



Methodisch-technische Aspekte der Evaluation erweiterten Zusammenwirkens

Reinhold Haux und Nicole C. Karafyllis

Zusammenfassung

Im Querschnittsfeld Technik und Medizin wird aus Informatik und Philosophie die Frage bearbeitet, welche Werte und Evaluationskriterien beim erweiterten Zusammenwirken von Menschen und Maschinen zu berücksichtigen sind. Das in der VDI-Richtlinie 3780 zur Technikbewertung enthaltene Werteoktagon, das acht grundlegende Werte technischen Handelns zueinander in Beziehung setzt und das Abwägungsentscheidungen für Politik und Gesellschaft ermöglichen soll, wird besprochen. Nach Einführung der aktuell verwendeten Evaluationsmethodik in der klinischen Medizin, dort insbesondere in der Therapieforschung, wird diskutiert, inwieweit diese Evaluationsansätze sich auch auf Fragen bestmöglicher Diagnostik und Therapie, Prävention und Nachsorge im erweiterten Zusammenwirken von Menschen und Maschinen angewandt werden können. Es wird ausgeführt, dass zu der Evaluation dieses Zusammenwirkens ein hoher interdisziplinärer Forschungsbedarf besteht und dass adäquate Ausbildungsangebote vorhanden sein sollten.

R. Haux (✉)

Peter L. Reichertz Institut für Med. Informatik der TU Braunschweig und der Med.
Hochschule Hannover, Braunschweig, Deutschland

E-Mail: Reinhold.Haux@plri.de

N. C. Karafyllis

Seminar für Philosophie, TU Braunschweig, Braunschweig, Deutschland

E-Mail: n.karafyllis@tu-braunschweig.de

© The Author(s) 2021

R. Haux et al. (Hrsg.), *Zusammenwirken von natürlicher und künstlicher
Intelligenz*, https://doi.org/10.1007/978-3-658-30882-7_13

175

Schlüsselwörter

Evaluation · Zusammenwirken · Natürliche Intelligenz · Menschliche Intelligenz · Künstliche Intelligenz · Technikethik · Technikbewertung · Langzeitstudien

1 Einleitung (Karafyllis)

Mit Blick auf das Querschnittsfeld Technik und Medizin werden die Autoren im *gemischten Doppel* aus Informatik und Philosophie die Frage bearbeiten, welche Werte und Evaluationskriterien beim erweiterten Zusammenwirken lebender und nicht-lebender Systeme bzw. Entitäten als Minimalbedingung zu berücksichtigen sind und wo sich dabei Probleme ergeben. Dies beginnt beim eingedeutschten Begriff *Evaluation*, der inhaltlich nicht klar z. B. von dem der Qualitätssicherung abgegrenzt ist (Ditton 2010). Die Begriffssemantik stammt aus dem Feld der empirischen Bildungsforschung und meint allgemein eine prozess- und ergebnis-orientierte Bewertung, womit sich Anknüpfungspunkte zur unten erläuterten VDI-Richtlinie 3780 (*Technikbewertung*) ergeben. Diese wird jüngst medizinethisch für den Bereich *E-Health* nutzbar gemacht (Groß und Schmidt 2018), was mit ein Grund ist, sie in Abschn. 2 kurz vorzustellen. *E-Health* wiederum versammelt sehr verschiedene digitale Gesundheitstechniken, Anwendungen und Nutzergruppen (von Assistenz-Systemen wie im Beispiel 2 im vierten Abschnitt unten, über Telemedizin bis zu Gesundheitsapps), weshalb sich u. a. Groß und Schmidt für genauere Begriffsklärungen zur Digitalisierung im Medizinbereich, spezifische Technologiebetrachtung und eine problemorientierte statt einer technikinduzierten Technikbewertung aussprechen. Erkenntnisleitend für die Lösungen sollte also das Problem sein und nicht die vorhandene oder avisierte Technik. Dies ist eine Grundsatzentscheidung.

Im gängigen Verständnis ist eine Evaluation auf einen definierten Gegenstand und Zeitraum bezogen und wird von ExpertInnen durchgeführt, die auf Basis empirischer Datenerhebung und anhand präzise fest- und offengelegter Kriterien bewerten. Bei systematischen Informationsbewertungen geschieht dies anhand bestimmter Regeln, etwa zum fach- und sachgerechten Umgang mit Statistiken, zum Datenschutz und zur Einwilligung in die Erhebung personenbezogener Daten durch das Forschungssubjekt *Mensch*, das auch, wenn es als Objekt beforcht wird, Subjekt bleibt. *Werte* bedeuten hier Regeln und Kriterien. Da Evaluationen meist fach- und gegenstandsspezifisch sind, stehen sie normativ im Zusammenhang zunächst mit der *beruflichen* Verantwortung und nicht unbedingt notwendig mit *gesellschaftlicher* Verantwortung. Will man

diese Verantwortungsebene erreichen (und dies sollte man tun), muss man sich explizit für eine *normative* Technikbewertung entscheiden. Das bedeutet, dass Ethik, Wissenschafts- und Gesellschaftstheorie zu integralen Bestandteilen der Evaluation werden. In diesem Verständnis sind Werte den Regeln und Kriterien vorgeschaltet und normativ vorausgesetzt. Dabei wird gesellschaftlich um Begriffe und Konzepte gerungen, etwa um *Gesundheit* und *Lebensqualität*, im Bereich künstlichen Zusammenwirkens auch oft um *Freiheit*. Diese nicht vermeidbare philosophische Spannung innerhalb der *Wertfrage* – Werte sind Kriterien und Regeln sowohl vorausgesetzt wie inhärent – durchzieht auch den vorliegenden Artikel.

