



Einschätzung der Qualität digitaler Gesundheitsangebote: Wie können informierte Entscheidungen gefördert werden?

Einleitung

Im Jahr 2016 wurden im Gutachten „Digitale Welt und Gesundheit“ des Sachverständigenrats für Verbraucherfragen drei notwendige Voraussetzungen für eine nutzbringende Digitalisierung des Gesundheitswesens identifiziert [1]: weniger Defensivmedizin (Ärzte treffen Entscheidungen gegen das Wohl des Patienten, um sich vor Haftungsrisiken zu schützen [2]), gesicherte Solidarität trotz personalisierter Belohnungssysteme für gesundheitsförderndes Verhalten [3] und die Risikokompetenz der Nutzer¹.

Die Risikokompetenz ist aus drei Gründen im Zusammenhang mit der Digitalisierung besonders bedeutsam [1]:

1. Die angestrebte Partizipation der Patienten in der Versorgung erfordert eine transparente Aufklärung zu konkreten medizinischen Angeboten.
2. Umfangreiche digitale Entwicklungen im Gesundheitswesen erfordern einen evidenzbasierten Schutz vor unnötigen Tests, Diagnosen und Behandlungen.
3. Für eine gesundheitsförderliche Selbstvermessung (Apps, Wearables) müssen Messdaten hinsichtlich ihrer

Fehleranfälligkeit und der Reichweite ihrer Implikationen verstanden werden.

Dieser narrative Übersichtsartikel nimmt, nach einer begrifflichen Einordnung, zwei Herausforderungen für die digitale Risikokompetenz in den Fokus: Wie finden Laien verlässliche und verständliche digitale Gesundheitsinformationen und wie können sie die Qualität von algorithmischen Entscheidungssystemen (z. B. in Apps) besser beurteilen? Für jede Herausforderung werden jeweils kompetenzfördernde Lösungsansätze beleuchtet.

Digitale Risikokompetenz

Risikokompetenz (Risk Literacy) bedeutet, in ungewissen Situationen die Entscheidungsoptionen und möglichen Konsequenzen beurteilen und zum eigenen Nutzen entscheiden zu können. Digitale Risikokompetenz bedeutet speziell, den Nutzen und Schaden von digitalen Technologien und Informationen kritisch beurteilen zu können, die digitalen Angebote kritisch nutzen zu können und für diese auch Evidenz aufzufinden, also z. B. zu verstehen, was eine App, ein Onlinetest oder eine personalisierte Behandlung wirklich leisten können [4, 5]. Risikokompetenz grenzt sich trotz Überschneidungen von der Ge-

sundheitskompetenz (Health Literacy) ab: Maßnahmen für mehr Gesundheitskompetenz möchten das Wissen, die Motivation und die Kompetenzen stärken, relevante Informationen zu finden, diese zu verstehen, zu beurteilen und im Sinne der eigenen Lebensqualität anzuwenden [6]. Speziell im digitalen Kontext gibt es auch eine erhebliche Überschneidung mit Medienkompetenz, etwa in der E-Health Literacy [7].

Forschung zur Risikokompetenz untersucht speziell die Auseinandersetzung mit statistischer Evidenz, deren Auffinden in der digitalen und analogen Umwelt, deren kritische Bewertung und deren Nutzung für das informierte Entscheiden: Edward Cokely und seine Kollegen [8] fokussieren beispielsweise auf die Diagnostik von Kompetenzen im Umgang mit Statistiken, Laura Martignon und ihre Kollegen [9] erforschen Kompetenzinterventionen im Rahmen der institutionellen Bildung.

Informiertes Entscheiden verlangt eine Abwägung der Evidenz möglichen Nutzens und Schadens, die aus den jeweiligen Optionen resultieren. Hierfür sind evidenzbasierte und verständliche, qualitätsgesicherte Gesundheitsinformationen erforderlich. Deshalb ist entscheidend zu wissen, wie und wo diese Informationen zu finden sind.

¹ Um das Leseverständnis zu erleichtern, wird nur eine sprachliche Form genutzt. Hierbei sind Angehörige aller Geschlechter eingeschlossen.

Tab. 1 Nutzeranleitungen zur Überprüfung von Gesundheitsinformationen im deutschsprachigen Raum

Herausgeber	Name der Nutzeranleitung	Quelle	Letzte Aktualisierung	Ziel	Format	Anzahl Prüfmale	Nutzertestung, Validierung an Gesundheitsinformationen, Güte
Gesellschaft für Versicherungswissenschaft und -gestaltung (GVG)	Qualitätskriterien für Gesundheitsinformationen	Gesundheitsziele.de	2011	Qualität von Gesundheitsinformationen einschätzen	Checkliste zur Qualitätsprüfung	9	<i>Nutzertestung:</i> Verständlichkeit und Nutzerfreundlichkeit
Verbraucherzentrale Nordrhein-Westfalen (VZ NRW)	10 Kriterien für die Bewertung der Transparenz und Verlässlichkeit	Vz-nrw.de	2015	Transparenz und Verlässlichkeit von Gesundheitsinformationen prüfen	Checkliste	10	Nicht bekannt
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)	Wie finde ich gute Gesundheitsinformationen im Internet?	gesundheitsinformation.de	2018	Gute Gesundheitsinformationen zu finden	Merkmalslisten und Checkliste	18	Nicht bekannt
Department für Evidenzbasierte Medizin, Donau-Universität Krems, Redaktion Medizinttransparent.at	Wie erkenne ich vertrauenswürdige Gesundheitsinformationen?	Medizinttransparent.at	2018	Vertrauenswürdige Gesundheitsinformationen erkennen	Checkliste	10	Nicht bekannt
Ärztliches Zentrum für Qualität in der Medizin (ÄZQ, basierend auf dem DISCERN-Instrument, 2005)	Qualität von Gesundheitsinformationen im Internet	Patienteninformation.de	2019	Informationsqualität erkennen	Merkmalslisten mit Erläuterungen	9	Nicht bekannt
Bertelsmann Stiftung	Kriterien zur Beurteilung des Schadenspotenzials von Gesundheitsinformationen	Bertelsmannstiftung.de	2019	Schadenspotenzial überprüfen	Kriterienraster zur Prüfung	16	<i>Validierung</i> trotz Praxistest nicht quantifiziert; <i>Güte:</i> „nicht für objektive, reliable und valide Einschätzungen“ (S. 37)
Harding-Zentrum für Risikokompetenz	RisikoAtlas: Fast-and-frugal Tree: Gesundheitsinformationen im Internet	Risikoatlas.de	2019	Gesundheitsinformationen identifizieren, die informiertes Entscheiden verhindern	Entscheidungsbaum mit Erläuterungen	4	<i>Nutzertestung:</i> Qualitative und quantitative Nutzertests, experimentelle Validierung; <i>Validierung:</i> einfache Kreuzvalidierung <i>Güte:</i> 91 % Sensitivität, 57 % Spezifität

Wie finden Menschen qualitätsgesicherte Gesundheitsinformationen?

