicht gerechtfertigt“). Als „voll und ganz gerechtfertigt“ für die Tarifierung ordneten 8% und als „eher gerechtfertigt“ ordneten dieses Merkmal 18% ein (0,7% verweigerte die Aussage, 0,6% wählte die Kategorie „weiß nicht“). Ebenso fand die deutliche Mehrheit das Merkmal *Schlafdauer* zur Berücksichtigung für die Tarifierung „eher nicht“ (35%) bzw. „überhaupt nicht gerechtfertigt“ (50%). Lediglich 5% bzw. 9% fanden dieses Merkmal für die Tarifierung „voll und ganz“ bzw. „eher gerechtfertigt“. Die Berücksichtigung des *Alkoholkonsums* für die Tarifierung ordnete gut die Hälfte der Befragten als „eher nicht“ (24%) bzw. „überhaupt nicht gerechtfertigt“ (32%) ein (17% fand dieses Merkmal „voll und ganz“ und 26% „eher gerechtfertigt“; 0,5% wählten die Antwort „weiß nicht“ und 1% verweigerte die Aussage). Ebenso überwiegend auf Ablehnung für die Tarifierung stieß das Merkmal *Körpergewicht* (33% hielten dieses Merkmal für „eher nicht“ und 32% für „überhaupt nicht gerechtfertigt“;

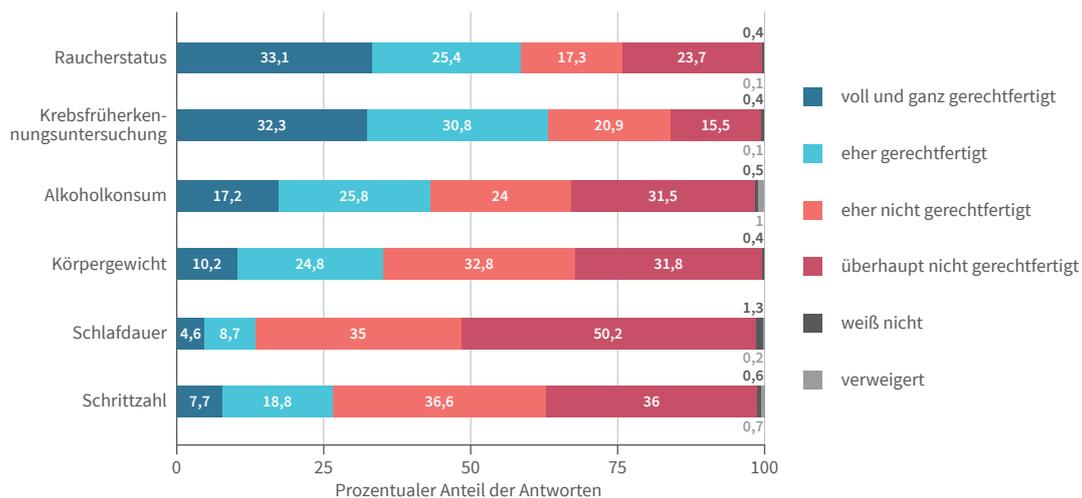


Abbildung V.7

Prozentualer Anteil der Akzeptanz (blau) bzw. Ablehnung (rot) der Befragten (n = 570) bezüglich der Erfassung und Nutzung von Merkmalen für einen Gesundheits-Telematiktarif mit Bonus-Framing.

10% fanden das Merkmal „voll und ganz“ und 25% „eher gerechtfertigt“). Die Berücksichtigung der *Teilnahme an Krebsfrüherkennungsuntersuchungen* als Merkmal für die Tarifierung sahen nahezu zwei Drittel der Befragten als „voll und ganz“ (32%) bzw. „eher gerechtfertigt“ (31%), 21% fanden das „eher nicht“ bzw. 16% als „überhaupt nicht gerechtfertigt“ (0,4% antworteten mit „weiß nicht“ und 0,1% verweigerten die Aussage). Die Berücksichtigung des *Raucherstatus* für die Tarifierung ordnete mehr als die Hälfte der Befragten als „voll und ganz“ (33%) bzw. „eher gerechtfertigt“ (25%) ein; 17% fanden die Berücksichtigung dieses Merkmals für die Tarifierung als „eher nicht“ bzw. 24% als „überhaupt nicht gerechtfertigt“ (0,4% antworteten mit „weiß nicht“ und 0,1% verweigerten die Aussage). Der prozentuale Anteil der Akzeptanz bzw. Ablehnung in Bezug auf die zuvor genannten Merkmale für die Tarifierung sind für den Gesundheit-Bonus-Split zusammenfassend in Abbildung V.7 dargestellt.

Im *Gesundheits-Malus-Split* (n = 541) stieß die Berücksichtigung des Merkmals *tägliche Schrittzahl* für die Tarifierung mit sehr großer Mehrheit auf Ablehnung (37% wählten die Kategorie „eher nicht gerechtfertigt“ und 55% die Kategorie „überhaupt nicht gerechtfertigt“). Als „voll und ganz gerechtfertigt“ für die Tarifierung ordneten dieses Merkmal lediglich 4% der Befragten und als „eher gerechtfertigt“ ordneten es 4% ein (0,6% wählte die Kategorie „weiß nicht“). Ebenso fand die überwiegende Mehrheit die Berücksichtigung der *Schlafdauer* für die Tarifierung „eher nicht“ (30%) bzw. „überhaupt nicht gerechtfertigt“ (64%); lediglich 1% der Befragten ordneten dieses Merkmal für die Tarifierung als „voll und ganz“ und 4% als „eher gerechtfertigt“ ein. Darüber hinaus standen die Befragten der Berücksichtigung des *Alkoholkonsums* für die Tarifierung mehrheitlich ablehnend gegenüber: 31% wählten die Antwort „eher nicht“ und 32% die Antwort „überhaupt nicht gerechtfertigt“. Zusammengefasst

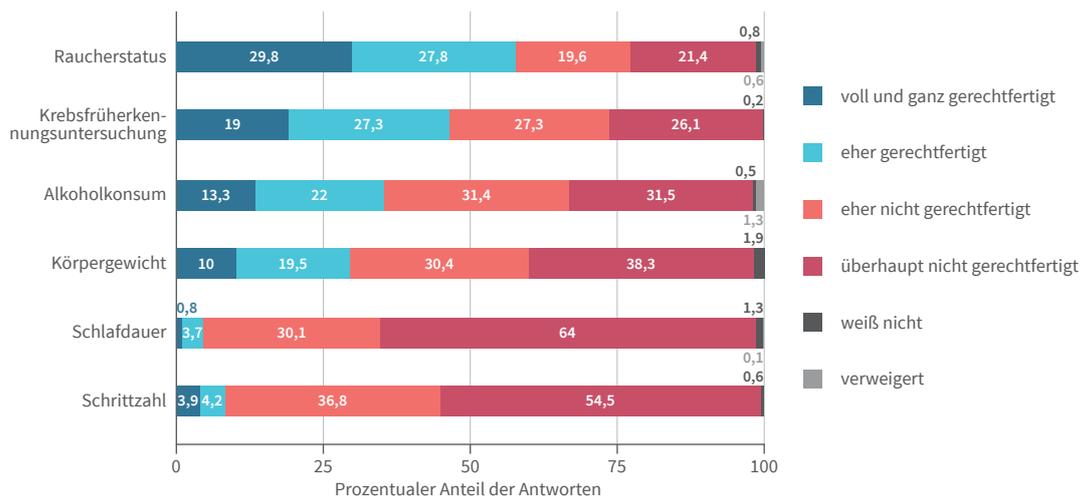


Abbildung V.8

Prozentualer Anteil der Akzeptanz (blau) bzw. Ablehnung (rot) der Befragten (n = 541) bezüglich der Erfassung und Nutzung von Merkmalen für einen Gesundheits-Telematiktarif mit Malus-Framing.

ein Drittel würde dieses Merkmal für die Tarifierung als „voll und ganz“ (13%) bzw. „eher gerechtfertigt“ (22%) einordnen (1% antworteten mit „weiß nicht“ und 1% verweigerten die Aussage). Die Berücksichtigung des Merkmals *Körpergewicht* für die Tarifierung stieß beim überwiegenden Teil der Befragten auf Ablehnung (30% hielten dieses Merkmal für „eher nicht“ und 38% für „überhaupt nicht gerechtfertigt“; 10% fanden das Merkmal für die Tarifierung „voll und ganz“ und 20% als „eher gerechtfertigt“, 2% wählten die Antwort „weiß nicht“). Die Berücksichtigung der *Teilnahme an Krebsfrüherkennungsuntersuchungen* als Merkmal für die Tarifierung sah etwas mehr als die Hälfte der Befragten als „eher nicht“ (27%) bzw. „überhaupt nicht gerechtfertigt“ (26%); 19% fanden dieses Merkmal „voll und ganz“ und 27% als „eher gerechtfertigt“. Der Berücksichtigung des *Raucherstatus* für die Tarifierung wurde durch die Befragten mehrheitlich Akzeptanz entgegen gebracht (30% wählten die Kategorie „voll und ganz“ bzw. 28%

„eher gerechtfertigt“; 20% wählten die Kategorie „eher nicht“ und 21% „überhaupt nicht gerechtfertigt“; 0,8% antworteten mit „weiß nicht“ und 0,6% verweigerten die Aussage). Der prozentuale Anteil der Akzeptanz bzw. Ablehnung in Bezug auf die zuvor genannten Merkmale für die Tarifierung sind für den Gesundheit-Malus-Split zusammenfassend in Abbildung V.8 dargestellt.

Potenzielle Nutzung des verhaltensbasierten Tarifs

Gut ein Drittel (34%) der Befragten im Gesundheits-Split (n = 1.111) können sich vorstellen, einen verhaltensbasierten Tarif zu nutzen. Im Gegensatz zum Kfz-Bereich zeigte sich hier im Antwortverhalten kein nennenswerter Effekt des Bonus- bzw. Malus-Framings. Die Kategorisierung der Antworten für die Zustimmung bzw. Ablehnung in Bezug auf die potenzielle Nutzung eine verhaltensbasierten Krankenversicherung Tabelle V.3 dargestellt.

Tabelle V.3

Kategorisierte erstgenannte Gründe der Zustimmung bzw. Ablehnung zu einer verhaltensbasierten Krankenversicherung.