Je interdisziplinärer der Forschungsgegenstand und die Zusammensetzung der Expertenkommission, desto eher wird die Ebene gesellschaftlicher Verantwortung erreicht. Gleichzeitig verkomplizieren sich die Regeln und Kriterien der Evaluation; und vorab schon die Frage nach dem konkreten Gegenstand und seiner Normierung a priori. So ist in der Wissenschaftsevaluation jedem das konflikträchtige Beispiel bekannt, dass Natur- und Technikwissenschaftler unter bewertungswürdigen *Publikationen* Artikel in Fachzeitschriften mit peer review verstehen, Geisteswissenschaftler hingegen meist Monografien in anerkannten Fachverlagen, vulgo: Bücher. Ähnlich hat sich auch der Begriff von *Dissertationsschrift* gewandelt: vom Buch zur kumulativen Verbindung von Fachartikeln sowie von der gedruckten zur elektronischen Publikation. Dass die Digitalisierung mit zum Sterben zahlreicher Fachverlage und Monopolisierung anderer beigetragen hat, mag gleich zu Beginn ihren weitreichenden Einfluss und Rückkopplungseffekt auf tradierte Bewertungskriterien und -methoden zeigen. Dabei ist angesichts der subversiven Wirkung von Digitaltechnologien die Frage, wann, wie und ob die Gestaltung der einzelnen Techniken überhaupt gelingen kann, schon in der Philosophie umstritten. So vergleicht Paul Virilio die Langzeitfolgen mit denen eines atomaren GAUs, u. a. mit Blick auf „strukturbedingte Massenarbeitslosigkeit“ als „Fall out dieser Informationsbombe“. Friedrich Kittler sieht die Lage optimistischer, denn: „Die Computertechnologie ist die einzige Technologie, die ich kenne, die wirklich radikal unprogrammierbar ist, wo ständig neue Sachen gemacht werden könnten, im Unterschied zur Fabrikationsstraße, die damals Henry Ford in Detroit errichtete [...]“ (Kittler 2017, S. 45). Hier wird also der programmierende Mensch als autonom gegenüber den Maschinen und Programmen gesehen und damit seine Handlungsfreiheit betont. Man beachte dennoch den Konjunktiv *könnten*: Denn ob und wie gut neue Sachen gemacht werden, hängt von den verantwortenden Menschen und ihrer Fähigkeit zum rekursiven Lernen ab. Dafür sind Evaluationen und

Technikbewertungen heuristische Werkzeuge, in die Wertentscheidungen eingehen (vgl. die Beiträge in Rapp 1999) – zuvorderst die Entscheidung, die Bedingungen von menschlicher Autonomie erhalten zu *wollen*.

Die Wissenschaftsphilosophie des 20. Jahrhunderts ist durch die Einsicht geprägt, dass bereits die Einschätzung, was überhaupt *definierter Gegenstand* der jeweiligen Forschung sein soll, Wertentscheidungen voraussetzt. In unserer auf die Medizin ausgerichteten Betrachtung, aber auch in der Bildungsforschung, meint dies nicht weniger als eine Antwort auf die Kantische Frage: Was ist der Mensch? Der gesunde Mensch ist aber ebenso wenig zu definieren wie der gebildete; es handelt sich also nicht um Absoluta, sondern um begriffliche Näherungen, die auf Maßverhältnissen beruhen und sich in Normen und Standards zunächst praktisch bewährt haben, bevor sie mit naturwissenschaftlich-mathematischen Methoden ‘*objektivier*’ und dann informatisch technisiert werden (s. u.). Auch die Auswahl der Evaluationskriterien ist wertgeladen. Sie fußt u. a. auf dem Wissenschaftskriterium der Relevanz und damit auf denjenigen Erkenntnisinteressen, die eine Forschungsgemeinde zu einem bestimmten Zeitpunkt für relevant hält. Wir haben es also mit *verstecktem* normativem, zum Teil auch historisch kontingentem Wissen auf allen Ebenen der Bewertung zu tun, zuvorderst bei der Entscheidung, überhaupt etwas bewerten zu *wollen* – oder dies zu unterlassen. Dabei gilt: Auch unterlassene Handlungen sind zu verantworten (Birnbacher 1995). Diese Aussage ist für das Folgende wichtig, weil der beschleunigte technologische Wandel von Gesellschaften unserer Ansicht nach begleitende Evaluationen mit angemessener Methodik benötigt, um den Wandel aktiv und verantwortungsbewusst gestalten zu können. Das bedeutet, dass ein an juristischen Haftungsfragen orientiertes Handeln für die Übernahme von Verantwortung nicht ausreicht, sondern Sorge und Vorsorge einzuschließen hat. Ansonsten handelt es sich um dasjenige, was der Soziologe Wolfgang Krohn (2007) als „Realexperimente“ mit Gesellschaft kennzeichnet und für die paradigmatisch der wissenschaftliche und politische Umgang mit der frühen Atomtechnologie und der Medikalisation (Arzneimittelforschung) steht. Dabei wird imaginär das Labor¹ auf die ganze Gesellschaft oder auch die Natur ausgedehnt, ohne noch eine schützende Trennwand für die unerwarteten Nebenfolgen der Experimente zur Verfügung zu haben.