Eine Reihe von Studien zeigt, dass nach wie vor die wenigsten digitalen Gesundheitsinformationen informiertes Entscheiden ermöglichen. Schon zu Beginn des Jahrtausends waren evidenzbasierte Informationen international nicht die Regel [10], auch nicht bei großen deutschen Krankenkassen [11]. Früherkennungsinformationen deutschsprachiger Krebsgesellschaften entsprachen auch 2014 nur teilweise [12] den Qualitätskriterien evidenz-

basierter Patienteninformationen [13] ebenso 2015 Nierenkrebsinformationen [14], 2016 Diabetesinformationen [15] und 2017 Früherkennungsinformationen [16] im englischen Sprachraum. Privatwirtschaftliche Internetseiten waren zudem seltener korrekt bezüglich Therapieoptionen als öffentliche und akademische Angebote zum Thema „Bauchspeicheldrüsenkrebs“ [17].

Viele Angebote, die Laien bei ihrer Suche nach Gesundheitsinformationen finden, sind der Vorbereitung informierter Entscheidungen nicht dienlich. Zwar werden die Bewertungen oft von den Fakten abgegrenzt, aber diese Fakten sind

selbst nur unklar dargestellt [18, 19]. Solche Angebote helfen kaum, den möglichen Nutzen und Schaden von Tests und Behandlungen zu verstehen. Zusammenfassend kann festgestellt werden, dass sich das von analogen Medien (z. B. Gesundheits- und Werbebroschüren) bekannte Problem irreführender Gesundheitsinformationen in den digitalen Medien nicht verbessert, sondern eher verstärkt hat.

Bisherige Lösungsansätze adressieren die Verbesserung des Angebots von deutschsprachigen Gesundheitsinformationen wie auch die Unterstützung der Nutzerseite. Das Angebot

evidenzbasierter und verständlicher Informationen wächst: Neben öffentlichen Einrichtungen wie dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG (gesundheitsinformation.de) bieten auch Krankenversicherer evidenzbasierte Gesundheitsinformationen an (z. B. igel-monitor.de; aok.de/faktenboxen; helsana.ch/de/blog/stichworte/gesundheitskompetenz). Für ein auffindbares qualitätsgeprüftes Angebot mit verständlichen Gesundheitsinformationen wurde mit einem Stellungnahmeverfahren das „Nationale Gesundheitsportal“ durch das IQWiG auf den Weg gebracht und soll unter Federführung des Bundesministeriums für Gesundheit bis Mitte 2021 den Regelbetrieb aufnehmen.

Zur Unterstützung der Nutzer wurden verschiedene Maßnahmen auf den Weg gebracht, die sie befähigen sollen, qualitätsgesicherte Gesundheitsinformationen von problematischen zu unterscheiden. So wollen verschiedene Akteure mit Qualitätssiegeln Orientierung bieten, z. B. die schweizerische Stiftung „Health on the Net“ (HON; [20]) oder die deutsche „Stiftung Gesundheit“ [21]. Doch ist die Bekanntheit und Wirksamkeit angesichts der Vielzahl von Internetsiegeln kaum gegeben. Schwer aufzufinden sind Positivlisten vertrauenswürdiger Internetquellen (z. B. wissenwaswirkt.org/qualitaet-bei-gesundheitsinformationen-ist-moeglich). Initiativen wie das „Aktionsbündnis Gesundheitskompetenz“ [22] oder das Projekt „OriGes“ am Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health (ceres; [23]) verfolgen direkte Bildungsansätze und bieten komplexe Interventionen.

Zeit und Ressourcen für dieses spezifische Bildungsangebot sind jedoch vor allem in institutionellen Settings zu erwarten. Um allen anderen Menschen, die bislang Informationen nicht kritisch genug geprüft haben [24], auch im Alltag das Auffinden von evidenzbasierten Informationen zu ermöglichen und sie damit bei einer informierten Entscheidung zu unterstützen, wurden von verschiedenen Institutionen sogenannte Nutzeranleitungen mit Prüfkriterien entwickelt ([25]; [Tab. 1](#)).

Bundesgesundheitsbl 2020 · 63:665–673 <https://doi.org/10.1007/s00103-020-03146-3>
© Der/die Autor(en) 2020

F. G. Rebitschek · G. Gigerenzer

Einschätzung der Qualität digitaler Gesundheitsangebote: Wie können informierte Entscheidungen gefördert werden?

Zusammenfassung

Eine wichtige Voraussetzung für das Gelingen der Digitalisierung des Gesundheitswesens ist die digitale Risikokompetenz der Nutzer, also ihre Fähigkeit, Nutzen und Schaden von digitalen Technologien und Informationen zu beurteilen, digitale Angebote kritisch zu nutzen und sich auch mit statistischer Evidenz auseinanderzusetzen. Wie finden Menschen qualitätsgesicherte Gesundheitsinformationen und wie können sie die Qualität von algorithmischen Entscheidungssystemen besser beurteilen? In diesem narrativen Beitrag sollen zwei Ansätze aufgezeigt werden, wie die Fähigkeit zum informierten Entscheiden gefördert werden kann. Evidenzbasierte und verlässliche Gesundheitsinformationen existieren im Internet, müssen aber von einer Vielzahl unzuverlässiger Informationen unterschieden werden. Verschiedene Institutionen im deutschen Sprachraum haben deshalb Anleitungen bereitgestellt, um Laien eine informierte Entscheidung zu erleichtern. Beispielsweise hat das Harding-Zentrum

für Risikokompetenz in Potsdam für diese Zwecke einen Entscheidungsbaum („fast-and-frugal tree“) entwickelt. Im Umgang mit Algorithmen können natürliche Häufigkeitsbäume (NFTs) helfen, die Güte und Fairness eines algorithmischen Entscheidungssystems zu beurteilen.

Neben zuverlässigen und verständlichen digitalen Angeboten sollten weitere Werkzeuge für Laien zur Beurteilung von Informationen und Algorithmen entwickelt und bereitgestellt werden. Diese können auch in Schulungsprogramme zur digitalen Kompetenzförderung aufgenommen werden. Damit wäre ein wichtiger Schritt zum Gelingen der Digitalisierung in der Prävention und Gesundheitsförderung getan.

Schlüsselwörter

Risikokompetenz · Informiertes Entscheiden · Digitale Gesundheitsinformationen · Algorithmische Entscheidungssysteme · Algorithmen Güte

Assessing the quality of digital health services: How can informed decisions be promoted?

Abstract

An important prerequisite for the success of the digitisation of the healthcare system are risk-literate users. Risk literacy means the ability to weigh potential benefits and harms of digital technologies and information, to use digital services critically, and to understand statistical evidence. How do people find reliable and comprehensible health information on the Internet? How can they better assess the quality of algorithmic decision systems? This narrative contribution describes two approaches that show how the competence to make informed decisions can be promoted.

Evidence-based and reliable health information exists on the Internet but must be distinguished from a large amount of unreliable information. Various institutions in the German-speaking world have therefore provided guidance to help laypersons make informed decisions. The Harding Center for

Risk Literacy in Potsdam, for example, has developed a decision tree (“fast-and-frugal tree“). When dealing with algorithms, natural frequency trees (NFTs) can help to assess the quality and fairness of an algorithmic decision system.

Independent of reliable and comprehensible digital health services, further tools for laypersons to assess information and algorithms should be developed and provided. These tools can also be included in institutional training programmes for the promotion of digital literacy. This would be an important step towards the success of digitisation in prevention and health promotion.