BEWERTUNG	KATEGORIE	N	
Zustimmung	es ist gerechter bzw. gesunde Lebensweise wird belohnt	18	
	man würde gesünder leben, mehr auf die Gesundheit achten (z. B. Kontrolluntersuchungen nutzen)	21	
	Geld sparen durch eigenes Verhalten	13	
	dient der Selbstkontrolle	6	
	unspezifischer, persönlicher Vorteil antizipiert	24	
	generelle Zustimmung, sonstige Nennungen	21	
	TOTAL	103	
Ablehnung	zu viel Überwachung und Kontrolle, Eingriff in die Privatsphäre	91	
	allgemeine Datenschutzgründe, es werden zu viele Daten gesammelt	20	
	ungerecht, für Krankheit kann man nichts	39	
	ändert Verhalten nicht bzw. möchte Verhalten nicht ändern	5	
	nicht umsetzbar, zu ungenau, fehlerhaft, zu undurchsichtig	14	
	brauche ich nicht, lebe schon gesund	6	
	unspezifischer, persönlicher Nachteil antizipiert	12	
	generelle Ablehnung, sonstige Nennungen	36	
TOTAL	223		
	Keine Angabe bzw. kann oder will ich nicht sagen, weiß nicht	34	
	TOTAL	360	
	Frage wurde nicht gestellt	1.855	
	TOTAL	2.215	

	ANTEIL PRO BEWERTUNG	BEISPIEL
	18 %	„Höhere Gerechtigkeit gegenüber den anderen Beitragszahlern“
	20 %	„es wird mehr auf Gesundheit geachtet“
	13 %	„Weil ich mir finanzielle Vorteile erhoffe“
	6 %	„Manchmal ist es gut, wenn jemand da ist um zu sehen was los ist.“
	23 %	„würde davon profitieren“
	20 %	„positiver Eindruck“
	100,0 %	
	41 %	"Weil ich mich dann der vollen Überwachung ausliefern"
	9 %	„der Datenschutz rechtfertigt dies nicht, und verletzt diesen“
	18 %	„Weil man nichts dafür kann, für die jeweiligen Krankheitsbilder“
	2 %	„weil ich mich nicht danach richten würde, zum Beispiel rauche ich“
	6 %	„Weil es schwierig nachzuweisen ist, ob man diese Angaben erfüllt.“
	3 %	„derzeit genug bezahlt und das ist so gut“
	6 %	„bin nicht der gesunde Typ“
	16 %	„Weil ich nicht oder zu wenig laufe“
	100,0 %	

17 % monatlich und 8 % wöchentlich. Lediglich 3 % der Befragten würden täglich eine Rückmeldung begrüßen.

Veröffentlichung des Punktwerts des verhaltensbasierten Tarifs

Nur sehr wenige Befragte können es sich vorstellen, den Punktwert aus einem verhaltensbasierten Krankenversicherungstarifs zu veröffentlichen – so sie einen nutzen würden. Auf die Frage, ob sie den Punktwert veröffentlichen würden antworteten 95 % mit „nein“ und 5 % mit „ja“ (n = 1.111). Die Veröffentlichung des Punktwertes aus der Krankenversicherung aller Personen in Deutschland lehnte ebenfalls fast jeder Befragte ab (94 %).

Einstellung zur verpflichtenden Aufzeichnung der täglichen Schrittzahl

Für die verpflichtende Aufzeichnung der täglichen Anzahl von Schritten plädierten lediglich 8 % der Befragten, 91 % lehnten dieses ab (n = 1.111). Bei der Darbietung von positiven Konsequenzen einer Schrittzahlaufzeichnung konnte die Akzeptanzrate jedoch erhöht werden, wenngleich die Mehrheit weiterhin ablehnend ist: Der

Aussage, dass durch die Aufzeichnung der täglichen Anzahl von Schritten die Gesundheit der Bevölkerung verbessert werden würde, steht die Mehrheit der Befragten ablehnend gegenüber (gut ein Drittel (34 %) stimmte dem „eher nicht“ bzw. knapp ein Drittel (32 %) „überhaupt nicht“ zu, 11 % stimmten dieser Aussage „voll und ganz“ bzw. „eher“ (22 %) zu). Die Aussage, dass durch die tägliche verpflichtende Aufzeichnung von Schritten die Gerechtigkeit erhöht werden könnte, traf bei der überwiegenden Mehrheit der Befragten auf Ablehnung (37 % stimmten „eher nicht“ bzw. 45 % stimmten „überhaupt nicht“ zu; 5 % stimmten „voll und ganz“ und 12 % stimmten „eher“ zu). Die Aussage, dass durch die verpflichtende Aufzeichnung der täglichen Schrittzahl die persönliche Freiheit eingeschränkt werden könnte, fand bei der Mehrheit der Befragten Zustimmung (53 % stimmten „voll und ganz“ zu, 19 % stimmten „eher zu“; 14 % stimmten „eher nicht“ und 14 % „überhaupt nicht“ zu). Der prozentuale Anteil der Zustimmung bzw. Ablehnung der Befragten zu einer verpflichtenden Aufzeichnung der Schrittzahl und daraus resultierender potenzieller Konsequenzen ist zusammenfassend in Abbildung V.10 dargestellt.

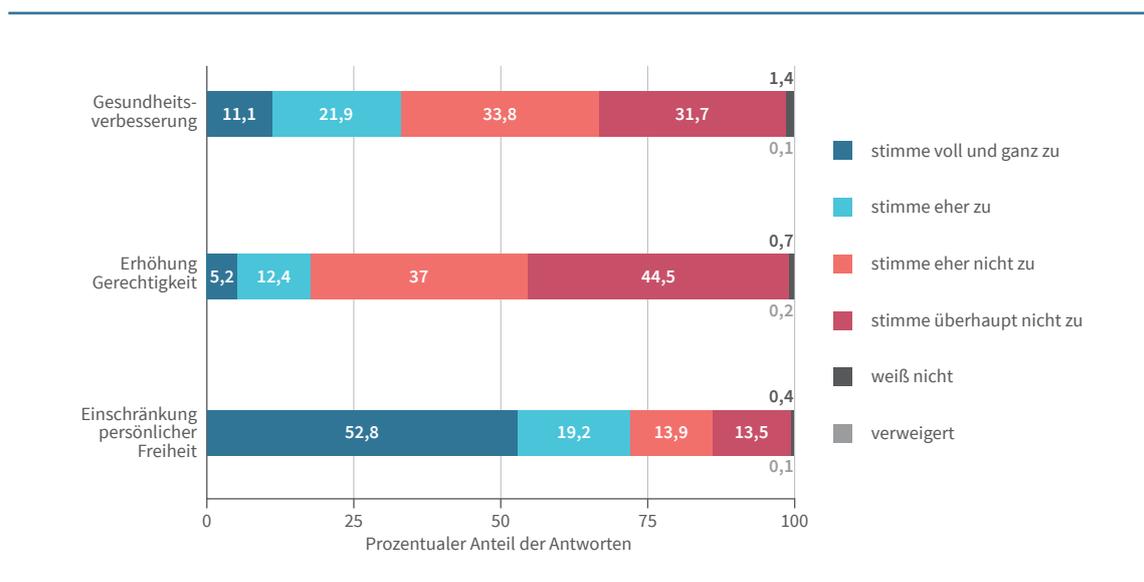


Abbildung V.10

Prozentualer Anteil der Zustimmung (blau) bzw. Ablehnung (rot) der Befragten (n = 1.111) zu einer verpflichtenden Aufzeichnung der Schrittzahl und daraus resultierender potenzieller Konsequenzen.

Tabelle V.4

Kategorisierte erstgenannte Gründe der Zustimmung bzw. Ablehnung zu einem KFZ-Gesundheitskombinationstarif.			
BEWERTUNG	KATEGORIE	N	
Zustimmung	es ist gerechter, gesunde Lebens- bzw. Fahrweise wird belohnt	10	
	mehr Sicherheit auf den Straßen bzw. weniger Unfälle bzw. sicherer fahren	22	
	Geld sparen durch eigenes Verhalten	26	
	dient der Selbstkontrolle	11	
	unspezifischer, persönlicher Vorteil antizipiert	20	
	generelle Zustimmung, Vorteile, sonstige Nennungen	60	
	TOTAL	149	
Ablehnung	zu viel Überwachung und Kontrolle, Eingriff in die Privatsphäre	234	
	allgemeine Datenschutzgründe, es werden zu viele Daten gesammelt bzw. von Versicherungen zusammengeführt	156	
	ungerecht, denn Fahrverhalten ist z. B. auch situationsabhängig	6	
	nicht umsetzbar, zu ungenau, fehlerhaft, zu undurchsichtig	22	
	brauche ich nicht, fahre schon vorsichtig bzw. sicher genug	13	
	unspezifischer, persönlicher Nachteil antizipiert	15	
	generelle Ablehnung, sonstige Nennungen	107	
TOTAL	553		
	Keine Angabe bzw. kann ich nicht sagen, fahre kein Auto bzw. weiß nicht / verweigert	53	
	TOTAL	755	
	Frage wurde nicht gestellt	1.460	
	TOTAL	2.215	

	ANTEIL PRO BEWERTUNG	BEISPIEL
	7%	„Weil das eigene gute Verhalten durch Vergütung finanziell belohnt wird“
	15%	„Sicher fahren und weniger Stress sind gesünder und führen zu weniger Unfällen“
	17%	„Weil ich davon ausgehe, dass ich sicher fahre und dadurch Geld sparen kann“
	7%	„damit man mehr darauf achtet“
	13%	„weil ich mich vernünftig im Verkehr verhalte“
	40%	„Da ich mit Kindern jetzt vorsichtiger fahre und mit dem Alter auch immer vorsichtiger werde und auch auf meine Gesundheit achte“
	100 %	
	42%	„möchte nicht elektronisch überwacht werden“
	28%	„viel zu viel Datensammelei“
	1%	„aus Solidaritätsgründen“
	4%	„weil die Kombination zwischen den beiden Tarifen scheint nicht sinnvoll“
	2%	„Weil meine KFZ Versicherungen schon seit Jahrzehnten bestehen und ich nichts mehr ändern möchte.“
	3%	„denn ich liebe hohe Geschwindigkeiten“
	19%	„Weil das leicht zu manipulieren ist“
	100 %	

Potenzielle Nutzung des Kombitarifs

Für die potenzielle Nutzung eines Kombitarifs, bei dem für eine Preisersparnis auch der Krankenversicherung Daten zum Fahrverhalten zur Verfügung gestellt werden müsste, gaben lediglich 5 % der Befragten an (n = 1.104), dass sie so einen Tarif mit Sicherheit nutzen würden. Knapp ein Fünftel wählte die Antwort „ja, möglicherweise“ (19%). Mehrheitlich wurde sich jedoch gegen die Nutzung eines solchen Kombitarifs ausgesprochen (25 % wählten die Kategorie „nein, eher nicht“ und 51 % „nein, sicher nicht“).

Für die potenzielle Nutzung eines Kombitarifs, bei dem für eine Preisersparnis auch der Kfz-Versicherung Daten zum Gesundheitsverhalten zur Verfügung gestellt werden müsste, gaben 8 % der Befragten an (n = 1.111), dass sie dieses mit Sicherheit tun würden. Knapp ein Viertel wählte die Antwort „ja, möglicherweise“ (23%). Mehrheitlich wurde sich jedoch gegen die Nutzung eines solchen Kombitarifs ausgesprochen (22 % wählten die Kategorie „nein, eher nicht“ und 46 % „nein, sicher nicht“, 1 % wählte die Kategorie „weiß nicht“).

Die Auswertung der Gründe in Bezug auf die Zustimmung bzw. Ablehnung zum Kombitarif ist für beide Splits (Gesundheit und Kfz) zusammenfassend in Tabelle V.4 dargestellt.

Die Analyse der erstgenannten Gründe für die Zustimmung zeigt, dass neben der potenziellen finanziellen Belohnung vor allem der antizipierte Zugewinn an Sicherheit beim Fahren oder in Form reduzierter Unfallrisiken, vorgebracht wurde. Gerechtigkeit sowie mögliche Unterstützung der persönlichen Selbstkontrolle wurden nur von jeweils sechs bis zehn Probanden als Begründung herangezogen.