Strenggenommen gibt es sie also gar nicht: die bei Evaluationen vorausgesetzte Trennung von vorgeschalteter empirischer (datenerhebender) und

¹Unter Einschluss des (natur)wissenschaftlichen Reduktionismus.

nachgeschalteter bewertender Ebene. Deshalb führt auch die gängige Vorstellung ins Leere, dass Ingenieurinnen und Informatiker zunächst wertfrei forschen und Modelle, Prototypen oder ähnliches entwickeln und Ethikerinnen und Philosophen dann die Produkte technischen Handelns bewerten (vgl. Ropohl 2009). Dass Ingenieure die Verantwortung den politischen Entscheidern überlassen (sollen), sehen jene selbst als kritisch (vgl. Kammeyer 2014). Vergleichbares gilt auch für Mediziner und Informatikerinnen. Notwendig ist deshalb eine Ethik *in* den Wissenschaften (Ammicht et al. 2015), denn Moral hat keine Arbeitsteilung, sondern betrifft jeden einzelnen Menschen – auch in seiner Funktion als Wissenschaftler und Wissenschaftlerin. In diesem Sinne meinte der Informatiker und Vordenker am MIT Joseph Weizenbaum, dass „jeder einzelne für die ganze Welt verantwortlich ist“ (Weizenbaum 1977, S. 349). Aber dafür braucht er normative Hilfestellungen.

2 Technikbewertung (Karafyllis)

Es war diese Einsicht, die den Verein Deutscher Ingenieure (VDI) zu Beginn der 1970er Jahre bewog, eine Richtlinie zur Technikbewertung zu etablieren (zur Historie, König 2013). Sie sollte sich, gerade eingedenk der Beiträge von deutschen Ingenieuren zum Vernichtungskrieg, an den Ingenieur und sein Handeln richten, aber so offen formuliert sein, dass sie keine starre Handlungsanleitung ist. So wurde gewährleistet, dass man auf neue technische Entwicklungen reagieren kann, z. B. aktuell die Digitalisierung. Im zuständigen Ausschuss waren neben Ingenieuren zahlreiche Philosophen vertreten,² die unter *Bewertung* keine rigide DIN-Norm oder ähnliches verstanden wissen wollten und um diejenigen versteckten, da gesellschaftstheoretisch disponierten Normen und Werturteile stritten, die im ersten Abschnitt nur angedeutet werden konnten. Es dauerte fast 20 Jahre, bis die VDI-Richtlinie 3780 „Technikbewertung“ 1991 offiziell verabschiedet wurde (VDI 1991, Nachdruck 2000). Sie bildet bis heute den berufsständischen Codex der deutschen Ingenieurinnen und Ingenieure sowie die Grundlage fortgesetzter technikethischer Reflexion zum Umgang mit ihr. Seit 2002 wird sie um die „Ethische[n] Grundsätze des Ingenieurberufs“ ergänzt (VDI 2002), die sich mit einem Berufseid am tugendethischen Vorbild

²Das relevante VDI-Gremium hieß damals Hauptgruppe *Mensch und Technik*, beteiligte Philosophen waren u. a. Friedrich Rapp, Hans-Heinz Holz, Alois Huning und Günter Ropohl.

des Hippokratischen Eids in der Medizin orientieren (weiterführend Hubig und Reidel 2003). In der kanadischen *Society for Civil Engineering* hat der Berufseid schon seit 1925 Tradition: Beim Aufnahmeeritus mit Schwur zur Anerkennung der berufsständischen Werte und Normen wird ein eiserner Ring (Iron Ring) verliehen. Dieser ist am kleinen Finger der Arbeitshand zu tragen und erinnert jeden Tag an die professionelle Demut im Umgang mit der Technik. Die ersten Ringe sollen aus den Eisenträgern der 1907 eingestürzten Quebec Brücke gewesen sein, Resultat von Planungs- und Konstruktionsfehlern mit Tod zahlreicher Menschen.

Der argumentative Ausgangspunkt der VDI-Richtlinie ist folgender: „Die Existenz und die Beschaffenheit der technischen Mittel gehen auf menschliche Zielsetzungen, Entscheidungen und Handlungen zurück, in denen Werte zum Ausdruck kommen.“ (VDI 1991, S. 341). Hier wird klar: Technik ist nicht wertfrei und der Mensch ist für die Technik verantwortlich. Weil also Ingenieure nicht nur Artefakte, sondern auch Werte erzeugen und verändern, müssen sie ihr Wertebewusstsein aktiv schulen. Dies wird in den formulierten Zielen der Richtlinie deutlich, unter anderem:

- IngenieurInnen für ihre gesellschaftliche Verantwortung sensibilisieren,
- Zielkonflikte zwischen verschiedenen Werten der Technik verdeutlichen,
- Abwägungsentscheidungen zur Wahl der angemessenen Technik ermöglichen.

Wenn Ingenieure Abwägungsentscheidungen auch für Politik und Gesellschaft ermöglichen sollen, müssen sie in Alternativen denken können. Wie aber kann man Techniken abwägen? Dazu legt die Richtlinie ein Werteoktagon vor, das acht grundlegende Werte technischen Handelns zueinander in Beziehung setzt (s. Abb. 1): Funktionsfähigkeit, Sicherheit, Gesundheit, Wirtschaftlichkeit, Umweltqualität, Wohlstand, Persönlichkeitsentfaltung und Gesellschaftsqualität.