Keywords

Risk literacy · Informed decision-making · Digital health information · Algorithmic decision systems · Algorithm performance

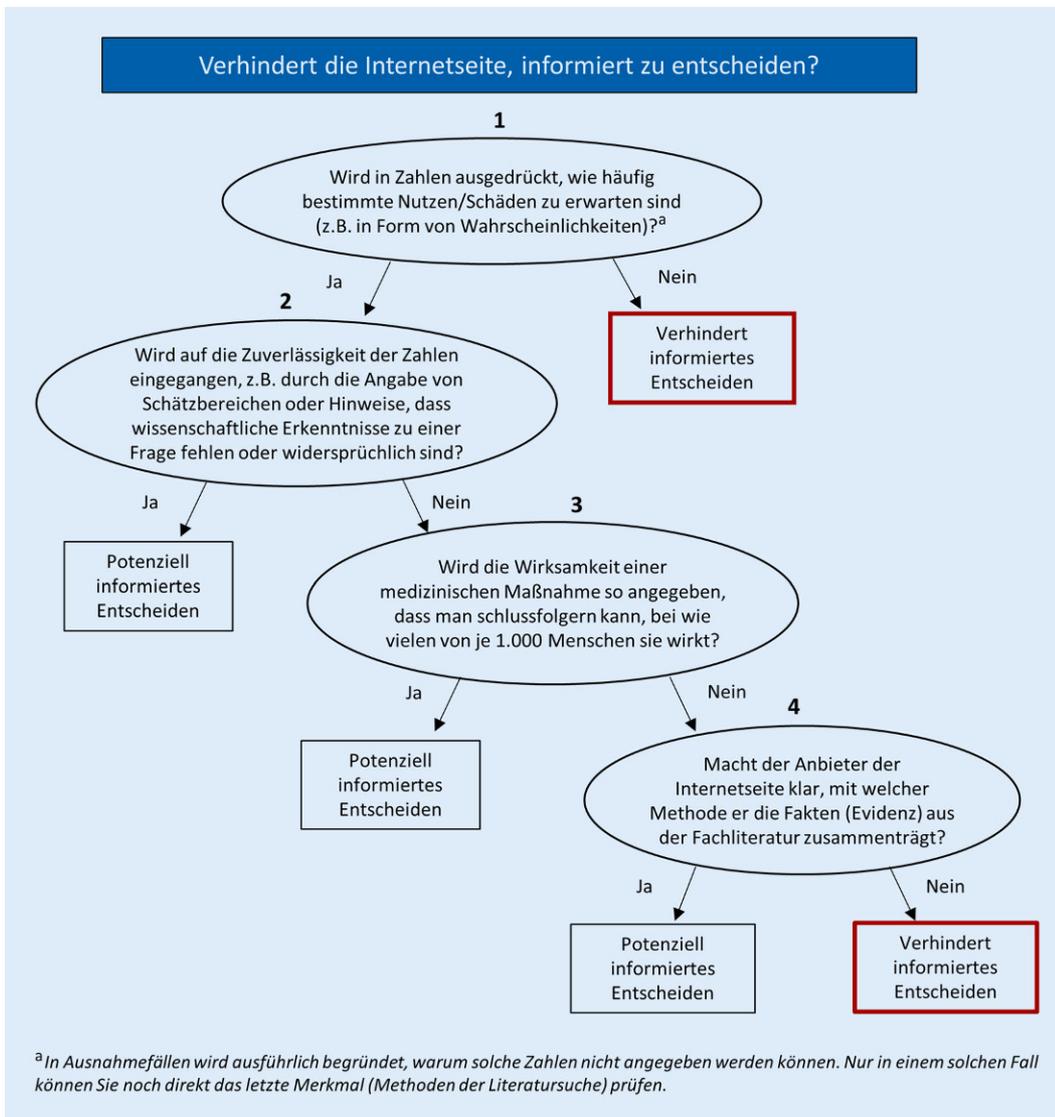


Abb. 1 ◀ Nutzeranleitung zum Erkennen von Gesundheitsinformationen, die eine informierte Entscheidung verhindern. Entscheidungsbaum (Fast-and-frugal Tree). Quelle: Projekt „RisikoAtlas“ des Harding-Zentrums für Risikokompetenz

Nutzeranleitungen ermöglichen keine Prüfung, inwieweit Leitlinien zur evidenzbasierten Gesundheitsinformation [26] erfüllt sind, aber sie reduzieren die Ungewissheit, bezogen auf die Qualität eines konkret vorliegenden Informationsangebotes. Sie helfen, die Spreu vom Weizen zu trennen. Die wiederholte Verwendung dieser Nutzeranleitungen kann zudem kompetenzfördernd sein, wenn bestimmte Prüfmerkmale verinnerlicht werden. Als nichtmonetäre und nichtregulatorische Interventionsansätze für gesellschaftliche Herausforderungen gehören sie damit zu den Boosts, in Abgrenzung von den Nudges. Während Methoden des Nudging (zu Deutsch: Schubsen) durch die Gestaltung von Entscheidungssituationen die Entschei-

dungen des Einzelnen im Sinne des Gestalters lenken sollen (z.B. wenn das Obst im Vergleich zu den Süßigkeiten in der Cafeteria besser erreichbar platziert wird), dienen Boosts der Kompetenzförderung des Einzelnen für seine Entscheidungen in seinem bzw. ihrem eigenen Sinne ([27]; z.B. Materialien, um zu lernen, warum Obst so wichtig ist).

Herausgegeben werden Nutzeranleitungen nicht nur von klassischen Akteuren im Gesundheitswesen (wie dem IQWiG), sondern auch von thematisch breiter aufgestellten Institutionen (wie der Bertelsmann Stiftung, Gesellschaft für Versicherungswissenschaft und -gestaltung (GVG), Verbraucherzentrale). In der Regel haben die Nutzeranlei-

tungen die Form von Checklisten oder Rastern, resultieren teilweise aus der Arbeit der Akteure an Anbietervorgaben (Ärztliches Zentrum für Qualität in der Medizin (ÄZQ), IQWiG) und variieren in den Nutzungszielen. So zielen sie auf Transparenz, Verlässlichkeit bzw. Vertrauenswürdigkeit ab (Verbraucherzentrale, Medizin-transparent.at vom Department für Evidenzbasierte Medizin und Evaluation an der Donau-Universität Krems) oder auf Qualitätseinschätzungen (ÄZQ, GVG, IQWiG) oder dienen als Warnsysteme (Bertelsmann Stiftung, Harding-Zentrum für Risikokompetenz). Zu unterscheiden sind hierbei die komplexeren Verfahren mit 16 bzw. 18 zu prüfenden Merkmalen (Bertelsmann Stiftung, IQWiG)

von einfacheren mit 9 bis 10 Merkmalen (ÄZQ, GVG, Medizin-transparent, Verbraucherzentrale) und schnellen mit maximal 4 Merkmalen (Harding-Zentrum für Risikokompetenz).