Die Analyse der erstgenannten Gründe für die Ablehnung zeigt, dass allgemeine Datenschutzbedenken am zweithäufigsten vorgebracht wurden, die darüber hinaus konkretisierte Ablehnung des Eingriffs in die

Privatsphäre und Überwachung mit Kontrolle jedoch mit 42 % der Teilnehmer, die eine Ablehnung begründeten, am häufigsten. Somit begründeten mehr als zwei Drittel dieser Teilnehmer ihre Ablehnung des Kombinationstarifs primär mit der Abgrenzung ihrer privaten bzw. persönlichen Daten und Verhaltensweisen von anderen. Ein kleinerer Teil sah Validitäts- und Reliabilitätsprobleme einer solchen Tarifierungsmethode als maßgeblich an. Eine Vielzahl von Gründen wurde nur von Einzelnen genannt, wie Ungerechtigkeit oder die Manipulationsgefahr.

Umfassendes soziales Bürgerscoring

Lediglich knapp ein Zehntel (9%) der Befragten in der Gesamtstichprobe (n = 2.215) befürworteten ein umfassendes Bürgerscoring, während die überwiegende Mehrheit dieses ablehnte (91%). Der Aussage, dass durch ein solches Bürgerscoring die Sicherheit erhöht werden könnte, stimmte knapp ein Drittel der Befragten „voll und ganz“ (9%) bzw. „eher zu“ (22%). Die Mehrheit stimmte dieser Aussage jedoch „eher nicht“ (34%) bzw. „überhaupt nicht“ zu (34%), während 1% die Kategorie „weiß nicht“ wählte und 0,1% die Antwort verweigerte. Der Aussage, dass durch ein solches System die Gerechtigkeit erhöht werden könnte, stimmte gut ein Viertel (eher) zu (7% „voll und ganz“; 19% „eher“), während drei Viertel der Befragten dieser Aussage (eher) nicht zustimmten (35% „eher nicht“ bzw. 39% „überhaupt nicht“); 0,7% wählte die Kategorie „weiß nicht“ und 0,2% verweigerten die Aussage. Der Aussage, dass durch ein solches System die persönliche Freiheit eingeschränkt werden könnte stimmte die überwiegende Mehrheit zu (63% stimmten „voll und ganz“ und 19% stimmten „eher zu“). 11% der Befragten stimmten dieser Aussage „eher nicht“ und 8% „überhaupt nicht“ zu. Der prozentuale Anteil der Zustimmung bzw. Ablehnung der Befragten zu einem sozialen Bürgerscoring und daraus resultierender potenzieller Konsequenzen ist zusammenfassend in Abbildung V.11 dargestellt.

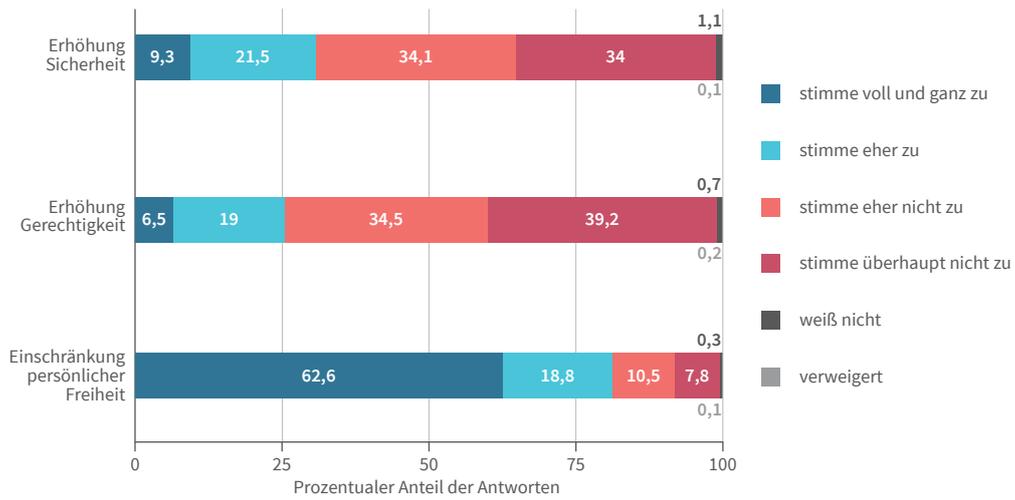


Abbildung V.11

Prozentualer Anteil der Zustimmung (blau) bzw. Ablehnung (rot) der Befragten (n = 2.215) zu einem sozialen Bürgerscoring und daraus resultierender potenzieller Konsequenzen.

Literatur

- Abel, R. B.** (2018). Automatisierte Entscheidungen im Einzelfall gem. Art. 22 DS-GVO. Anwendungsbereich und Grenzen im nicht-öffentlichen Bereich. *Zeitschrift für Datenschutz*, 8(7), 304–307.
- acatech & Körber-Stiftung.** (2018). *TechnikRadar 2018. Was die Deutschen über Technik denken*. München/Hamburg: acatech, Körber-Stiftung.
- Adam, L. & Micklitz, H.-W.** (2016). *Information, Beratung und Vermittlung in der digitalen Welt. Rechtsfragen in Finanzen, Gesundheit und Handel. Veröffentlichung des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Adomeit, K.** (2002). Diskriminierung – Inflation eines Begriffs. *Neue Juristische Wochenschrift*, 55(22), 1622–1623.
- Al-Ani, A.** (2018). Das chinesische Social Credit System – ein digitales Gesellschaftsmodell mit globalen Auswirkungen. In Bundesministerium der Verteidigung (Hrsg.), *China – Perspektiven und Herausforderungen. Bericht zur 3. Sitzung des Netzwerks „Strategie und Vorausschau“*. Berlin: BMVG.
- Alber, M., Lapuschkin, S. & Seeger, P.** (2018). Investigate neural networks! *arXiv preprint arXiv:1808.04260v1*.
- Albrecht, U.** (Hrsg.) (2016). *Chancen und Risiken von Gesundheits-Apps (CHARISMA)*. Hannover: Medizinische Hochschule Hannover.
- Ashley, K. D.** (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge: Cambridge University Press.
- Auer, B. & Rottmann, H.** (2015). *Statistik und Ökonometrie für Wirtschaftswissenschaftler: Eine anwendungsorientierte Einführung*. Wiesbaden: Gabler Verlag/Springer Fachmedien Wiesbaden GmbH.
- Bäcker, M.** (2018). Kommentierung des Art. 13 DSGVO. In B. Buchner & J. Kühling (Hrsg.), *Datenschutz-Grundverordnung und BDSG. Kommentar*. München: Verlag C. H. Beck.
- Baecke, P. & Bocca, L.** (2017). The value of vehicle telematics data in insurance risk selection processes. *Decision Support Systems*, 98, 69–79.
- Bala, C. & Schuldzinski, W.** (2017). Einleitung: Schöne neue Verbraucherwelt? Big Data, Scoring und das Internet der Dinge. In C. Bala & W. Schuldzinski (Hrsg.), *Schöne neue Verbraucherwelt? Big Data, Scoring und das Internet der Dinge – Beiträge zur Verbrauchersforschung Band 5* (S. 7–20). Düsseldorf: Verbraucherzentrale NRW.
- Bambaucher, J. M. & Zarsky, T.** (2018). The algorithm game. *Notre Dame Law Review*, (im Erscheinen).
- Barocas, S. & Selbst, A. D.** (2016). Big data's disparate impact. *California Law Review*, 104, 671–732.
- Beckhusen, G. M.** (2004). *Der Datenumgang innerhalb des Kreditinformationssystems der SCHUFA*. Baden-Baden: Nomos Verlagsgesellschaft.
- Beckhusen, G. M.** (2005). Das Scoring-Verfahren der SCHUFA im Wirkungsbereich des Datenschutzrechts. *Zeitschrift für Bank- und Kapitalmarktrecht*, 5(9), 335–344.
- Ben-Shahar, O. & Schneider, C. E.** (2014). *More than you wanted to know: the failure of mandated disclosure*. Princeton, N.J., USA: Princeton University Press.
- Berg, T., Burg, V., Gombović, A. & Puri, M.** (2018). *On the rise of fintechns – credit scoring using digital footprints (No. w24551)*. National Bureau of Economic Research.
- Berk, R., Heidari, H., Jabbari, S., Kearns, M. & Roth, A.** (2017). Fairness in criminal justice risk assessments: The state of the art. *arXiv preprint arXiv:1703.09207*.
- Bian, Y., Yang, C., Zhao, J. L. & Liang, L.** (2018). Good drivers pay less: A study of usage-based vehicle insurance models. *Transportation Research Part A: Policy and Practice*, 107, 20–34.
- Bilger, A., Löwel, V. & Tomaszewski, L.** (2018). *Wer steuert die Künstliche Intelligenz? Ein Podcast der Heinrich Böll Stiftung vom 30. Januar 2018*. Abgerufen am 12. Juni 2018 von <https://www.boell.de/de/2018/01/30/wer-steuert-die-kuenstliche-intelligenz>
- Bitkom** (2014). *Pay-as-u-drive-Versicherung mit Potenzial*. Abgerufen am 28. September 2018 von <https://www.bitkom.org/Presse/Presseinformation/Pay-as-u-Drive-Versicherungen-mit-Potenzial.html>
- Bitter, P. & Uphues, S.** (2017). *Big Data und die Versicherten-gemeinschaft – „Entsolidarisierung“ durch Digitalisierung?* Abgerufen am 26. September 2018 von www.abida.de/sites/default/files/13%20Entsolidarisierung.pdf
- Blackbaud** (2014). *Target analytics prospectpoint*. Abgerufen am 14. Mai 2018 von https://www.blackbaud.com/files/resources/downloads/01.14.ANLY_%20Prospect-Point.datasheet.pdf
- Block, J.** (2018). Kommentierung von § 3 AGG. In B. Gsell, W. Krüger, S. Lorenz & C. Reymann (Hrsg.), *Beck-Online Großkommentar zum AGG (Stand 01.03.2018)*.
- BMWi & BMJV** (2015). *BMWi/BMJV-Maßnahmenprogramm „Mehr Sicherheit, Souveränität und Selbstbestimmung in der digitalen Wirtschaft“*. Herausforderungen und Handlungselemente für Gesellschaft, Wirtschaft und Verbraucher. Berlin: Bundesministerium für Wirtschaft und Energie, Bundesministerium der Justiz und für Verbraucherschutz.
- Boehme-Neßler, V.** (2016). Das Ende der Anonymität. Wie Big Data das Datenschutzrecht verändert. *Datenschutz und Datensicherheit*, 40(7), 419–423.
- Boetius, J.** (2017). Kommentierung von § 203 VVG. In T. Langheid & M. Wandt (Hrsg.), *Münchener Kommentar zum VVG*. München: Verlag C. H. Beck.
- Bolton, R. J. & Hand, D. J.** (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235–249.