Zwischen diesen Werten gibt es häufig Konkurrenzbeziehungen in der Zielerreichung, z. B. zwischen Wirtschaftlichkeit und Umweltqualität. Eine befördernde Instrumentalbeziehung wird hingegen zwischen Sicherheit und Gesellschaftsqualität gesehen, v. a. was die Unterkategorie der Versorgungssicherheit betrifft (die VDI-Richtlinie entstand zur Zeit der Ölkrisen). Heute würde man angesichts umfassender, digitaler Überwachungstechnologien nicht mehr nur eine generell förderliche Beziehung ausmachen. Hier wird deutlich, dass auch die Relationen zwischen Werten im Urteil historisch kontingent sein können. Generelles Ziel bleibt es, Ingenieure für ihre Verantwortung bezüglich möglicher Technikfolgen weiträumig zu sensibilisieren. Dafür liefert die VDI-Richtlinie einen guten Überblick über das gesellschaftliche Wertengeflecht,

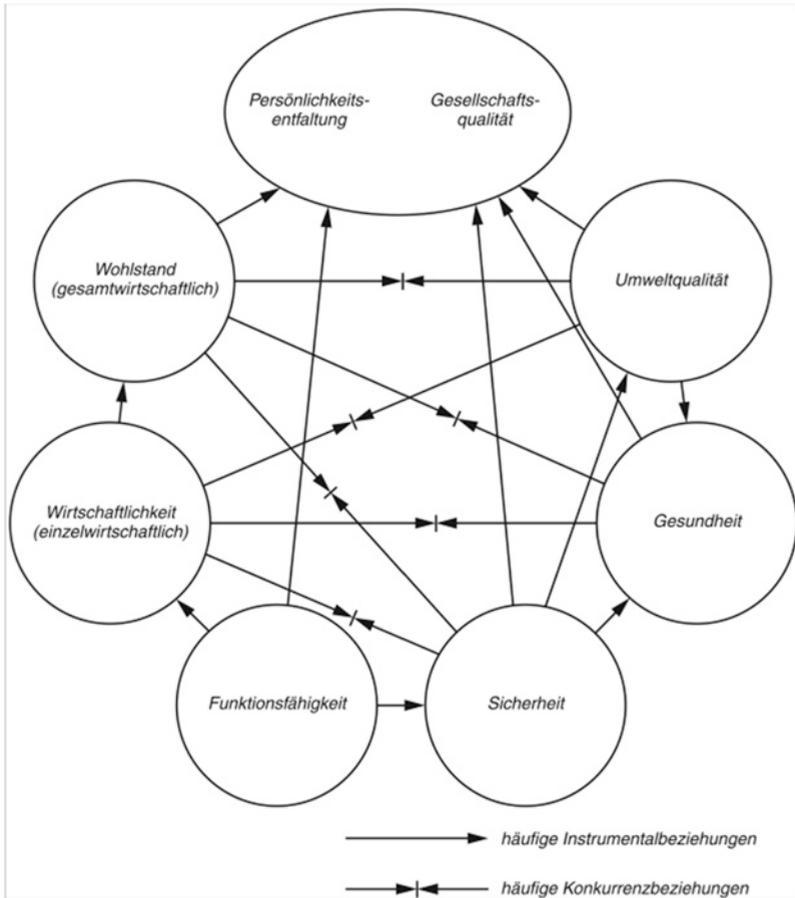


Abb. 1 Wertektogon der VDI-Richtlinie 3780. (Quelle: VDI 2000, S. 21)

aber sie erlaubt noch keine moralischen Priorisierungen, welchen Werten im Konfliktfall der Vorzug zu geben ist. Dies leisten erst die 2002 vorgelegten „Ethischen Grundsätze im Ingenieurberuf“ (VDI 2002). Im Mittelpunkt steht das *Prinzip des Bedingungserhalts*. Es bedeutet, dass prioritär immer Menschen und demokratische Gesellschaften als Bedingungen jeglichen technischen Handelns zu schützen sind. Im Wortlaut (§ 2.4):

„In Wertkonflikten achten Ingenieurinnen und Ingenieure den Vorrang der Menschengerechtigkeit vor einem Eigenrecht der Natur, von Menschenrechten vor Nutzenerwägungen, von öffentlichem Wohl vor privaten Interessen sowie von hinreichender Sicherheit vor Funktionalität und Wirtschaftlichkeit. Dabei sind sie sich bewußt, daß Kriterien und Indikatoren für die unterschiedlichen Wertbereiche nicht dogmatisch vorauszusetzen, sondern nur im Dialog mit der Öffentlichkeit zu ermitteln, abzuwägen und abzugleichen sind.“ (VDI 2002)

3 Evaluationsmethodik in der Medizin ... (Haux)

In der VDI-Richtlinie 3780 ist *Gesundheit* als einer von mehreren Werten im Hinblick auf die Bewertung von Technik aufgeführt (VDI 2000). In dem dort enthaltenen Werteoktagon haben *Sicherheit* und *Umweltqualität* häufige *Instrumentalbeziehungen* zu *Gesundheit*, die selbst eine solche Beziehung zu *Gesellschaftsqualität* und *Persönlichkeitsentfaltung* hat. Zudem hat *Gesundheit* dort häufige *Konkurrenzbeziehungen* zu *Wirtschaftlichkeit* und zu *Wohlstand*. *Gesundheit* wird bezogen auf ihre Bewertung in einen gesamtgesellschaftlichen Zusammenhang gestellt, wobei in dieser Richtlinie die Bewertung ausschließlich auf die Bewertung von Technik ausgerichtet ist.

In der auf den Patienten und auf Gesundheitsversorgung ausgerichteten *klinischen Medizin* hat sich die Vorgehensweise zur Bewertung von *Gesundheit* in den vergangenen Jahrzehnten erheblich weiterentwickelt. Dabei entstand eine spezielle Evaluationsmethodik, die heute dem Fachgebiet *Medizinische Biometrie* bzw. *Medizinische Statistik* zugeordnet werden kann und die im Folgenden vorgestellt werden soll. Evaluation in der klinischen Medizin ist auf den Menschen, insbesondere den Patienten ausgerichtet. Sie schließt die Bewertung von Technik mit ein, falls diese Teil der Gesundheitsversorgung ist, was den o. g. Wert *Versorgungssicherheit* berührt.