Jedoch ist die Validität bei den meisten dieser Werkzeuge nicht überprüft worden. Das gilt sowohl für ihre Fähigkeit, zuverlässige und unzuverlässige Gesundheitsinformationen jeweils korrekt zu klassifizieren, als auch ihre Effektivität, die Informationssuche von Nutzern positiv zu verändern. Der Entscheidungsbaum (Fast-and-frugal Tree, FFT) zur Erkennung von webseitenbasierten Gesundheitsinformationen, die informiertes Entscheiden verhindern, ist hier die Ausnahme. FFTs ermöglichen generell schnelle und zuverlässige Entscheidungen bei Unsicherheit [28]. Beispiele gibt es in der Notfallmedizin [29], aber auch in der Finanzwelt [30]. FFTs stehen in Form einer grafisch aufgearbeiteten, einfachen Baumstruktur sowohl digital (App, Internetseite) als auch analog für Laien zur Verfügung (Poster, Broschüren). Der FFT in **Abb. 1** wurde mithilfe von Expertenbewertungen von Gesundheitsinformationen modelliert. Wie eine von uns noch nicht veröffentlichte Studie zeigt, ist er effizient, indem er bei 9 von 10 Informationsangeboten, die keine informierte Entscheidung ermöglichen würden, warnt. Er ist zudem effektiv, da er seinen Nutzern eher Angebote nahelegt, die informiert zu entscheiden helfen [18, 19]. Der FFT ist in eine Verbraucher-App (Risiko-Kompass) integriert und kann so im Alltag eingesetzt werden. Die dadurch ermöglichte kritischere Informationssuche kann sich positiv auf bislang problembehaftete Arzt-Patienten-Gespräche zu Internetinformationen [31] auswirken und damit eine wesentliche Voraussetzung für eine nutzbringende Digitalisierung im Bereich Prävention und Gesundheitsförderung sein.

Wie können Menschen die Qualität von algorithmischen Entscheidungssystemen beurteilen?

Algorithmische Entscheidungssysteme („algorithmic decision making“, ADM;

[32]) liefern Informationen zur Vorbereitung einer Entscheidung (unterstütztes Entscheiden) oder einen Wert, der eine bestimmte Entscheidung auslöst (automatisiertes Entscheiden). ADM-Systeme personalisieren dabei auf Basis vergangener und gegenwärtiger Merkmale Entscheidungsgrundlagen bzw. Entscheidungen. Das Anwendungsfeld dieser Systeme ist sehr breit: Bonusprogramme der Krankenkassen belohnen kontinuierlich persönliche Aktivitäten, Smartphone-Apps geben diagnostische Risikohinweise und private Berufs-, Kranken- und Lebensversicherer bieten personalisierte Risikotarife. Im medizinischen Sektor ist es die 4P-Medizin, die personalisiert, prädiktiv und präventiv sowie partizipativ sein möchte (z. B. polygene Risikoscores für eine Fülle persönlicher Erkrankungsrisiken [33]). Alle Systeme beeinflussen die Entscheidungsautonomie, bis hin zu der Frage, inwieweit der Einzelne noch an einer Entscheidung beteiligt ist.

Evidenz ist wesentlich für partizipative Gesundheitsentscheidungen. Nicht nur die verständliche Kommunikation eines diagnostischen oder prädiktiven Risikowertes in Verbindung mit Handlungsoptionen ist hier relevant [34], sondern auch die Rahmenbedingungen seines Zustandekommens. Für eine informierte Auseinandersetzung sind erstens der Nutzen und Schaden durch den Einsatz des Algorithmensystems (auf individueller, sozialer und gesellschaftlicher Ebene) im Vergleich zur Standardversorgung ohne Algorithmus zu beleuchten. Zweitens muss analysiert werden, welche Merkmale welchen Einfluss haben, und drittens muss untersucht werden, wie gut das System überhaupt funktioniert und inwieweit es bestimmte Personengruppen diskriminiert [35]. Die letzteren beiden Punkte werden später wieder aufgegriffen.

Zur Frage, wie gut ein System überhaupt funktioniert: Jegliche Algorithmen, die Entscheidungen unterstützen sollen, sind anhand von Daten aus der Vergangenheit und einer spezifischen Population entwickelt (trainiert) worden. Ihre Güte muss unter realen Umständen (Populationen) geprüft werden. Sie ist vor allem bei der Analyse von visuellem

und sprachlichem Input hoch. Solcher Input bleibt prinzipiell recht stabil über die Zeit und verschiedene Umwelten hinweg: Röntgenaufnahmen von Nieren beispielsweise sehen 1999 und 2019 sehr ähnlich aus. Tatsächlich können Krankheiten von Deep-Learning-Algorithmen, welche sich eigene neue Indikatoren aus Patientenbildern ableiten, genauso gut wie von medizinischen Fachkräften erkannt werden [36]. Die 2019 in einem Arzt-Patienten-Gespräch verwendeten Worte wären auch 1999 verstanden worden. So eine stabile Situation ist ideal für Algorithmen. Niedrige Fehlalarmraten wären bei der Erkennung von seltenen genetischen Phänotypen aus Bildern von Gesichtern möglich (Cornelia-de-Lange-Syndrom, 100 % Spezifität mit einem 95 %-Konfidenzintervall von 100–100 %, bei einer Sensitivität von 96 % [87 %; 100 %]; [37]).

Gerade viele bekannte Klassifikationsalgorithmen zeichnen sich jedoch durch „innovative“, weniger stabile Inputdaten in dynamischen komplexen Umwelten der echten Welt aus. Beim Erschließen von Homosexualität aus Facebook-Likes [38] oder auch beim Erkennen von Traurigkeitszuständen aus Tastaturanschlägen [39] sind „Treffer“ mit 30–40 %iger Wahrscheinlichkeit korrekt. Dass dies nicht einem unscharfen oder schlecht zu messendem Zielmerkmal geschuldet sein muss, lässt sich auch an der begrenzten Güte von Big-Data-Algorithmen zur Vorhersage von Bauchspeicheldrüsenkrebsdiagnosen erkennen [40]. Sind schlicht die Trainingsdaten eines Algorithmus für den Anwendungsfall ungeeignet, werden z. B. Asthmatikern im Fall einer Lungenentzündung schon mal bessere Überlebenschancen als Nichtasthmatikern ausgerechnet [41]. Die Asthmatiker, welche in den Trainingsdaten enthalten waren, erhielten tatsächlich eine andere Versorgung als üblich.