- Braun, M. & Nürnberg, V.** (2015). Verhaltensbasierte Versicherungstarife – innovative E-Health-Initiative oder Ausstieg aus der Solidargemeinschaft? *Gesundheits- und Sozialpolitik*, 69(1), 70–75.
- Brighton, H. & Gigerenzer, G.** (2015). The bias bias. *Journal of Business Research*, 68(8), 1772–1784.
- Britz, G.** (2008). *Einzelfallgerechtigkeit versus Generalisierung. Verfassungsrechtliche Grenzen statistischer Diskriminierung*. Tübingen: Mohr Siebeck.
- Brömmelmeyer, C.** (2017). Belohnungen für gesundheitsbewusstes Verhalten in der Lebens- und Berufsunfähigkeitsversicherung? Rechtliche Rahmenbedingungen für Vitalitäts-Tarife. *Recht und Schaden*, 44(5), 225–232.
- Buchner, B.** (2017). Datenverarbeitung im nicht-öffentlichen Bereich. In M.-T. Tinnefeld, B. Buchner, T. Petri & H.-J. Hof, *Einführung in das Datenschutzrecht. Datenschutz und Informationsfreiheit in europäischer Sicht* (S. 393–468). Berlin: Oldenbourg Verlag.
- Buchner, B.** (2018). Kommentierungen von Art. 4 Nr. 4 DSGVO u. a. In B. Buchner & J. Kühling (Hrsg.), *Datenschutz-Grundverordnung und BDSG. Kommentar*. München: Verlag C. H. Beck.
- Budzinski, O. & Schneider, S.** (2017). Smart Fitness: Ökonomische Effekte einer Digitalisierung der Selbstvermessung. *List Forum für Wirtschafts- und Finanzpolitik*, 43(2), 89–124.
- Bull, H. P.** (2011). *Informationelle Selbstbestimmung – Vision oder Illusion? Datenschutz im Spannungsverhältnis von Freiheit und Sicherheit*. Tübingen: Mohr Siebeck.
- Bundesversicherungsamt** (2016). *Tätigkeitsbericht 2016*. Abgerufen am 26. September 2018 von www.bundesversicherungsamt.de/fileadmin/redaktion/Presse/epaper2016/index.html
- Bundesversicherungsamt** (2018). *Sonderbericht: Zum Wettbewerb in der gesetzlichen Krankenversicherung*. Bonn: Bundesversicherungsamt.
- Burrell, J.** (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.
- Butzer, H.** (2001). *Fremdlasten in der Sozialversicherung. Zugleich ein Beitrag zu den verfassungsrechtlichen Vorgaben für die Sozialversicherung*. Tübingen: Mohr Siebeck.
- Bygrave, L. A.** (2001). Automated profiling: Minding the machine: Article 15 of the ec data protection directive and automated profiling. *Computer Law & Security Review*, 17(1), 17–24.
- Calabresi, G. & Bobbitt, P.** (1978). *Tragic choices. The fels lectures on public policy analysis*. New York, USA: W.W. Norton & Company.
- Calliess, G.-P.** (2006). *Grenzüberschreitende Verbraucherverträge: Rechtssicherheit und Gerechtigkeit auf dem elektronischen Weltmarktplatz*. Tübingen: Mohr Siebeck.
- Carr, A.** (2016). *I found out my secret internal tinder rating and now I wish I hadn't*. Abgerufen am 21. September 2018 von <https://www.fastcompany.com/3054871/whats-your-tinder-score-inside-the-apps-internal-ranking-system>
- CDU/CSU/SPD** (2018). *Ein neuer Aufbruch für Europa. Eine neue Dynamik für Deutschland. Ein neuer Zusammenhalt für unser Land. Koalitionsvertrag zwischen CDU, CSU und SPD. 19. Legislaturperiode*. Abgerufen am 25. September 2018 von www.bundesregierung.de/Content/DE/_Anlagen/2018/03/2018-03-14-koalitionsvertrag.pdf
- Center for International Governance Innovation** (2017). *Global survey on internet security and trust*. Abgerufen am 27. September 2018 von <https://www.cigionline.org/internet-survey-2017>
- Chege, V.** (2012). EU-Antidiskriminierungsrichtlinien und EU-Gleichstellungsrecht: Praktische Erfahrungen bei Fällen mehrdimensionaler Diskriminierung. *Neue Justiz*, 13(12), 503–511.
- Chittaranjan, G., Blom, J. & Gatica-Perez, D.** (2011). Who's who with big-five: analyzing and classifying personality traits with smartphones. *15th Annual International Symposium on Wearable Computers*, (S. 29–36). San Francisco, CA.
- Chouldechova, A.** (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153–163.
- Christl, W.** (2014). *Kommerzielle digitale Überwachung im Alltag. Studie im Auftrag der Bundesarbeitskammer Wien*. Abgerufen am 23. Mai 2018 von http://crackedlabs.org/dl/Studie_Digitale_Ueberwachung.pdf
- Christl, W. & Spiekermann, S.** (2016). *Networks of control*. Wien: Facultas.
- Christoph, G. & Hackel, H.** (2002). *Starthilfe Stochastik*. Stuttgart/Leipzig/Wiesbaden: B.G. Teubner.
- Consumer Federation of America & Vantage Score Solutions** (2016). *6th annual credit score survey*. Washington/ Stamford.
- Consumer Reports** (2015). *The secret score behind your rates*. *Consumer Reports* (30. Juli 2015). Abgerufen am 26. Juni 2018 von <https://www.consumerreports.org/cro/car-insurance/credit-scores-affect-auto-insurance-rates/index.htm>
- Dammann, U.** (2016). Erfolge und Defizite der EU-Datenschutzgrundverordnung – Erwarteter Fortschritt, Schwächen und überraschende Innovationen. *Zeitschrift für Datenschutz*, 6(7), 307–314.
- Davis, K. & Fisher, A. (Hrsg.)** (2012). *Governance by indicators. Global power through quantification and rankings*. Oxford/New York: Oxford University Press.
- De Montjoye, Y.-A., Hidalgo, C., Verleysen, M. & Blondel, V.** (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, Artikel 1376.

- De Montjoye, Y.-A., Quoidbach, J., Robic, F. & Pentland, A.** (2013). Predicting personality using novel mobile phone-based metrics. In A. Greenberg, W. Kennedy & N. Bos (Hrsg.), *Proceedings of the 6th international conference on social computing, behavioral-cultural modeling and prediction (SBP'13)*. (S. 48–55). Berlin, Heidelberg: Springer Verlag.
- Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit** (2016). *25. Tätigkeitsbericht Datenschutz des Hamburgischen Beauftragten für Datenschutz und Informationsfreiheit 2014/2015*. Abgerufen am 4. September 2018 von http://daten.transparenz.hamburg.de/Dataport.HmbTG.ZS.Webservice.GetResource100/GetRessource100.svc/19c7c7c5-4960-48e8-81c7-9f3ba6963278/Akte_T20.pdf
- Der Hessische Datenschutzbeauftragte** (2014). *Dreiundvierzigster Tätigkeitsbericht des Hessischen Datenschutzbeauftragten Professor Dr. Michael Ronellenfitsch*. Abgerufen am 4. September 2018 von https://datenschutz.hessen.de/sites/datenschutz.hessen.de/files/content-downloads/2014_43_TB.pdf
- Der Hessische Datenschutzbeauftragte** (2017). *Sechsendvierzigster Tätigkeitsbericht des Hessischen Datenschutzbeauftragten Professor Dr. Michael Ronellenfitsch*. Abgerufen am 4. September 2018 von https://datenschutz.hessen.de/sites/datenschutz.hessen.de/files/2017_46_TB_0.pdf
- Deutsche Aktuarvereinigung e. V.** (2017). *Ergebnisbericht des Ausschusses Schadenversicherung. Aktuarieller Umgang mit Big Data in der Schadenversicherung*. Abgerufen am 27. September 2018 von https://aktuar.de/unsere-themen/fachgrundsaeetze-oeffentlich/2017-11-20_DAV-Ergebnisbericht_Data-Science-in-der-Kompositversicherung.pdf
- Deutscher Ethikrat** (2017). *Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme*. Berlin: Deutscher Ethikrat.
- Dixon, P. & Gellmann, R.** (2014). *The scoring of america: How secret consumer scores threaten your privacy and your future*. Abgerufen am 7. Juni 2018 von http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf
- Domurath, I. & Neubeck, I.** (2018). *Verbraucherscoring aus Sicht des Datenschutzrechts. Veröffentlichungen des Sachverständigenrat für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Dreher, M.** (1991). *Die Versicherung als Rechtsprodukt*. Tübingen: Mohr Siebeck.
- Dreher, M.** (2018). Kommentierung der Einleitung zum VAG u. a. In E. R. Prölss & M. Dreher (Hrsg.), *Versicherungsaufsichtsgesetz mit Nebengesetzen*. München: Verlag C.H. Beck.
- Dressel, J. F.** (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580. DOI: 10.1126/sciadv.aao5580.
- Dreyer, S. & Schulz, W.** (2018). *Was bringt die Datenschutz-Grundverordnung für automatisierte Entscheidungssysteme?* Gütersloh: Bertelsmann Stiftung.
- Duhigg, C.** (2012). *How companies learn your secrets*. Abgerufen am 14. Mai 2018 von https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&pagewanted=all
- Durand, D.** (1941). *Risk elements in consumer instalment financing*. New York: National Bureau of Economic Research.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R.** (2012). Fairness through awareness. *Proceedings of the 3rd innovations in theoretical computer science conference* (S. 214–226). ACM.
- Equifax** (2018). *Discretionary spending index*. Abgerufen am 14. Mai 2018 von <https://www.equifax.com/business/discretionary-spending-index/>
- Erdmann, E.** (2018). *SCHUFA, öffne dich!* Abgerufen am 26. September 2018 von www.zeit.de/digital/datenschutz/2018-03/openSCHUFA-SCHUFA-transparenz-score-algorithmen
- Executive Office of the President of the United States** (2016). *Big data: A report on algorithmic systems, opportunity, and civil rights*. Abgerufen am 26. September 2018 von https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
- Federal Trade Commission** (2014). *Data brokers. A call for transparency and accountability*. Abgerufen am 23. Mai 2018 von <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>
- Fischer, S. & Petersen, T.** (2018). *Was Deutschland über Algorithmen weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage im Auftrag der Bertelsmann Stiftung*. Gütersloh: Bertelsmann Stiftung.
- French, J. R.** (1953). Experiments in field settings. In L. Festinger & D. Katz (Hrsg.), *Research methods in the behavioral sciences* (S. 98–135). New York: Holt, Rinehart and Winston.
- Frenzel, E. M.** (2018). Kommentierung des Art. 5 DSGVO. In B. P. Paal & D. A. Pauly (Hrsg.), *Beck'sche Kompakt-Kommentare. Datenschutz-Grundverordnung und Bundesdatenschutzgesetz*. München: Verlag C. H. Beck.
- Fries, M.** (2016). *Verbraucherrechtsdurchsetzung*. Tübingen: Mohr Siebeck.
- Fritzsche, K. P.** (2017). Zur Begründung des Diskriminierungsverbots. Historische Erfahrungen, philosophische Argumente und menschenrechtliche Entwicklungen. In A. Scherr, A. El-Mafaalani & G. Yüksel (Hrsg.), *Handbuch Diskriminierung* (S. 3-23). Wiesbaden/Berlin: Springer VS.
- Gardner, J.** (1998). On the ground of her sex(uality). *Oxford Journal of Legal Studies*, 18(1), 167–187.