Als Medizininformatiker mit engem Bezug zur Medizinischen Biometrie möchte ich am Beispiel der Therapieforschung exemplarisch einen wichtigen, wenn nicht den aktuell wichtigsten Evaluationsansatz in der klinischen Medizin beschreiben. Dies insbesondere deshalb, da dieser auch bei dem „Zusammenwirken lebender und nicht lebender Entitäten im Zeitalter der Digitalisierung“ (BWG-SYnENZ 2020) nicht nur in Medizin und Gesundheitsversorgung, sondern auch in anderen gesellschaftlichen Bereichen eine wichtige Rolle einnehmen könnte und vielleicht auch sollte.

Um dies näher erläutern zu können, wird zunächst über (klinische) Medizin, über Therapieforschung und über deren Evaluationsmethoden berichtet. Deren Bedeutung für die Evaluation des Zusammenwirkens lebender und nicht lebender

Entitäten und damit vor allem des Zusammenwirkens von Menschen und (funktional umfassenden, *intelligenten*) Maschinen wird im nächsten Abschnitt ausgeführt, der damit die Bewertung von Technik wieder notwendig mit umfasst.

Karl Jaspers beschrieb *Medizin* (inklusive Gesundheitsversorgung) folgendermaßen: „Die Medizin dient der Gesundheit, dem Leibeswohl des Einzelnen und der Hygiene der Zustände der gesamten Bevölkerung“ (Jaspers 1946, S. 10 f.). Deshalb stehen Mediziner, v. a. die forschenden an Universitäten, vor der besonderen Aufgabe, den Menschen als individuelles Subjekt, Objekt und überindividuell zu erfassen: „Die medizinische Fakultät lebt in der Spannung der Auffassung des Menschen als Leib, der mit naturwissenschaftlichen Mitteln vollständig zu begreifen ist und dem allein mit diesen geholfen werden kann, und der Kommunikation mit dem Menschen als Freiheit der Existenz, dem ich als Arzt Schicksalsgefährte, nicht mehr nur naturforschender Helfer bin.“ Der forschende Mediziner bleibt wesentlich Arzt mit der dafür notwendigen Empathie und ist deshalb immer mehr als ein Generator, Nutzer oder Anwender von empirischen Daten. Jaspers hebt nicht auf den technischen Aspekt ab, der noch im alten Wort *Heilkunst* zum Ausdruck kommt, wenngleich er mit dem Hinweis auf die Schicksalsgefährten Arzt und Patient ein metaphysisches Element der Beziehung stark macht, das geistesgeschichtlich entweder in der Kunst oder in der Natur verortet wird. Im letzteren Fall ergibt sich eine Brücke zur Vernaturwissenschaftlichung und Technisierung der Medizin, die Schicksal und Zufall versucht, berechenbar und teilweise auch kontrollierbar zu machen. Diese Transformation von der Medizin als heilender Praxis zur Medizin als Wissenschaft ist die theoretische Basis für die unten genannte Einführung statistischer Methoden und darauf basierender Evaluationen. Durch die zunehmende Spezialisierung und Arbeitsteilung in den Gesundheitsberufen, die nicht mehr für alle Patientenkontakt voraussetzen, muss hier die Frage offen bleiben, wie Jaspers' umfassender Berufsethos für den Arzt sich in den verschiedenen heute medizinisch arbeitenden Berufsgruppen durchdeklinieren lässt – sicherlich nur in Allianz mit Berufsethiken für technische Berufe. Wie die „Erhaltung der Gesundheit“ und „Behandlung von Kranken“ (vgl. Gross und Löffler 1997, S. 1), wie Diagnostik und Therapie, Prävention und Nachsorge bestmöglich durchgeführt und kommuniziert werden können, und welche wissenschaftlichen Ansätze passend für deren Bewertung sind, bleibt von zentraler Bedeutung in der medizinischen Forschung und für die Praxis der Gesundheitsversorgung.

In der *klinischen Medizin* spielen aufgrund der Komplexität und Variabilität des Menschen empirische Evaluationsansätze eine besonders wichtige Rolle: „Jeder Mensch ist einmalig [...] Deshalb können wir nicht erwarten, dass diagnostische Verfahren immer den richtigen Befund liefern und Therapien

immer gleich wirken“ (Gaus und Muche 2013, S. 5). Die traditionell subjektbasierte Beurteilung wurde zunehmend kritisch gesehen und als unwissenschaftlich eingestuft. So schrieb Carl Reinhold August Wunderlich³ schon Mitte des 19. Jahrhunderts: „Jeder Arzt sollte Statistiker sein. Jeder Arzt soll Buch führen über Erfolge und Nichterfolge“ (Wunderlich 1851, S. 110).

Im 20. Jahrhundert gewann die auf empirischen Ansätzen basierende klinische Forschung weiter an Bedeutung. Diese war und ist in Deutschland eng verbunden mit Paul Martini⁴ und seiner Methodenlehre (Martini 1932). Nicht ohne Grund stammt der erste Artikel der ältesten, auf Medizinische Informatik und Biometrie spezialisierten Zeitschrift ‚Methods of Information in Medicine‘ (McCray et al. 2011) von Paul Martini. Er befasste sich mit der Methodik der „therapeutisch-klinischen Versuchsplanung“ (Martini 1962). In seiner Aussage „Das Grundgesetz jeder therapeutisch-klinischen Versuchsanordnung ist der therapeutisch-klinische Vergleich“ wird eine wichtige Erkenntnis zusammengefasst: Aufgrund der Komplexität des Menschen lassen sich Erkenntnisse, *wie* Gesundheitsversorgung bestmöglich ausgestaltet werden kann und *welche* therapeutischen Maßnahmen bestmöglich für Patienten geeignet sind, am besten durch den fairen *Vergleich* mehrerer Maßnahmen gewinnen.