Zum anderen stellt sich das Problem der Fairness des Algorithmus, selbst wenn ein ADM-Algorithmus geprüft und entlastet wurde: Stammen Trainingsdaten aus der echten Welt und enthalten diskriminierende Aspekte, dann manifestieren und intensivieren sich diese Aspekte leicht. Patienten können nach ethni-

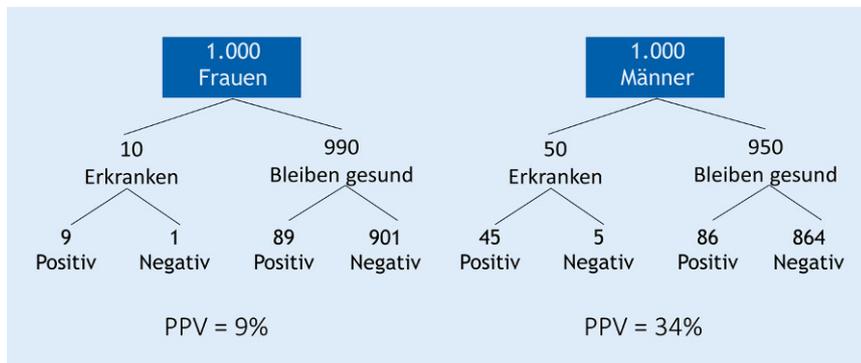


Abb. 2 ▲ Zwei natürliche Häufigkeitsbäume, aus denen sich der positive prädiktive Wert (PPV) berechnen lässt (Formel: Anzahl der Richtig-Positiven/(Anzahl der Richtig-Positiven + Anzahl der Falsch-Positiven) in Prozent). Der PPV gibt an, wie viele Personen, bei denen eine bestimmte Krankheit im Testverfahren angezeigt wurde, auch tatsächlich krank sind. Hier variieren diese Werte für Männer und Frauen aufgrund unterschiedlicher Erkrankungsgrundraten, trotz konstanter Sensitivität und Spezifität

scher Herkunft auch dann diskriminiert werden, wenn Größen, die mit der ethnischen Herkunft korrelieren (z. B. Gesundheitskosten, die systembedingt verschiedene Ethnien unterschiedlich betreffen), mit in die Berechnung eingehen [42]. Stammen Trainingsdaten hingegen nicht aus einer repräsentativen Umgebung, können algorithmische Entscheidungssysteme, voreilig eingesetzt, ebenso leicht diskriminieren, wenn ihre Zuverlässigkeit nur für eine verzerrte Population gesichert werden kann (6% Frauenanteil als Basis zur Vorhersage akuter Nierenschädigungen [43]). Um die ADMs hinsichtlich Güte und Diskriminierung zu hinterfragen und zu überprüfen, erfordert es risikokompetente Menschen. Konkret besteht eine zur Risikokompetenz gehörende Fähigkeit darin, natürliche Häufigkeitsbäume (Natural Frequency Trees, NFTs; [44]) eigenständig auf Problemstellungen anzuwenden (Abb. 2).

In NFTs stellt man Betroffene (z. B. von einer Krankheit) und Nichtbetroffene als Ziele eines Algorithmus (oder Tests) einander gegenüber. Dies lässt sich am links gezeigten Häufigkeitsbaum in Abb. 2 kurz erläutern. Die erste Frage, welche man stellen würde, wäre: Wie viele betroffene Frauen gibt es eigentlich (als Neuerkrankungsrate, Inzidenz oder Prävalenz von Erkrankungen)? Wenn im dort gezeigten Beispiel 1% der Frauen erkrankt, entspricht das 10 von 1000 Frauen. Danach fragt man, wie viele der 10 betroffe-

nen Frauen erkannt werden. Ausgehend von einer vorgegebenen 90% Sensitivität des Algorithmus, sind es hier 9 Frauen. Es wird daher auch klar, wie viele von ihnen nicht erkannt werden; also hier 1 erkrankte Frau. Die dritte und letzte Frage lautet analog, wie viele der Nichtbetroffenen (990 Frauen) einen Fehlalarm erhalten. Ein Fehlalarm liegt genau dann vor, wenn eine Nichtbetroffene ein positives Ergebnis erhält. In diesem Beispiel ist eine Fehlalarmrate von 9% vorgegeben (entspricht 100–91% Spezifität), wodurch 89 der 990 Frauen einen falsch-positiven Befund erhalten. Man kann nun erkennen, wie viele von jenen, die positiv testen (nämlich 9 + 89 Frauen), tatsächlich betroffen sind (nämlich 9 Frauen): Der positive prädiktive Wert (PPV) entspricht hiernach 9 von 98 Frauen, d. h. etwa 9%. Der Nutzer, aber auch eine vom Algorithmus betroffene Person kann mit NFTs im Vorhinein mit vertretbarem Aufwand die persönlich relevante Frage beantworten, wie wahrscheinlich ein Algorithmus- oder Testergebnis denn auch richtig ist. Das bayesianische Schlussfolgern auf diesen PPV wird durch Häufigkeitsbäume erleichtert [45] und auch Schulkinder können den Umgang damit lernen [46]. Häufigkeitsbäume sind den meisten Laien jedoch bislang nicht bekannt.

Nicht nur zur Ermittlung der Güte von Algorithmen, sondern gerade auch bezüglich ihres Diskriminierungspotenzials stellen NFTs ein wichtiges Werk-

zeug dar. Anhand von Abb. 2 ist vor allem die Bedeutung der NFTs zur Prüfung der Fairness (im Sinne der Nichtdiskriminierung verschiedener Gruppen) hervorzuheben. Hierzu vergleicht man mittels NFTs zwei Gruppen im Hinblick auf ihre Gleichbehandlung. Wie alle Tests unterliegen Algorithmen dem Problem multipler Fairnessanforderungen. Konkret kann ein Algorithmus für zwei Gruppen (z. B. Männer und Frauen), in denen die Grundrate des avisierten Ziels (z. B. die Häufigkeit einer Krankheit) verschieden ist, nicht gleichzeitig genauso empfindlich sein (Sensitivität), genauso spezifisch sein und genauso wahrscheinlich richtig liegen (PPV). Die Algorithmenentwickler müssen also, mit Blick auf die Fairness gegenüber beiden Gruppen, eine Balance von verschiedenen Fairnessindikatoren finden [32]. Diese Balance bedarf einer Aushandlung verschiedener Anspruchsgruppen, da verschiedene Irrtümer eines Algorithmus mit unterschiedlichen Kosten für die beteiligten Akteure verbunden sind. Die Abbildung macht deutlich, wie sich der PPV aufgrund der Grundrate für Männer und Frauen stark unterscheidet, obwohl Sensitivität und Spezifität konstant gehalten sind. Um beiden Gruppen die gleiche Irrtumswahrscheinlichkeit im Fall eines Verdachts durch den Algorithmus zu ermöglichen, müsste eine Gruppe ein verändertes Fehlerverhältnis gegenüber der anderen akzeptieren. Mithilfe der NFTs sind diese Abwägungen auch für Laien diskutierbar und die Fähigkeit, sie mit Zettel und Stift zu nutzen, erleichtert die kritische Auseinandersetzung mit Algorithmen-Systemen.

Diskussion und Ausblick

Angesichts der Herausforderungen verlässlicher, verständlicher digitaler Gesundheitsinformationen für Patientenentscheidungen sowie der informierten Auseinandersetzung mit algorithmischen Entscheidungssystemen wurde schon 2016 mehr Risikokompetenz bei den Menschen gefordert [1]. In diesem Beitrag werden zwei kompetenzfördernde Techniken (Boosts) vorgestellt: Nutzeranleitungen und natürliche Häufigkeitsbäume.

Die Literatur zu ADM-Systemen verdeutlicht, dass diese Systeme immer auch fehlerhaft sind, wobei das Verhältnis von Fehlerarten (z. B. weniger übersehen, dafür mehr Fehlalarme) für den Anwender bzw. den Betroffenen persönlich relevant ist (z. B. fürchtet der eine, dass ein Krankheitsrisiko übersehen wird, während die andere verunsichert wird, wenn bei vielen Krankheitsrisiken fälschlich alarmiert wird). Um die Güte von ADM-Systemen kritisch und quantitativ beurteilen zu können, ist ein informierter Umgang vonnöten. Derselbe wird für eine ausgewogene Fairnessabwägung zwischen verschiedenen Zielgruppen des Algorithmus benötigt. Mithilfe von sogenannten Natural Frequency Trees können Laien die Güte als auch die jeweilige Fairnessbalancierung verständlich nachvollziehen.