- Gärtner, S.** (2014). Anmerkung zu BGH, Urteil vom 28.1.2014 – VI ZR 156/13. *Zeitschrift für Bank- und Kapitalmarktrecht*, 14(5), 197–199.
- Gesamtverband der Deutschen Versicherungswirtschaft e. V.** (2016). *So setzt sich der Versicherungsbeitrag für einen Pkw zusammen*. Abgerufen am 8. Juni 2018 von <https://www.gdv.de/de/themen/news/so-setzt-sich-der-versicherungsbeitrag-fuer-einen-pkw-zusammen-11804>
- Gesamtverband der Deutschen Versicherungswirtschaft e. V.** (2018). *Positionspapier: Verwendung von Algorithmen in der Versicherungswirtschaft*. Abgerufen am 8. Juni 2018 von <https://www.gdv.de/resource/blob/32438/18f083017cbfd52d21e898e5499a145b/nutzung-von-algorithmen-stellungnahme-down-load-data.pdf>
- Gesellschaft für Informatik** (2018). *Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Geslevich-Packin, N. & Lev-Aretz, Y.** (2016). On social credit and the right to be unnetworked. *Columbia Business Law Review*, 31(2), 339–425.
- Gigerenzer, G., Schlegel-Matthies, K. & Wagner, G.** (2016). *Digitale Welt und Gesundheit. eHealth und mHealth – Chancen und Risiken der Digitalisierung im Gesundheitsbereich. Gutachten des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Gigerenzer, G., Wagner, G. & Müller, K.-R.** (2018). *Wie man Licht in die Black Box wirft*. Abgerufen am 28. September 2018 von <http://www.faz.net/aktuell/feuilleton/debatten/wie-man-algorithmen-transparent-machen-kann-15652267.html>
- GKV-Spitzenverband** (2017). *Leitfaden Prävention*. Berlin: GKV-Spitzenverband.
- Goldmedia** (2017). *Ökonomischer Wert von Verbraucherdaten für Adress- und Datenhändler*. Studie im Auftrag des Bundesministeriums der Justiz und für Verbraucherschutz. Abgerufen am 05. April 2018 von https://www.goldmedia.com/fileadmin/goldmedia/2015/Studien/2017/Verbraucherdaten_BMJV/Studie_Wert_Daten_Adresshaendler_Goldmedia_BMJV_2017.pdf
- Goodfellow, I., Bengio, Y. & Courville, A.** (2016). *Deep learning*. Cambridge: MIT Press.
- Grünberger, M.** (2013). *Personale Gleichheit. Der Grundsatz der Gleichbehandlung im Zivilrecht*. Baden-Baden: Nomos Verlagsgesellschaft.
- Gründerszene.de** (2018). *Kreditech*. Abgerufen am 23. Mai 2018 von <https://www.gruenderszene.de/datenbank/unternehmen/kreditech>
- Hacker, P.** (2017). Personal data, exploitative contracts, and algorithmic fairness: autonomous vehicles meet the internet of things. *International Data Privacy Law*, 7(4), 266–286.
- Hacker, P.** (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, 55(4), 1143–1186.
- Hammersen, J. & Eisenried, U.** (2014). Ist „Redlining“ in Deutschland erlaubt? Plädoyer für eine weite Auslegung des Auskunftsanspruchs. *Zeitschrift für Datenschutz*, 4(7), 342–345.
- Hand, D. J.** (2005). Good practice in retail credit scorecard assessment. *Journal of the Operational Research Society*, 56(9), 1109–1117.
- Hand, D. J.** (2006). Classifier technology and the illusion of progress. *Statistical Science*, 21(1), 1–14.
- Hand, D. J. & Henley, W. E.** (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523–541.
- Hardy, Q.** (2012). *Just the facts. Yes, all of them*. Abgerufen am 31. Mai 2018 von <https://www.nytimes.com/2012/03/25/business/factuality-gil-elbaz-wants-to-gather-the-data-universe.html>
- Härting, N.** (2015). *Vier Thesen zur neu entbrannten Scoring-Debatte*. Abgerufen am 27. September 2018 von <https://www.cr-online.de/blog/2015/05/20/vier-thesen-zur-neu-entbrannten-scoring-debatte/>
- Heinemann, O. & Wäble, F.** (2010). Datenschutzrechtlicher Auskunftsanspruch bei Kredit scoring. Inhalt und Grenzen des Auskunftsanspruchs nach § 34 BDSG. *MultiMedia und Recht*, 13(9), 600–604.
- Heintz, B.** (2007). Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven. In A. Mennicken & H. Vollmer (Hrsg.), *Zahlenwerk. Kalkulation, Organisation und Gesellschaft* (S. 65–85). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hellgardt, A.** (2016). *Regulierung und Privatrecht. Staatliche Verhaltenssteuerung mittels Privatrecht und ihre Bedeutung für Rechtswissenschaft, Gesetzgebung und Rechtsanwendung*. Tübingen: Mohr Siebeck.
- Heuzeroth, T. & Seibel, K.** (2018). *Die gefürchtete SCHUFA-Formel gerät ins Wanken*. Abgerufen am 25. September 2018 von <https://www.welt.de/wirtschaft/article175643056/DSGVO-Neue-Datenschutzregeln-koennten-SCHUFA-zu-schaffen-machen.html>
- Hoeren, T.** (2016). Thesen zum Verhältnis von Big Data und Datenqualität. Erste Raster zum Erstellen juristischer Standards. *MultiMedia und Recht*, 19(1), 8–11.
- Hoffmann, S.** (2018). Social credit: technology-enhanced authoritarian control with global consequences. *ASPI International Cyber Policy Centre Policy Brief* 6.

- Hoffmann-Riem, W.** (1998). Informationelle Selbstbestimmung in der Informationsgesellschaft. *Archiv des öffentlichen Rechts*, 123(3), 513–540.
- Hoffmann-Riem, W.** (2016). *Innovation und Recht – Recht und Innovation: Recht im Ensemble seiner Kontexte*. Tübingen: Verlag Mohr Siebeck.
- Hofmann, J. & Bergemann, B.** (2017). *Die informierte Einwilligung: Ein Datenschutzphantom*. Abgerufen am 12. Juni 2018 von <https://netzpolitik.org/2017/die-informierte-einwilligung-ein-datenschutzphantom/>
- Hofstetter, Y.** (2016). *Das Ende der Demokratie. Wie die künstliche Intelligenz die Politik übernimmt und uns entmündigt*. Gütersloh: C. Bertelsmann Verlag.
- Holm, S.** (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Homonoff, T., O’ Brien, R. & Sussman, A.** (2017). Does access to FICO scores influence financial behavior? Evidence from a field experiment with student loan borrowers. In A. Gneezy, V. Grieskevicus & P. Williams (Hrsg.), *NA – Advances in Consumer Research*, 45 (S. 170–174). Duluth, MN: Association for Consumer Research.
- Hurley, M. & Adebayo, J.** (2016). Credit scoring in the era of big data. *Yale Journal of Law and Technology*, 18(1), 148–216.
- Ilic, D., Djulbegovic, M., Jung, J. H., Hwang, E. C., Zhou, Q. & Cleves, A.** (2018). Prostate cancer screening with prostate-specific antigen (psa) test: A systematic review and meta-analysis. *BMJ*, 362, k3519.
- Infas** (2018). *Verbraucher-Scoring: Repräsentativbefragung zur Akzeptanz und Kenntnis über (neuartige) Scoring-Methoden. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- James, G.; Witten, D., Hastie, T. & Tibshirani, R.** (2013). *An introduction to statistical learning*. New York: Springer
- Jenny, M. A., Pachur, T., Williams, S. L., Becker, E. & Margraf, J.** (2013). Simple rules for detecting depression. *Journal of Applied Research in Memory and Cognition*, 2(3), 149–157.
- Jentsch, N.** (2016). Kreditwürdigkeitsanalysen im Zeitalter von Big Data: Innovation oder Revolution? *Wirtschaftsdienst*, 96(9), 644–647.
- Jentsch, N.** (2018). *Hysterie und Hype um KI*. Abgerufen am 17. August 2018 von <http://www.fr.de/kultur/netz-tv-kritik-medien/netz/kuenstliche-intelligenz-hysterie-und-hype-um-ki-a-1532908>
- Ji, S., Li, W., Srivatsa, M., He, J. & Beyah, R.** (2016). General graph data de-anonymization: From mobility traces to social networks. *ACM Transactions on Information and System Security*, 18(4), 12:1–12:29.
- Joeres, A.** (2018). *Parcoursup – das außerirdische Universitätsauswahlssystem der französischen Regierung*. Abgerufen am 15. Juni 2018 von <https://algorithmenethik.de/2018/05/30/parcoursup-das-ausserirdische-universitaetsauswahlssystem-der-franzoesischen-regierung/>
- Jordan, S., von der Lippe, E., Starker, A., Hoebel, J. & Franke, A.** (2015). Einflussfaktoren für die Teilnahme an Bonusprogrammen der gesetzlichen Krankenversicherung. Ergebnisse der Studie „Gesundheit in Deutschland aktuell“. *Das Gesundheitswesen* 2015, 77(11), 861–868.
- Junge, B.** (2012). *Wer hat meine Daten?* Abgerufen am 31. Mai 2018 von <https://www.tagesspiegel.de/medien/digitale-welt/vernetzung-wer-hat-meine-daten/7223236-all.html>
- Just, N. & Latzer, M.** (2016). Governance by algorithms: Reality construction by algorithmic selection on the internet. *Media, Culture & Society*, 39(2), 238–258.
- Kalis, B.** (2018). Leistungs- und Vertragsrecht (§ 44). In H. Sodan (Hrsg.), *Handbuch des Krankenversicherungsrechts*. München: Verlag C. H. Beck.
- Kamp, M. & Weichert, T.** (2005). *Scoringssysteme zur Beurteilung der Kreditwürdigkeit. Chancen und Risiken für Verbraucher*. Kiel: Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD).
- Kerler, W., Köppen, U., Schnuck, O. & Zierer, M.** (2018). *Lücken im Prüfsystem. Auskunfteien beauftragen Gutachten selbst*. Abgerufen am 16. Mai 2018 von www.br.de/nachrichten/luecken-im-pruefsystem-von-auskunfteien-100.html
- Kettner, S.-E., Thorun, C. & Vetter, M.** (2018). *Wege zur besseren Informiertheit. Verhaltenswissenschaftliche Ergebnisse zur Wirksamkeit des One-Pager-Ansatzes und weiterer Lösungsansätze im Datenschutz*. Berlin: ConPolicy GmbH. Institut für Verbraucherpolitik.
- Kim, P.** (2017). Data-driven discrimination at work. *William & Mary Law Review*, 48(6), 857–936.
- Kingreen, T.** (2003). *Das Sozialstaatsprinzip im europäischen Verfassungsverbund. Gemeinschaftsrechtliche Einflüsse auf das deutsche Recht der gesetzlichen Krankenversicherung*. Tübingen: Verlag Mohr Siebeck.
- Kleinberg, J., Mullainathan, S. & Raghavan, M.** (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.
- Klement, J. H.** (2006). *Verantwortung. Funktion und Legitimation eines Begriffs im Öffentlichen Recht*. Tübingen: Mohr Siebeck.
- Klimke, D.** (2015). Telematik-Tarife in der Kfz-Versicherung. *Recht und Schaden*, 42(5), 217–225.
- Kloepfer, M. & Kutzschbach, G.** (1998). SCHUFA und Datenschutzrecht. *MultiMedia und Recht*, 1(12), 650–659.