In der zweiten Hälfte des 20. Jahrhunderts etablierte sich eine auf dem fairen Vergleich basierende Therapieforchung mittels *kontrollierter klinischer Studien*, die in erheblichem Maße zum medizinischen Fortschritt beitrug. Grundvoraussetzung des fairen Vergleichs ist die sogenannte *Strukturgleichheit*: Bis auf die zu untersuchenden Therapien sollen alle anderen Eigenschaften von Patienten, die den Therapieerfolg beeinflussen könnten (ob bekannt oder unbekannt) in den jeweiligen Therapiegruppen möglichst gleich verteilt sein (Leiner et al. 2012, S. 231). Die *Randomisation*, die *streng zufällige Zuteilung* von Patienten zu Therapien, entwickelte sich zur Methode der Wahl für diesen fairen Vergleich. Eine systematische Planung von Studien, formalisierte Methoden (insbesondere statistische Hypothesen-Tests) und rechnerbasierte Auswertungssysteme trugen dazu bei und markieren einen ersten Höhepunkt für die 1970er Jahre – in dieses Jahrzehnt fällt nicht nur die erste *Digitalisierungswelle* der Wissenschaften, sondern auch die sog. *Medikalisierung* in Industriegesellschaften.

³Arzt, von 1850–1877 Professor an der Medizinischen Klinik an der Universität Leipzig.

⁴Arzt, von 1932–1959 Professor für Innere Medizin an der Universität Bonn.

Exemplarisch seien für die methodischen Neuerungen genannt: das Lehrbuch über Medizinische Statistik von Herbert Immich⁵ (1974), das unter Federführung von Hans Joachim Jesdinsky⁶ (1978) entstandene „Memorandum zur Planung und Durchführung kontrollierter klinischer Studien“ (Jesdinsky 1978) und das kurz darauf unter Federführung von Norbert Victor⁷ herausgegebene Buch über Therapiestudien (Victor et al. 1981). Im dortigen Aufsatz über Therapiestudien von Karl Überla⁸ wird ausgeführt: „Der wesentliche Bestandteil der empirischen Erkenntnisgewinnung ist die Wiederholung derselben Ereignisse unter den gleichen Bedingungen“ sowie: „Im biologischen Vergleich treten die Ereignisse nicht mit schöner Regelmäßigkeit auf wie das Aufgehen der Sonne. [...] Therapiestudien sind der Versuch, mit dieser Variabilität, die einen hilflos läßt, rational fertig zu werden“ (Überla 1981, S. 10).

In einem jüngeren Überblick über die „Entwicklung klinischer Studien von Paul Martini bis heute“ schrieb Martin Schumacher: „Der Prototyp der randomisierten klinischen Studie hat als ‚Gold-Standard‘ in den letzten 40 Jahren das Bild der klinischen Therapieforschung weltweit und auch in Deutschland bestimmt.“ (Schumacher 2016).

Letztendlich ging es bei diesen Evaluationsansätzen darum, trotz der Einmaligkeit eines jeden Menschen in der klinischen Medizin Entscheidungen auf wissenschaftlicher Basis⁹ herbeizuführen, dies zum Wohl des Patienten. Dabei wurden der Patient und die Patientin notwendig zu einem standardisierbaren Typus.

Eine Typisierung klinischer Studien, wie sie von Gaus und Muche (2013) vorgeschlagen wird, ist in Abb. 2 dargestellt. Kontrollierte klinische Studien sind dort bei Eingriffen in Diagnostik und Therapie als *Interventionsstudien* und bei Fragestellungen als *Therapien beurteilen* eingeordnet. Für die Planung, Durchführung und Auswertung klinischer Studien, die gerade bei Therapiestudien

⁵Arzt, von 1971–1982 Professor für Medizinische Statistik, Dokumentation und Datenverarbeitung an der Universität Heidelberg.

⁶Arzt, Professor für Biomathematik und Statistik an der Universität Düsseldorf.

⁷Mathematiker, damals Professor für Biomathematik an der Universität Gießen, von 1983–2007 Professor für Medizinische Biometrie und Informatik an der Universität Heidelberg.

⁸Arzt und Psychologe, von 1974–2005 Professor für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie an der Ludwig Maximilians-Universität München und von 1981 bis 1985 Präsident des Bundesgesundheitsamtes.

⁹Kriterien dazu, u. a. Reproduzierbarkeit, sind in Haux (2003) beschrieben. Dort befindet sich auch weitere Literatur zu dieser Thematik.