Bis heute stehen evidenzbasierte und verlässliche Gesundheitsinformationen einer Vielzahl von Informationsangeboten gegenüber, die informiertes Entscheiden nicht unterstützt. Verschiedene Nutzeranleitungen, die in deutscher Sprache verfügbar sind, sollen deshalb helfen, qualitätsgesicherte und problematische Informationsangebote voneinander zu unterscheiden. Untersucht wurde die Wirksamkeit nach unserer Kenntnis nur für eine Nutzeranleitung, einen „fast-and-frugal tree“. So finden Nutzer eher evidenzbasierte Gesundheitsinformationen, wenn sie einen solchen Entscheidungsbaum nutzen können [18, 19]. Dies bietet eine Chance, informiertes Entscheiden zu fördern. Nach ihrer Erfahrung geben 84% der Hausärzte an, dass Patienten durch Internetinformationen zu Gesundheit und Krankheit verwirrt und verunsichert werden, 72% meinen, dass Patienten nervöser und ängstlicher würden. Umgekehrt sehen wenige, dass die Patienten sich sicherer fühlen würden (4%) und informierter wären und die Ärzte besser verstehen würden (29%; [31]). Die Informationen, welche die Patienten tatsächlich nutzen, sollten also in Zukunft verständlicher sein und informiertes Entscheiden ermöglichen. Die Angebotsseite könnte sich mit einem nationalen Gesundheitsinformationsportal und der geförderten Bekanntheit von Qualitätssiegeln eben-

falls verbessern. Nutzer, die institutionell tätig sind (Kindertageseinrichtungen, Schulen, Berufsschulen, Hochschulen, Arbeitgeber), können von Initiativen für mehr digitale Gesundheits- und Risikokompetenz profitieren. Für alle anderen, also vor allem für Laien, bieten sich kompetenzfördernde Nutzeranleitungen an. Die Herausforderungen in Bezug auf Nutzeranleitungen sind jedoch vielfältig.

Ein Kritikpunkt besteht darin, dass interessen geleitete Anbieter von Gesundheitsinformationen auf transparente Nutzeranleitungen durch *Gaming* reagieren können (zu Deutsch die Anleitungen *ausspielen*, z. B. erfundene Statistiken einfügen, die wissenschaftlich aussehen, aber vom Nutzer nicht überprüft werden können). Das bedeutet, dass sie versuchen, die Erfüllung der einzelnen Prüfmerkmale (von Checklisten, Kriterienrastern und Entscheidungsbäumen) soweit nachzuahmen („scheinzu erfüllen“), dass die Anwender der Nutzeranleitungen zu falschen Schlussfolgerungen gelangen. Das damit verbundene Problem sind aufwendige Erstellungs- und Aktualisierungsprozesse, um Merkmale, die durch *Gaming*, aber auch durch ein verändertes Gesamtangebot entwertet wurden, durch bessere Indikatoren zu ersetzen. Damit eine Nutzeranleitung valide im Hinblick auf tatsächlich von Laien gefundene Gesundheitsinformationen ist, muss sie an diesen entwickelt und unabhängig von der Entwicklungsgrundlage getestet werden. Das bedeutet, für jede Aktualisierung müssen neue Fälle echter Gesundheitsinformationen systematisch erhoben werden. Deren Prüfmerkmale und neue Kandidatenmerkmale müssen erfasst werden. Außerdem muss ein Kriteriumswert für jeden Fall ermittelt werden, z. B. bewerten Experten unabhängig voneinander jeden Fall. Durch einen solchen Aktualisierungsprozess verändert sich die Nutzeranleitung über die Zeit, alte Indikatoren werden entwertet, neue kommen hinzu.

Ein weiterer Kritikpunkt betrifft die Voraussetzungen, welche Nutzer mitbringen müssen. Das Bildungsniveau und die Lesefähigkeiten müssen ausreichen, um im Internet Recherchen durchführen und Texte lesen zu können.

Durch Texte kaum angesprochene Nutzer wenden sich alternativ Gesundheitsvideos zu. Dieses Gesamtangebot ist jedoch ebenfalls von problematischer Qualität (z. B. anorexiefördernde englischsprachige Youtube-Videos), wodurch auch hier evidenzbasierte Varianten im Angebot benötigt werden (Helsana, Stiftung Gesundheit). Die Nutzer von textbezogenen Anleitungen müssen ferner Gesundheitsinformationen auf Merkmale prüfen können. Die Nutzeranleitungen von der Bertelsmann Stiftung und dem Harding-Zentrum für Risikokompetenz (Abb. 1) wurden nutzergetestet. Andere Nutzeranleitungen werfen die Frage auf, ob einzelne Prüfkriterien für Laien ausreichend verständlich sind, etwa ob sie wissen, was „große klinische Studien“ sind (ÄZQ), ob eine Autorin „die nötige Ausbildung“ hat (Medizin-transparent) oder was „Prozesse der Erstellung“ sind (IQWiG). Es wird deutlich, dass die Lücke zwischen dem Fachkriterium hinter dem Merkmal und der Prüfbarkeit durch Laien nicht in jedem Fall perfekt geschlossen werden kann, sofern auch der Aufwand für Laien begrenzt werden soll. Umso wichtiger ist die validierte Auswahl der trennschärfsten und zugleich laienverständlichen Prüfkriterien.

Grundlegender könnte man für bestimmte Bevölkerungsgruppen die Zielstellung von Nutzerentscheidungen auf Basis von statistischer Evidenz infrage stellen. Gerade Personen mit geringem Zahlenverständnis haben Schwierigkeiten, Tabellen zu interpretieren. Hier könnten jedoch geeignete Präsentationsformate helfen. Tatsächlich profitieren gerade sie von grafischen Präsentationsformaten, wie den Icon Arrays (in Feldern gruppierte Symbole; [47]), welche sich in evidenzbasierte Informationsangebote integrieren lassen [48]. Flankiert werden können diese Schlüsselinformationen von zielgruppengerechten Narrativen oder Videos. Gerade mit Blick auf Evidenzsuche und Evidenznutzung hilft zudem die Vermittlung von Konzepten zum informierten Entscheiden und zur Erkennung, ob eine Entscheidung wirklich informiert ist [49].

Mit Blick auf die informierte Auseinandersetzung mit ADM-Systemen stehen ebenfalls spezifische Voraussetzungen in-

frage. Die Bevölkerung gibt mehrheitlich an, Algorithmen nicht zu verstehen [50], und Statistik- und Programmierkenntnisse sind nur begrenzt verbreitet. Ein Zusammenspiel von Bildungsinterventionen in Kombination mit gesetzlichen Anforderungen an die Präsentation von ADM-Systemen und ihren Ergebnissen sollte nicht in Abrede gestellt werden. Interventionen zur Algorithmenkompetenz genauso wie Informationsbroschüren zu spezifischen ADM-Systemen könnten jedoch durch Prüfschemata wie die NFTs praxistauglich ergänzt werden. Verständlich gefragt, zeigt die Bevölkerung in Deutschland durchaus Präferenzen hinsichtlich der Gestaltung von ADM-Systemen [35]. Außerdem können Nebenfolgen des Einsatzes durch eine informierte Auseinandersetzung diskutiert und politisch bzw. möglicherweise auch durch die Betroffenen selbst kompensiert werden.