- Koch, R.** (2017). Kommentierung der AKB 2015 (außer Kfz-Unfallversicherung). In H. Baumann, R. M. Beckmann, K. Johannsen, R. Johannsen & R. Koch (Hrsg.), *Versicherungsvertragsgesetz. Großkommentar* (Bd. 12). Berlin: De Gruyter.
- Kolany-Raiser, B.** (2016). Der Verbraucher als Datenlieferant: Rechtliche Aspekte von „smarten“ Produkten. In C. Bala & W. Schuldzinsk (Hrsg.), *Schöne neue Verbraucherwelt? Big Data, Scoring und das Internet der Dinge* (S. 47–66). Düsseldorf: Verbraucherzentrale NRW.
- Korczak, D. & Wilken, M.** (2008). *Scoring im Praxistest: Aussagekraft und Anwendung von Scoringverfahren in der Kreditvergabe und Schlussfolgerungen*. München: Verbraucherzentrale Bundesverband e. V.
- Kosinski, M., Stillwell, D. & Graepel, T.** (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 5802–5805.
- Kosinski, M., Stillwell, D., Kohli, P., Bachrach, Y. & Graepel, T.** (2012). *Personality and website choice*. In: *ACM Conference on Web Sciences (Hrsg.), ACM Web Sciences 2012*. Abgerufen am 29. Mai 2018 von https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/person_WebSci_final.pdf
- Kostka, G.** (2018). *China's social credit systems and public opinion: Explaining high levels of approval*. Abgerufen am 27. September 2018 von <https://ssrn.com/abstract=3215138>
- Kraft, M. & Hering, J.** (2017). Potenziale von Telematik-Tarifen in der Kfz-Versicherung in Deutschland. *Zeitschrift für die gesamte Versicherungswissenschaft*, 106(5), 503–524.
- Kraftfahrt-Bundesamt** (2017). *Rund um den Punkt – kurz gefasst*. Abgerufen am 1. Oktober 2018 von https://www.kba.de/SharedDocs/Publikationen/DE/Presse/rund_um_den_Punkt_kurz_gefasst_faltblatt_pdf.pdf?__blob=publicationFile&v=6
- Lazer, D., Kennedy, R., King, G. & Vespignani, A.** (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205.
- Lessmann, S., Baesens, B., Seow, H. V. & Thomas, L. C.** (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136.
- Levinger, B., Benton, M. & Meier, S.** (2011). The cost of not knowing the score: Self-estimated credit scores and financial outcomes. *Journal of Family and Economic Issues*, 32(4), 566–585.
- Lintorf, K.** (2012). Zur individuellen und gesellschaftlichen Bedeutung von Schulnoten. In K. Lintorf, *Wie vorhersehbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale* (S. 19–35). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lipton, Z. Z.** (2017). The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*.
- Litman, T.** (2005). Pay-as-you-drive pricing and insurance regulatory objectives. *Journal of Insurance Regulation*, 23(3).
- Looschelders, D.** (2015). Fragmentierung der Kollektive in der Privatversicherung – juristische Implikationen. *Zeitschrift für die gesamte Versicherungswissenschaft*, 104(4), S. 481–499.
- Lowry, S. & Macpherson, G.** (1988). A blot on the profession. *British Medical Journal (Clinical Research ed.)*, 296(6623), 657.
- Lunt, C.** (2014). *Patentnr. 8,806,584 B2*. Washington, DC: U.S. Patent and Trademark Office.
- Lupton, D.** (2016). *The quantified self*. Cambridge: Polity.
- Lüttringhaus, J. D.** (2018). Mehr Freiheit wagen im Versicherungsrecht durch daten- und risikoadjustierte Versicherungstarife – „Pay-as-you-drive“- und „Pay-as-you-live“- und „Smart-Home“-Tarife als Herausforderung für das Versicherungsvertragsrecht. *Max Planck Private Law Research Paper No. 17/18*.
- Marschall, T., Morawitzky, D., Reutter, M., Schwartz, R. & Baars, H.** (2015). Netzwerkanalysen für die Betrugserkennung im Online-Handel. In O. Thomas & F. Teuteberg (Hrsg.), *Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015)* (S. 1859–1873). Osnabrück: Universität Osnabrück.
- Martini, M.** (2017). Algorithmen als Herausforderung für die Rechtsordnung. *Juristenzeitung*, 72(21), 1017–1025.
- Martini, M.** (2018). Kommentierung von Art. 22 DSGVO u. a. In B. P. Paal, B. Paal & D. Pauly (Hrsg.), *Beck'sche Kompakt-Kommentare. Datenschutz-Grundverordnung und Bundesdatenschutzgesetz*. München: Verlag C. H. Beck.
- Mathis, K.** (2017). *Nachhaltige Entwicklung und Generationengerechtigkeit. Eine interdisziplinäre Studie aus rechtlicher, ökonomischer und philosophischer Sicht*. Tübingen: Mohr Siebeck.
- Mau, S.** (2017). *Das metrische Wir. Über die Quantifizierung des Sozialen*. Berlin: Suhrkamp Verlag.
- Maxwell, W. J.** (2015). Principles-based regulation of personal data: The case of 'fair processing'. *International Data Privacy Law*, 5(3), 205–216.
- McLaughlin, C.** (2013). *Die Besserwisser*. Abgerufen am 1. Juni 2018 von <https://www.zeit.de/2013/28/acxiom>
- Mengden, M.** (2018). *Zugangsfreiheit und Aufmerksamkeitsregulierung. Zur Reichweite des Gebots der Gewährleistung freier Meinungsbildung am Beispiel algorithmengestützter Zugangsdienste im Internet*. Tübingen: Mohr Siebeck.
- Merry, S. E.** (2016). *The seductions of quantification. Measuring human rights, gender violence, and sex trafficking*. Chicago, IL, USA/London: The University of Chicago Press.

- Merry, S. E., Davis, K. E. & Kingsbury, B. (Hrsg.). (2015). *The quiet power of indicators. Measuring governance, corruption, and rule of law*. Cambridge, UK: Cambridge University Press.
- Metz, R. (2009). Scoring: Licht im Tunnel. *Verbraucher und Recht*, 24(11), 403–408.
- Meyer, K. (2017). *Theorien der Intersektionalität zur Einführung*. Hamburg: Junius Verlag.
- Micklitz, H.-W. (2014). Unfair terms in consumer contracts. In N. Reich, H.-W. Micklitz, P. Rott & K. Tonner (Hrsg.), *European Consumer Law* (S. 125–164). Cambridge/Antwerpen/Portland: Intersentia.
- Micklitz, H.-W. (2017). Ungeheuerliche Neuigkeiten? *Verbraucher und Recht*, 32(2), 43–46.
- Micklitz, H.-W., Oehler, A., Piorkowsky, M.-B., Reisch, L. A. & Strünck, C. (2010). *Der vertrauende, der verletzte oder der verantwortungsvolle Verbraucher? Plädoyer für eine differenzierte Strategie in der Verbraucherpolitik. Stellungnahme des wissenschaftlichen Beirats Verbraucher- und Ernährungspolitik beim BMELV*. Berlin.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
- Möller, J. & Florax, B.-C. (2003). Datenschutzrechtliche Unbedenklichkeit des Scoring von Kreditrisiken? *Neue Juristische Wochenschrift*, 56(38), 2724–2726.
- Montavon, G., Samek, W. & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15.
- Moorcraft, B. (2018). *LGA launches 'fun and simple' selfiequote tool*. Abgerufen am 23. Mai 2018 von <https://www.insurancebusinessmag.com/us/news/technology/lga-launches-fun-and-simple-selfiequote-tool-90113.aspx>
- Moos, F. & Rothkegel, T. (2016). Nutzung von Scoring-Diensten im Online-Versandhandel. Scoring-Verfahren im Spannungsfeld von BDSG, AGG und DS-GVO. *Zeitschrift für Datenschutz*, 6(12), 561–658.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton, N.J., USA: Princeton University Press.
- Müller-Peters, H. & Wagner, F. (2017). *Geschäft oder Gewissen? Vom Auszug aus der Versicherung aus der Solidargemeinschaft*. Goslar: Studiengesellschaft für verbrauchergerichtetes Versichern e. V.
- Narayanan, A. & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets (How to break anonymity of the netflix prize dataset). *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, 111–125.
- Narayanan, A. & Shmatikov, V. (2009). De-anonymizing social networks. *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, 173–187.
- Nieber, K. (2014). Bedeutung von geschlechtsspezifischen Unterschieden in der Arzneimitteltherapie. In M. Gadebusch Bondio & E. Katsari (Hrsg.), *Gender-Medizin: Krankheit und Geschlecht in Zeiten der individualisierten Medizin* (S. 69–85). Bielefeld: transcript Verlag.
- Nilsson, N. (2009). *The quest for artificial intelligence. A history of ideas and achievements*. New York: Cambridge University Press.
- Norddeutscher Rundfunk (Hrsg.) (2016). *Nackt im Netz: Millionen Nutzer ausgespäht*. Abgerufen am 12. Juni 2018 von <https://www.ndr.de/nachrichten/netzwelt/Nackt-im-Netz-Millionen-Nutzer-ausgespaht,nacktimnetz100.html>
- OECD (2013). Exploring the economics of personal data: A survey of methodologies for measuring monetary value. *OECD Digital Economy Papers*, No. 220. Paris: OECD Publishing.
- Oehler, A. (2017). Grundsätze ordnungsgemäßer Bewertung durch Scoring (The basics of generally accepted scoring principles). *Wirtschaftsdienst*, 97(10), 784–751.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.
- OpenSCHUFA (2018). *OpenSCHUFA – wir knacken die SCHUFA*. Abgerufen am 25. September 2018 von www.startnext.com/openSCHUFA
- Overbeck, A. (2016). *Datenschutz und Verbraucherschutz bei Bonitätsprüfungen durch Wirtschaftsauskunfteien mittels Scoring*. Münster: MV Wissenschaft.
- Paal, B. P. & Hennemann, M. (2018). Kommentierung des Art. 13 DSGVO. In B. P. Paal & D. A. Pauly (Hrsg.), *Beck’sche Kompakt-Kommentare. Datenschutz-Grundverordnung und Bundesdatenschutzgesetz*. München: Verlag C. H. Beck.
- Palmethofer, W., Semsrott, A. & Alberts, A. (2016). *Der Wert persönlicher Daten: Ist Datenhandel der bessere Datenschutz? Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen (SVRV).
- Paraschiakos, C. (2017). *Bankenaufsicht zwischen Risikoverwaltung und Marktbegleitung. Eine rechtsdogmatische und verwaltungswissenschaftliche Untersuchung bankenaufsichtsrechtlicher Unsicherheitsbewältigung am Beispiel der Eigenmittelregulierung*. Tübingen: Mohr Siebeck.
- Passig, K. (2017). Fünfzig Jahre Black Box. *Merkur – Deutsche Zeitschrift für Europäisches Denken*, 71(823), 16–30.
- Peppet, S. R. (2011). Unraveling privacy: The personal prospectus and the threat of a full-disclosure future. *Northwestern University Law Review*, 105(3), 1153–1204.
- Perel, M. & Elkin-Koren, N. (2017). Black box tinkering: Beyond disclosure in algorithmic enforcement. *Florida Law Review*, 69, 181–221.