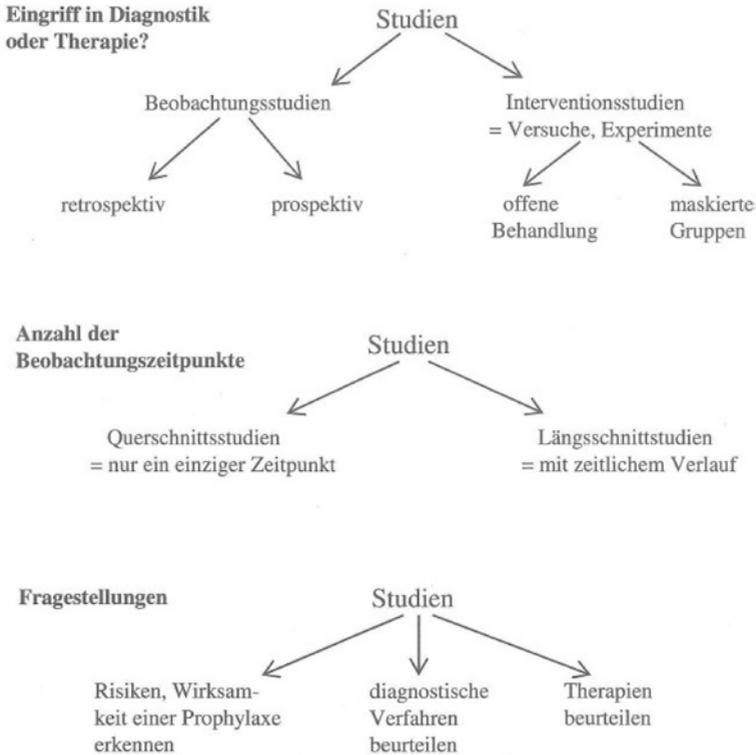


Abb. 2 Typen klinischer Studien anhand von drei Einteilungskriterien. Kontrollierte klinische Studien sind dort bei Eingriff in Diagnostik und Therapie als Interventionsstudien und bei Fragestellungen als Therapien beurteilen beschrieben. (Quelle: Gaus und Muche 2013, S. 38)

ja Experimente am Menschen sind, gelten u. a. strenge ethische Kriterien, für die die *Deklaration von Helsinki* (WMA 1964) eine wichtige Grundlage bildet. Bevor solche Studien durchgeführt werden, müssen hierfür eingerichtete Ethikkommissionen den Studienplänen zustimmen. Bei Studien mit Arzneimitteln müssen klinische Prüfungen zunächst die Phasen I (Pharmakokinetik, Dosis-Wirkungsbeziehung) und II (Verträglichkeit, prinzipielle Wirksamkeit) erfolgreich durchlaufen haben, bevor in einer Phase III mittels kontrollierter klinischer Studien ein Wirksamkeitsnachweis an einer größeren Zahl von

Patienten untersucht werden darf (Gaus und Muche 2013, S. 51). Erst danach darf ein Medikament in Deutschland zugelassen werden. Eine Phase IV nach Zulassung dient dann u. a. der Arzneimittelüberwachung. Zuständige Behörden, die hierfür eingerichtet wurden, sind in Deutschland das *Bundesinstitut für Arzneimittel und Medizinprodukte* (BfArM) und in der Europäischen Union die *European Medicine Agency* (EMA).

Anhand eines Beispiels, der CAO/ARO/AIO-04-Studie (Rödel et al. 2015)¹⁰, soll das Prinzip kontrollierter klinischer Studien verdeutlicht werden. Die Darstellung ist vereinfacht.

Beispiel 1: CAO/ARO/AIO-04-Studie

Die CAO/ARO/AIO-04-Studie diente der Untersuchung der Hypothese, dass bei Patienten mit fortgeschrittenem Rektumkarzinom die bisher etablierte Kombinationstherapie aus Bestrahlung, Chemotherapie und Operation – nachfolgend Verum genannt – oder dass eine Therapie, bei der zusätzlich das Medikament Oxaliplatin gegeben wird – nachfolgend Novum genannt –, überlegen ist. Zugelassen wurde die Studie von der Ethikkommission der Universität Erlangen-Nürnberg.

In dieser zweiarmligen, multizentrischen klinischen Phase-III-Studie wurden zwischen 2006 und 2010 in 88 beteiligten Zentren Patienten rekrutiert und untersucht. Patienten, die nach ihrer Zustimmung in die Studie aufgenommen werden konnten, mussten mindestens 18 Jahre alt sein und ein histologisch gesichertes, fortgeschrittenes, primäres Rektumkarzinom haben.

Die Randomisierung erfolgte geschichtet nach Schweregraden und Zentren. Primärer Endpunkt für einen Therapieerfolg war das dreijährige krankheitsfreie Überleben. Der Berechnung des Stichprobenumfangs lag zugrunde, dass die neue Therapie die bisherige 3-Jahres-Überlebensrate von 75 % auf mindestens 82 % erhöhen sollte. Bei einem Signifikanzniveau von 5 % (Wahrscheinlichkeit für den sog. Fehler 1. Art (α'), aufgrund der Stichprobe von Patienten Unterschiede nicht zu erkennen, obwohl sie in Wirklichkeit vorhanden sind) und bei einer Trennschärfe von 80 % (damit bei einer Wahrscheinlichkeit von 20 % für den sog. Fehler 2. Art (β'), aufgrund der Stichprobe von Patienten Unterschiede anzunehmen, obwohl sie

¹⁰Das Beispiel verdanke ich dem Direktor des Instituts für Medizininformatik, Biometrie und Epidemiologie, Prof. Olaf Gefeller, in dessen Institut die biometrische Studienplanung und Datenanalyse durchgeführt wurde.

in Wirklichkeit gar nicht vorhanden sind), wurde ein notwendiger Stichprobenumfang von 1200 Patienten berechnet. Als statistischer Test wurde ein geschichteter Log-Rank-Test (geschichtet nach Zentren und Schweregraden) verwendet. Die Null-Hypothese lautete, dass zwischen Verum und Novum keine Unterschiede im Hinblick auf das dreijährige krankheitsfreie Überleben zu erkennen sind. Die Alternativ-Hypothese war, dass die neue Therapie der bisher etablierten Therapie überlegen ist.