Ein Fallbeispiel für Nebenfolgen bietet die Symptomchecker-App Ada, die einen interaktiven algorithmenbasierten Assistenten mit diagnostischen Hinweisen anbietet. Hierbei besteht ein Schadenpotenzial. Nutzer, die trotz Disclaimer Aussagen von diagnostischen Algorithmen für wahre Diagnosen nehmen, laufen Gefahr, unnötige psychologische Belastungen zu erleben und vorschnell zu intervenieren. Aus hausärztlicher Sicht ist bekannt, dass schon jene, die sich im Internet intensiv zu Symptomen und Krankheit informieren, eher eine weiterführende apparative Diagnostik einfordern [31]. Die zugrunde liegenden Algorithmen von Ada machen Fehler und die enthaltenen diagnostischen Hinweise, die Ursachen, können falsch sein. Ada gehört im Vergleich mit 23 anderen Symptomcheckern zwar zu den akkuratesten Apps, jedoch weist der durchschnittliche PPV von 56 % (95 %-KI 41–69 %) darauf hin, dass fast jeder zweite „positive“ diagnostische Hinweis falsch sein kann [51], wobei die meisten Ergebnisse weiter abgeklärt werden müssten. Für das Gesundheitssystem als Ganzes würde vor allem die Abklärung von Auffälligkeiten und Fehlarmen von millionenfach genutzten Apps zu einem erheblichen Zeitverlust und Aufwand führen. Außerdem gibt es in Deutschland bereits jetzt ein Prob-

lem von Überdiagnosen [52]. Daher muss nicht nur für Gesundheits-Apps, sondern jedes ADM-System mit gesundheitlicher Relevanz das Nutzen-Schaden-Verhältnis seiner möglichen Implementierung in einem angemessenen Verfahren bestimmt werden. Angesichts der Relevanz von Gesundheitsdaten und des bereits beschriebenen Problemverhaltens [53] gehört auch eine datenschutzrechtliche Betrachtung dazu.

Informierte Nutzer wissen, woran sie hilfreiche Gesundheitsinformationen erkennen. Sie fragen, welche wissenschaftlichen Belege es für den Nutzen einer neuen Gesundheits-App gibt. Am Beispiel der Apps auf Rezept, die diesen medizinischen Nutzen oder eine patientenrelevante Struktur- und Verfahrensverbesserung in der Versorgung binnen eines Jahres nachweisen sollen, lässt sich in den nächsten 1 bis 2 Jahren ablesen, wie diese Informationen gewonnen, dargestellt und auch verstanden werden sollen. Hierzu sollten auch die Auswirkungen der digitalen Angebote auf die gesundheitsbezogene Chancenungleichheit analysiert werden [7]. Qualitätsgesicherte Evidenz zum Nutzen-Schaden-Verhältnis digitaler Innovationen sowie ein angemessener Gesundheitsdatenschutz bieten die Chance für eine überaus erfolgreiche Digitalisierung – ohne sie ist die Digitalisierung in der Prävention und Gesundheitsförderung für die Patienten von unbekanntem Wert und wird am Ende scheitern.

Korrespondenzadresse

Dr. Felix G. Rebitschek

Fakultät für Gesundheitswissenschaften
Brandenburg, Harding-Zentrum für
Risikokompetenz, Universität Potsdam
Virchowstr. 2, 14482 Potsdam, Deutschland
Rebitschek@mpib-berlin.mpg.de

Danksagung. Wir danken allen Gutachtern für die konstruktiven Rückmeldungen. Teile der Forschung wurde im Rahmen des Projektvorhabens RisikoAtlas durchgeführt. Die Förderung dieses Vorhabens und des Autors Felix G. Rebitschek erfolgte aus Mitteln des Bundesministeriums der Justiz und für Verbraucherschutz (BMJV) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgte über die Bundesanstalt für Landwirtschaft und Ernährung (BLE) im Rahmen des Programms zur Innovationsförderung.

Funding. Open Access funding provided by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. F.G. Rebitschek und G. Gigerenzer geben an, dass kein Interessenkonflikt besteht.

Psychologische Studien der Autoren mit menschlichen Teilnehmern, die hier zitiert wurden, wurden durch die Ethikkommission des Max-Planck-Instituts für Bildungsforschung geprüft. Sie erhob keine ethischen und wissenschaftlichen Bedenken gegen diese Studie.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Gigerenzer G, Schlegel-Matthies K, Wagner GG (2016) Digitale Welt und Gesundheit: eHealth und mHealth-Chancen und Risiken der Digitalisierung im Gesundheitsbereich (SVRV (Sachverständigenrat für Verbraucherfragen))
- Gerber M, Kraft E, Bosshard C (2016) Overuse – unnötige Behandlungen als Qualitätsproblem. Schweiz Arzteztg 97:236–243
- Gigerenzer G, Wagner GG, Gerberding J et al (2019) Verbraucher-Scoring: Fair und korrekt? In: Reiffenstein M, Blaschek B (Hrsg) Konsumentenpolitisches Jahrbuch: Trends, Rechtsentwicklung und Judikatur der letzten zwei Jahre. Verlag Österreich, Wien, 577–111
- Gigerenzer G (2015) Risk savvy: How to make good decisions. Penguin, London
- Gigerenzer G, Mata J, Frank R (2009) Public knowledge of benefits of breast and prostate cancer screening in Europe. J Natl Cancer Inst 101:1216–1220
- Sørensen K, Van Den Broucke S, Fullam J et al (2012) Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 12:80
- Bittlingmayer UH, Dadaczynski K, Sahrai D, Van Den Broucke S, Okan O (2020) Digitale Gesundheitskompetenz – Konzeptionelle Verortung, Erfassung und Förderung mit Fokus auf Kinder