- Petersen, S.** (2000). *Grenzen des Verrechtlichungsgebots im Datenschutz*. Hamburg: LIT Verlag.
- Petri, T. B.** (2003). Sind Score-Werte rechtswidrig? *Datenschutz und Datensicherheit*, 27(2), 631–636.
- Phillips, N. D., Neth, H., Woike, J. K. & Gaissmaier, W.** (2017). FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making*, 12(4), 344–368.
- Picker, E.** (2008). Das Allgemeine Gleichbehandlungsgesetz – Gebot allgemeiner Gleichbehandlung? In P. Hanau, J. T. Thau & H. P. Westermann (Hrsg.), *Gegen den Strich. Festschrift für Klaus Adomeit* (S. 541–562). Köln: Luchterhand.
- Piltz, G. & Holländer, C.** (2008). Scoring als modernes Orakel von Delphi. Wie die geplante Änderung des Bundesdatenschutzgesetzes (BDSG) Transparenz und Rechtssicherheit schaffen will. *Zeitschrift für Rechtspolitik*, 41(5), 143–146.
- Podszun, R., Busch, C. & Henning-Bodewig, F.** (2018). *Behördliche Durchsetzung des Verbraucherrechts? Darstellung und Systematisierung von Möglichkeiten und Defiziten der privaten Durchsetzung des Verbraucherschutzes sowie Einbeziehung der Kartellbehörden zu dessen Durchsetzung*. Berlin: Bundesministerium für Wirtschaft und Energie.
- Pollmann, M. & Kipker, D. K.** (2016). Eingeschränkte Selbstbestimmung im Onlineverkehr: Stärkung der Einwilligungserklärung durch Einführung vorformulierter Datenschutzbestimmungen. *Informatik 2016*.
- Porter, T. M.** (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton, N.J., USA: Princeton University Press.
- Pötters, S.** (2018). Kommentierung des Art. 5 DSGVO. In P. Gola (Hrsg.), *DS-GVO. Datenschutz-Grundverordnung VO (EU) 2016/679. Kommentar*. München: Verlag C. H. Beck.
- PricewaterhouseCoopers** (2018). *Ist Deutschland bereit für Social Scoring*. Abgerufen am 1. Oktober 2018 von <https://www.pwc.de/de/finanzdienstleistungen/studie-ist-deutschland-bereit-fuer-social-scoring.pdf>
- Rebitschek, F., Gross, C., Brümmer, M., Gigerenzer, G. & Wagner, G. G.** (2018). *Dokumentation einer empirischen Pilot-Studie zum Wissen über und zur Bewertung von Verbraucher-Scoring. Veröffentlichungen des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Ribeiro, M. T., Singh, S. & Guestrin, C.** (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D. et al.** (2015). *Strategies and best practies for data literacy education*. Abgerufen am 3. September 2018 von http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf
- Rixecker, R.** (2016). Kommentierung des § 213 VVG u. a. In T. Langheid & R. Rixecker (Hrsg.), *Versicherungsvertragsgesetz mit Einführungsgesetz und VVG-Infomationspflichtenverordnung. Kommentar*. München: Verlag C. H. Beck.
- Rona-Tas, A. & Hiss, S.** (2011). Forecasting as valuation: the role of ratings and predictions in the subprime mortgage crisis in the United States. In J. Beckert & P. Aspers (Hrsg.), *The Worth of Goods. Valutation and Pricing in the Economy* (S. 223–246). Oxford/New York: Oxford University Press.
- Ropohl, G.** (2009). *Allgemeine Technologie. Eine Systemtheorie der Technik*. Karlsruhe: Universitätsverlag Karlsruhe.
- Rossi, F.** (2018). „Künstliche Intelligenz muss fair sein“ – Interview, geführt von Corinna Budras. *Frankfurter Allgemeine Zeitung vom 24. Juli 2018*, 17.
- Roßnagel, A.** (2018). Datenschutzgrundsätze – unverbindliches Programm oder verbindliches Recht? Bedeutung der Grundsätze für die datenschutzrechtliche Praxis. *Zeitschrift für Datenschutz*, 7(8), 339–344.
- Rottenburg, R., Merry, S. E., Park, S.-J. & Mugler, J. (Hrsg.)** (2015). *The world of indicators. The making of governmental knowledge through quantification*. Cambridge, UK: Cambridge University Press.
- Rubinfeld, D. L. & Gal, M. S.** (2017). Access barriers to big data. *Arizona Law Review*, 59, 339–381.
- Rudkowski, L.** (2017). Vertragsrechtliche Anforderungen an die Gestaltung von Self-Tracking-Tarifen in der Privatversicherung. *Zeitschrift für die gesamte Versicherungswissenschaft*, 106(5), 453–502.
- Russell, S. J. & Norvig, P.** (2010). *Artificial intelligence: A modern approach*. Upper Saddle River: Prentice Hall.
- Saetnan, A. R., Schneider, I. & Green, N.** (2018). *The politics of big data. Big data, big brother?* Oxford: Routledge.
- Samek, W., Wiegand, T. & Müller, K.-R.** (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries*, 1, 1–10.
- Schaar, P.** (2008). Geodaten – Werkzeug der Verbraucherdiskriminierung? *Kommunikation & Recht*, 9, 1.
- Schantz, P.** (2016). Die Datenschutz-Grundverordnung – Beginn einer neuen Zeitrechnung im Datenschutzrecht. *Neue Juristische Wochenschrift*, 69(26), 1841–1847.
- Schantz, P. & Wolff, H. A.** (2017). *Das neue Datenschutzrecht. Datenschutz-Grundverordnung und Bundesdatenschutzgesetz in der Praxis*. München: Verlag C.H. Beck.
- Schiek, D.** (2007). Kommentierung des § 3 AGG. In D. Schiek (Hrsg.), *Allgemeines Gleichbehandlungsgesetz. Ein Kommentar aus europäischer Perspektive*. Baden-Baden: Nomos Verlagsgesellschaft.
- Schild, H.-H.** (2018). Kommentar zu Art. 4 DSGVO. In S. Brink & H.-A. Wolff (Hrsg.), *Beck'scher Online-Kommentar Datenschutzrecht*. München: Verlag C. H. Beck.

- Schlaudt, O.** (2018). *Die politischen Zahlen. Über Quantifizierung im Neoliberalismus*. Frankfurt a.M.: Vittorio Klostermann.
- Schmidt-Wudy, F.** (2018). Kommentierung des Art. 15 DSGVO. In S. Brink & H. A. Wolff (Hrsg.), *Beck'scher Online-Kommentar Datenschutzrecht*. München: Verlag C. H. Beck.
- Schneider, K.** (2018). *So greift „Open SCHUFA“ die Auskunft an*. Abgerufen am 25. September 2018 von Handelsblatt Online: www.handelsblatt.com/finanzen/steuern-recht/recht/initiative-will-algorithmus-knacken-so-greift-open-SCHUFA-die-auskunft-an/20973246.html
- Schröder, M. & Taeger, J. (Hrsg.)** (2014). *Scoring im Fokus: ökonomische Bedeutung und rechtliche Rahmenbedingungen im internationalen Vergleich*. Oldenburg: BIS-Verlag der Carl von Ossietzky Universität Oldenburg.
- Schröder, M., Lang, G., Lerbs, O. & Radev, D.** (2014). Ökonomische Bedeutung und Funktionsweise von Credit Scoring. In M. Schröder & J. Taeger (Hrsg.), *Scoring im Fokus: ökonomische Bedeutung und rechtliche Rahmenbedingungen im internationalen Vergleich*. Oldenburg: BIS-Verlag der Carl von Ossietzky Universität.
- SCHUFA Holding AG** (2018a). „OpenSCHUFA“-Kampagne irreführend und gegen Sicherheit und Datenschutz in Deutschland. Abgerufen am 25. September 2018 von www.SCHUFA.de/de/ueber-uns/themenportal/detailseite/themenportal-detailseite.9856.jsp
- SCHUFA Holding AG** (2018b). *SCHUFA Ombudsmann. Tätigkeitsbericht 2017*. Abgerufen am 4. September 2018 von https://SCHUFA-ombudsmann.de/media/misc/Taetigkeitsbericht_2017_v3.pdf
- Schüffner, M. & Franck, P.** (2018). Versicherungsaufsichtsrecht (§ 47). In H. Sodan (Hrsg.), *Handbuch des Krankenversicherungsrechts*. München: Verlag C. H. Beck.
- Schulte am Hülse, U. & Timm, M.** (2014). Anmerkung zu BGH, Urteil vom 28.1.2014 – VI ZR 156/13. *Neue Juristische Wochenschrift*, 67(17), 1238–1239.
- Schulz, S.** (2017). Kommentierung zu Art. 22 DSGVO. In P. Gola (Hrsg.), *Datenschutz-Grundverordnung. VO (EU) 2016/679. Kommentar*. München: Verlag C.H. Beck.
- Schulz, T., Müller, M. & Rosenbach, M.** (2013). *Die Daten-Bank*. Abgerufen am 23. Mai 2018 von <http://www.spiegel.de/netzwelt/netzpolitik/big-data-daten-bank-a-899538.html>
- Schulz-Schaeffer, I.** (2007). *Zugeschriebene Handlungen. Ein Beitrag zur Theorie sozialen Handelns*. Weilerswist: Velbrück Verlag.
- Schumann, D.** (2017). *Pay as you drive: Die rechtliche Zulässigkeit von Telematik-Tarifen im Privatkundensegment der Kraftfahrzeug-Haftpflichtversicherung*. Karlsruhe: Verlag Versicherungswirtschaft.
- Seibel, K.** (2015). *Gegen Kreditech ist die SCHUFA ein Schuljunge*. Abgerufen am 23. Mai 2018 von <https://www.welt.de/finanzen/verbraucher/article139671014/Gegen-Kreditech-ist-die-SCHUFA-ein-Schuljunge.htm>
- Seibt, P.** (2018). *Wie ich bei der SCHUFA zum „deutlich erhöhten Risiko“ wurde*. Abgerufen am 26. September 2018 von <http://www.spiegel.de/wirtschaft/service/SCHUFA-wie-ich-zum-deutlich-erhoehten-risiko-wurde-a-1193506.html>
- Selbst, A. D. & Barocas, S.** (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, im Erscheinen (ssrn.com/abstract=3126971 or dx.doi.org/10.2139/ssrn.3126971).
- Selbst, A. D. & Powles, J.** (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), 233–242.
- Selke, S.** (2014). *Lifelogging. Wie die digitale Selbstvermessung unsere Gesellschaft verändert*. Berlin: ECON.
- Society for Human Resource Management** (2012). *Background checking – The use of credit background checks in hiring decisions*. Abgerufen am 26. Juni 2018 von <https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/creditbackgroundchecks.aspx>
- Spindler, G.** (2011). *Unternehmensorganisationspflichten. Zivilrechtliche und öffentlich-rechtliche Regelungskonzepte*. Göttingen: Universitätsverlag Göttingen.
- Spindler, G., Thorun, C. & Wittmann, J.** (2017). *Rechtsdurchsetzung im Verbraucherdatenschutz. Bestandsaufnahme und Handlungsempfehlungen*. Berlin: Friedrich-Ebert-Stiftung.
- Srivatsa, M. & Hicks, M.** (2012). Deanonymizing mobility traces: Using social networks as a side-channel. *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, 628–637.
- Stiftung Warentest** (2014). *Telematik-Tarife in der Auto-Versicherung: Spion fährt mit*. Abgerufen am 28. September 2018 von <https://www.test.de/Telematik-Tarife-in-der-Autoversicherung-Spion-faehrt-mit-4692642-0/>
- Strathern, M.** (1997). ‘Improving ratings’: Audit in the british university system. *European Review*, 5, 305–321.
- Sunstein, C. R. & Reisch, L. A.** (2014). Automatically green: Behavioral economics and environmental protection. *Harvard Environmental Law Review*, 38, 127–158.
- SVRV** (2016). *Verbraucherrecht 2.0. Verbraucher in der digitalen Welt. Gutachten des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- SVRV** (2017). Sachverständigenrat für Verbraucherfragen: Lösungsoptionen. In H.-W. Micklitz, L. A. Reisch, G. Joost & H. Zander-Hayat (Hrsg.), *Verbraucherrecht 2.0 – Verbraucher in der digitalen Welt* (S. 9–43). Baden-Baden: Nomos Verlagsgesellschaft.

- SVRV (2017a). *Digitale Souveränität. Gutachten des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Sweeney, L. (2000). *Simple demographics often identify people uniquely (Data Privacy Working Paper 3)*. Pittsburgh: Carnegie Mellon University.
- Taeger, J. (2014). Anmerkung zu BGH, Urteil vom 28.1.2014 – VI ZR 156/13. *MultiMedia und Recht*, 17(7), 492–494.
- Taeger, J. (2016). Scoring in Deutschland nach der EU-Datenschutzgrundverordnung. *Zeitschrift für Rechtspolitik*, 49(3), 72–75.
- Tamm, M. (2011). *Verbraucherschutzrecht. Europäisierung und Materialisierung des deutschen Zivilrechts und die Herausbildung eines Verbraucherschutzprinzips*. Tübingen: Mohr Siebeck.
- Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. Yale: Yale University Press.
- Thomas, L., Crook, J. & Edelman, D. (2017). *Credit scoring and its applications, second edition*. Philadelphia: Society for Industrial and Applied Mathematics.
- Townley, C., Morrison, E. & Yeung, K. (2017). Big data and personalised price discrimination in EU competition law. *King's College London Law School Research Paper No. 2017–38*.
- Trove. (2018). *Predictive data science sharpens hourly load forecasting*. Abgerufen am 25. Juni 2018 von <http://trovedata.com/results/case-study/predictive-data-science-sharpens-hourly-load-forecasting>
- Ullrich, N.-D. (2018). Kommentierung des § 65a SGB V. In J. Barchtold, S. Huster & M. Rehborn (Hrsg.), *Gesundheitsrecht*. Baden-Baden: Nomos Verlagsgesellschaft.
- Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein & GP Forschungsgruppe (2014). *Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen*. Kiel, München: Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein; GP Forschungsgruppe.
- Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (2005). *Scoringssysteme zur Beurteilung der Kreditwürdigkeit. Chancen und Risiken für Verbraucher*. Bonn: Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft.
- Veil, W. (2018). Kommentierungen von Art. 4 Nr. 4 DSGVO u. a. In S. Gierschmann, K. Schlender, R. Stenzel & W. Veil (Hrsg.), *Kommentar Datenschutz-Grundverordnung*. Bonn: Bundesanzeiger Verlag.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229.
- Verbraucherzentrale Bayern (2016). *Pay as you drive – Telematik-Tarife in der Kfz-Versicherung*. Abgerufen am 28. September 2018 von https://www.vis.bayern.de/finanzen_versicherungen/versicherungen/kfztelematik.htm#was
- Verbraucherzentrale Bremen (2016). *15 Fragen und Antworten zur SCHUFA mit einem Musterbrief „Wie wehre ich mich gegen falsche Einträge bei der SCHUFA“*. Abgerufen am 27. September 2018 von <https://www.verbraucherzentrale-bremen.de/wissen/geld-versicherungen/kredit-schulden-insolvenz/15-fragen-und-antworten-zur-SCHUFA-mit-einem-musterbrief-wie-wehre-ich-mich-gegen-falsche-eintraege-bei-der-SCHUFA-11102>
- Verbraucherzentrale Bundesverband e. V. (2017). *Algorithmenbasierte Entscheidungsprozesse. Thesenpapier des vzbv*. Berlin: Verbraucherzentrale Bundesverband e. V.
- Verbraucherzentrale Niedersachsen (2015). *Scoring – Punkte zählen für Ihre Kreditwürdigkeit. Auskünfte einholen bei SCHUFA & Co*. Abgerufen am 27. September 2018 von <https://bit.ly/2On5HWn>
- Verbraucherzentrale Nordrhein-Westfalen (2015). *Bonusprogramme der Gesetzlichen Krankenkassen – Anreize für gesundheitsbewusstes Verhalten oder Prämienzahlung für Gesunde? Untersuchung der Verbraucherzentrale NRW*. Abgerufen am 9. August 2018 von https://www.verbraucherzentrale.nrw/sites/default/files/migration_files/media236794A.pdf
- Verbraucherzentrale Nordrhein-Westfalen (2018). *Scoring mit Kundendaten: So verlangen Sie Auskunft bei SCHUFA & Co*. Abgerufen am 27. September 2018 von <https://www.verbraucherzentrale.nrw/wissen/digitale-welt/datenschutz/scoring-mit-kundendaten-so-verlangen-sie-auskunft-bei-SCHUFA-co-12756>
- Versium Analytics Inc. (2018). *Custom predictive scores solve business challenges*. Abgerufen am 14. Mai 2018 von <https://versium.com/predictive-scores/>
- Voit, W. (2018). Kommentierung von § 203 VVG u. a. In J. Prölss & A. Martin (Hrsg.), *Versicherungsvertrags-gesetz*. München: Verlag C. H. Beck.
- von Lewinski, K. (2018). Kommentierung von § 28b BDSG (a.F.). In H.-A. Wolff & S. Brink (Hrsg.), *Beck'scher Online-Kommentar Datenschutzrecht*.
- von Lewinski, K. & Pohl, D. (2018). Auskunftfeien nach der europäischen Datenschutzreform. Brüche und Kontinuitäten der Rechtslage. *Zeitschrift für Datenschutz*, 7(1), 17–23.
- Vormbusch, U. (2012). *Die Herrschaft der Zahlen. Zur Kalkulation des Sozialen in der kapitalistischen Moderne*. Frankfurt am Main: Campus Verlag.
- Wachter, S., Mittelstadt, B. & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*. 7(2), 76–99.

- Wagner, G. G.** (2018). Die fragwürdige Rolle gesetzliche festgelegter sozial- und wirtschaftspolitischer Zielwerte. In N. Kosturkova & J. Rieger (Hrsg.), *Ordnungspolitisch stets auf Kurs* (S. 31–42). München: Verlag Franz Vahlen.
- Wagner-Braun, M.** (2002). *Zur Bedeutung berufsständischer Krankenkassen innerhalb der privaten Krankenversicherung in Deutschland bis zum Zweiten Weltkrieg*. Stuttgart: Franz-Steiner Verlag.
- Weichert, T.** (Hrsg.) (2018). *Big Data im Gesundheitsbereich. Assessing Big Data* (abida).
- Weidner, W. & Transchel, F. W.** (2015). Aktuarielle Besonderheiten bei der Kalkulation von Telematik-Tarifen in der Kfz-Versicherung. *Zeitschrift für die gesamte Versicherungswissenschaft*, 104(5), 595–614.
- Weingart, P. & Wagner, G. G.** (2015). *Wissenschaftliche Politikberatung im Praxistest*. Velbrück Wissenschaft.
- Welchering, P.** (2017). *Wenn Software über Leben und Tod entscheidet*. Abgerufen am 23. Mai 2018 von <https://www.zdf.de/nachrichten/heute/software-soll-ueberleben-und-tod-entscheiden-100.html>
- Wiebe, A.** (2002). *Die elektronische Willenserklärung. Kommunikationstheoretische und rechtsdogmatische Grundlagen des elektronischen Geschäftsverkehrs*. Tübingen: Mohr Siebeck.
- Wiegard, R.-B. & Breitner, M.** (2017). Smart services in healthcare: A risk-benefit-analysis of pay-as-you-live services from customer perspective in Germany. *Electronic Markets*, 1–17.
- Wischmeyer, T.** (2018). Regulierung intelligenter Systeme. *Archiv des öffentlichen Rechts*, 143(1), 1–66.
- Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J.** (2016). *Data mining: Practical machine learning tools and techniques*. Cambridge, MA: Morgan Kaufmann.
- WP 29** (2013). *Article 29 Data Protection Working Party. Advice paper on essential elements of a definition an a provision on profiling within the EU general data protection regulation*. Brussels: Article 29 Data Protection Working Party.
- WP 29** (2018). *Article 29 Data Protection Working Party. Guidelines on transparency under regulation 2016/679*. Brussels: Article 29 Data Protection Working Party.
- Wu, T.** (2016). *The attention merchants: The epic struggle to get inside our heads*. New York, USA: Verlag Alfred A. Knopf.
- Wundenberg, M.** (2012). *Compliance und die prinzipiengeleitete Aufsicht über Bankengruppen*. Tübingen: Mohr Siebeck.
- Wurmnest, W.** (2016). Kommentierung von § 307 BGB u. a. In F. J. Säcker, R. Rixecker, H. Oetker & B. Limperg (Hrsg.), *Münchener Kommentar zum Bürgerlichen Gesetzbuch, Band 2: Schuldrecht. Allgemeiner Teil*. München: Verlag C. H. Beck.
- Zander-Hayat, H., Domurath, I. & Gross, C.** (2016). *Personalisierte Preise. Veröffentlichungen des Sachverständigenrats für Verbraucherfragen*. Berlin: Sachverständigenrat für Verbraucherfragen.
- Zander-Hayat, H., Reisch, L. A. & Steffen, C.** (2016). Personalisierte Preise – Eine verbraucherpolitische Einordnung. *Verbraucher und Recht*, 31(11), 403–409.
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K. W.** (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *arXiv preprint arXiv:1707.09457*.
- Zweig, K. & Krafft, T. D.** (2018). Fairness und Qualität algorithmischer Entscheidungen. In R. M. Kar, B. Thapa & P. Parycek (Hrsg.), *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft* (S. 204–227). Berlin: Kompetenzzentrum Öffentliche IT.



Sachverständigenrat für Verbraucherfragen

Der Sachverständigenrat für Verbraucherfragen ist ein Beratungsgremium des Bundesministeriums der Justiz und für Verbraucherschutz (BMJV). Er wurde im November 2014 vom Bundesminister der Justiz und für Verbraucherschutz, Heiko Maas, eingerichtet. Der Sachverständigenrat für Verbraucherfragen soll auf der Basis wissenschaftlicher Erkenntnisse und unter Berücksichtigung der Erfahrungen aus der Praxis das Bundesministerium der Justiz und für Verbraucherschutz bei der Gestaltung der Verbraucherpolitik unterstützen.

Der Sachverständigenrat ist unabhängig und hat seinen Sitz in Berlin.

Vorsitzende des Sachverständigenrats ist Prof. Dr. Lucia Reisch.