- und Jugendliche. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. <https://doi.org/10.1007/s00103-019-03087-6>
8. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R (2012) Measuring risk literacy: the Berlin Numeracy test. *Judgm Decis Mak* 7:25–47
 9. Martignon L, Hoffrage U (2019) Wer wagt, gewinnt? Wie Sie die Risikokompetenz von Kindern und Jugendlichen fördern können. Hogrefe, Göttingen
 10. Eysenbach G, Powell J, Kuss O, Sa E-R (2002) Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 287:2691–2700
 11. Mühlhauser I, Oser F (2008) Sind medizinische und Gesundheitsinformationen auf den Internetseiten von Wikipedia evidenzbasiert? – Eine Inhaltsanalyse. *Z Evid Fortbild Qual Gesundheitsw* 102:441–448
 12. Hofmann J, Kien C, Gartlehner G (2015) Vergleichende Evaluierung von Informationsprodukten zu Krebscreening der deutschsprachigen Krebs-Organisationen. *Z Evid Fortbild Qual Gesundheitsw* 109:350–362
 13. Bunge M, Mühlhauser I, Steckelberg A (2010) What constitutes evidence-based patient information? Overview of discussed criteria. *Patient Educ Couns* 78:316–328
 14. Alsaiahi A, Joury A, Aljuaid M, Wazzan M, Pines JM (2017) The content and quality of health information on the Internet for patients and families on adult kidney cancer. *J Cancer Educ* 32:878–884
 15. Guan Y, Maloney KA, Roter DL, Pollin TI (2018) Evaluation of the informational content, readability and comprehensibility of Online health information on monogenic diabetes. *J Genet Couns* 27:608–615
 16. Okan Y, Smith SG, Bruine de Bruin W (2019) How is cervical cancer screening information communicated in UK websites? Cross-sectional analysis of content and quantitative presentation formats. *BMJ Open* 9:e29551
 17. Storino A, Castillo-Angeles M, Watkins AA et al (2016) Assessing the accuracy and readability of online health information for patients with pancreatic cancer. *JAMA Surg* 151:831–837
 18. Rebitschek FG, Gigerenzer G, Jenny MA (in Vorbereitung) Fast-and-frugal trees boost lay people to choose evidence-based health information
 19. Rebitschek FG, Jenny MA (2020) Ein Entscheidungsbaum warnt vor unverlässlichen Gesundheitsinformationen. In: *Jahrbuch der Max-Planck-Gesellschaft* 2019/20. MPG
 20. Stiftung-Hon (2019) HONcode. http://www.hon.ch/HONcode/Webmasters/index_de.html. Zugegriffen: 5. Apr. 2020
 21. Stiftung-Gesundheit (2020) Gütesiegel als Merkmal für Qualität: Vertrauenswürdige Informationen im Web kennzeichnen. <https://www.stiftung-gesundheit.de/zertifiziert-von-der-stiftung-gesundheit-guetesiegel>. Zugegriffen: 5. Apr. 2020
 22. Schaeffer D, Bauer U, Hurrelmann K (2019) Strategiepapier# 5 zu den Empfehlungen des Nationalen Aktionsplans. Gesundheitskompetenz systematisch erforschen
 23. Woopen C, Jünger S, Jannes M (2019) Projekt „Entwicklung einer Orientierungshilfe zur Stärkung der Verbraucherkompetenz beim Umgang mit digitalen Gesundheitsinformationsangeboten (OriGes)“. <https://ceres.uni-koeln.de/forschung/projekte/origes/>. Zugegriffen: 4. Febr. 2020
 24. Zschorlich B, Gechter D, Janßen IM, Swinehart T, Wiegand B, Koch K (2015) Gesundheitsinformatio-
nen im Internet: Wer sucht was, wann und wie? *Z Evid Fortbild Qual Gesundheitsw* 109:144–152
 25. Köpke S, Berger B, Steckelberg A, Meyer G (2005) In Deutschland gebräuchliche Bewertungsinstrumente für Patienteninformationen – eine kritische Analyse. *Z Evid Fortbild Qual Gesundheitsw* 99:353–357
 26. Lühnen J, Albrecht M, Hanßen K, Hildebrandt J, Steckelberg A (2015) Leitlinie evidenzbasierte Gesundheitsinformation: Einblick in die Methodik der Entwicklung und Implementierung. *Z Evid Fortbild Qual Gesundheitsw* 109:159–165
 27. Hertwig R, Grüne-Yanoff T (2017) Nudging and boosting: steering or empowering good decisions. *Perspect Psychol Sci* 12:973–986
 28. Martignon L, Katsipopoulos KV, Woike JK (2008) Categorization with limited resources: A family of simple heuristics. *J Math Psychol* 52:352–361
 29. Green L, Mehr DR (1997) What alters physicians' decisions to admit to the coronary care unit? *J Fam Pract* 45:219–226
 30. Aikman D, Galesic M, Gigerenzer G et al (2014) Taking uncertainty seriously: simplicity versus complexity in financial regulation. Bank of England financial stability paper, Bd. 28
 31. Wangler J, Jansky M (2019) Internetassoziierte Gesundheitsängste in der hausärztlichen Versorgung – Ergebnisse einer Befragung unter Allgemeinmedizinern und hausärztlich tätigen Internisten in Hessen. *Dtsch Med Wochenschr* 144:e102–e108
 32. Informatik GF (2018) Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen
 33. Gibson G (2019) On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet* 15:e1008060
 34. Rebitschek FG, Reisel D, Lein I, Wegwarth O (2019) Epigenetic risk assessment of female cancers: women's information needs and attitudes. *Public Health Genomics* 22:46–57
 35. Gigerenzer G, Rebitschek FG, Wagner GG (2018) Eine vermessene Gesellschaft braucht Transparenz. *Wirtschaftsdienst* 98:860–868
 36. Liu X, Faes L, Kale AU et al (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 1:e271–e297
 37. Gurovich Y, Hanani Y, Bar O et al (2019) Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 25:60
 38. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci USA* 110:5802–5805
 39. Epp C, Lippold M, Mandryk RL (2011) Identifying emotional states using keystroke dynamics. In: *Proceedings of the Sigchi conference on human factors in computing systems*. ACM, New York, 5715–724
 40. Gigerenzer G (2017) Can search engine data predict pancreatic cancer? *BMJ* 358:j3159
 41. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N (2015) Intelligent models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, 5 1721–1730
 42. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366:447–453
 43. Tomašev N, Glorot X, Rae JW et al (2019) A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572:116–119
 44. Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: frequency formats. *Psychol Rev* 102:684–704
 45. McDowell M, Jacobs P (2017) Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychol Bull* 143:1273
 46. Zhu L, Gigerenzer G (2006) Children can solve Bayesian problems: The role of representation in mental computation. *Cognition* 98:287–308
 47. Galesic M, Garcia-Retamero R, Gigerenzer G (2009) Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychol* 28:210–216
 48. McDowell M, Rebitschek FG, Gigerenzer G, Wegwarth O (2016) A Simple Tool for Communicating the Benefits and Harms of Health Interventions. *MDM Policy Pract* 1:2381468316665365
 49. Chalmers I, Oxman AD, Austvoll-Dahlgren A et al (2018) Key Concepts for Informed Health Choices: a framework for helping people learn how to assess treatment claims and make informed choices. *J Evid Based Med* 23:29–33
 50. Fischer S, Petersen T (2018) Was Deutschland über Algorithmen weiß und denkt. Bertelsmann, Gütersloh
 51. Nateqi J, Lin S, Krobath H et al (2019) Vom Symptom zur Diagnose – Tauglichkeit von Symptom-Checkern. *HNO* 67:334–342
 52. Grote Westrick M, Münch I, Volbracht E (2019) Überversorgung – Überflüssige medizinische Leistungen können Patienten schaden. <https://www.bertelsmann-stiftung.de/de/themen/aktuelle-meldungen/2019/november/ueberversorgung-schadet-den-patienten/>. Zugegriffen: 5. Febr. 2020
 53. Tremmel S, Gieselmann H, Eikenberg R (2019) Massive Datenschutzmängel in der Gesundheits-App Ada. In: *c't*. <https://www.heise.de/ct/artikel/Massive-Datenschutzmaengel-in-der-Gesundheits-App-Ada-4549354.html>. Zugegriffen: 21. Apr. 2020