Von den Daten der zwischen 2006 und 2010 rekrutierten 1265 Patienten konnten 1236 ausgewertet werden: 613 in der Verum-Gruppe, 623 in der Novum-Gruppe. In der Novum-Gruppe lag die krankheitsfreie Überlebensrate bei 75,9 % während sie in der Verum-Gruppe bei 71,2 % lag. Die Alternativhypothese war anzunehmen.

Das klare Ergebnis einer signifikanten Verbesserung im Hinblick auf das dreijährige krankheitsfreie Überleben in der Gruppe, die eine Kombinationstherapie erhielt, bei der zusätzlich Oxaliplatin gegeben wurde, hatte erhebliche Beachtung gefunden. Die klinischen Leitlinien zur Behandlung von Patienten mit Rektumkarzinom wurden nach dieser Studie entsprechend geändert. ◀

4 ... und ihr Potential für das Zeitalter der Digitalisierung (Haux)

Die Kommission Synergie und Intelligenz der Braunschweigischen Wissenschaftlichen Gesellschaft (BWG-SYnENZ 2020) befasst sich mit dem *Zusammenwirken lebender und nicht lebender Entitäten im Zeitalter der Digitalisierung*, insbesondere um das Zusammenwirken von Menschen und (funktional umfassenden, *intelligenten*) Maschinen. Im Hinblick auf Evaluation ist hier, wie in der VDI-Richtlinie 3780, die Bewertung von Technik wieder notwendig mit eingeschlossen.

Die Kommission benennt drei Anwendungsgebiete erweiterten Zusammenwirkens als Gegenstand ihrer Untersuchungen: erweitertes Zusammenwirken in Medizin und Gesundheitsversorgung, erweitertes Zusammenwirken mit Tieren und Pflanzen in der Landwirtschaft sowie physische und virtuelle Mobilitätsformen im erweiterten Zusammenwirken.

Im erstgenannten Anwendungsgebiet werden, wie in der vorher beschriebenen klinischen Forschung, Evaluationen über Studien durchgeführt, mit kontrollierten Studien als *Gold-Standard*. Dieser Evaluationsansatz lässt sich auch auf Fragen bestmöglicher Diagnostik und Therapie, Prävention und Nachsorge im erweiterten Zusammenwirken von Menschen und Maschinen anwenden.

Zwei Studien sollen dies exemplarisch verdeutlichen¹¹. Auch diese werden wieder vereinfacht dargestellt. In diesen Studien war das Peter L. Reichertz Institut für Medizinische Informatik die verantwortliche Einrichtung für die Studienplanung und Datenanalyse. In unserer Medizininformatik-Forschung arbeiten wir in dem Forschungsschwerpunkt assistierende Gesundheitstechnologien u. a. an ‚intelligenten Wohnungen‘. Als neuer diagnostisch-therapeutischer Raum sollen diese Wohnungen den Menschen assistieren und die Gesundheitsversorgung weiter verbessern (Haux 2016; Haux et al. 2016b; Mielke et al. 2017; Wolf et al. 2017).

Beispiel 2: GAL-NATARS-Studie

In der GAL-NATARS-Studie (Marscholke et al. 2014) ging es um das häusliche Langzeit-Monitoring von geriatrischen Patienten mit mobilitätseinschränkenden Frakturen. Primärziel war die Überprüfung der technischen Machbarkeit und die Akzeptanz technischer Assistenzsysteme, Sekundärziel die Identifikation von mittels assistiver Technologien ermittelten Risiko-Mustern im häuslichen Umfeld, die unter Praxisbedingungen abgeleitet und für das geriatrische Assessment genutzt werden können. Zugelassen wurde die Studie von der Ethik-Kommission der Medizinischen Hochschule Hannover.

An dieser prospektiven multizentrischen Beobachtungsstudie waren drei geriatrische Zentren beteiligt. Aufgenommen wurden Patienten mit mobilitätseinschränkenden Frakturen am Bewegungsapparat, die mindestens 70 Jahre alt waren und die einen Mini-Mental-State-Index von 20 oder mehr Punkten hatten. Bei Einwilligung der Patienten wurde in deren Wohnung für drei Monate nach Entlassung aus der stationären geriatrischen Rehabilitation körper- und insbesondere raumbezogene Sensorik installiert. Zudem wurden diese Personen zusätzlich klinisch-geriatrisch untersucht.

Zwischen 2012 und 2014 wurden 24 Personen mit einem durchschnittlichen Alter von 83,5 Jahren in die Studie aufgenommen. Die Studie belegte klar die technische Machbarkeit unter Praxisbedingungen (unseres Wissens zum ersten Mal). Zudem konnte (unseres Wissens ebenfalls zum ersten Mal) gezeigt werden, dass in Wohnungen über technische Assistenzsysteme Aktivitätsmuster gemessen werden können, die neue wichtige Informationen für eine verbesserte ärztliche und pflegerische Versorgung enthielten (Wang et al. 2014). ◀

¹¹Allgemeinere Überlegungen und Übersichten hierzu befinden sich in Martin-Sanchez (2016), Haux (2010) und Haux et al. (2016a, b).

Beispiel 3: AGT-Reha-Studie

In der AGT-Reha-WK-Studie (Wolf et al. 2016) wird die Frage untersucht, ob bei der poststationären Rehabilitation von Patienten mit Schulterschmerzen ein neu entwickeltes Informatik-Therapeutikum¹², das sogenannte AGT-Reha-System, der bisherigen Standardnachsorge, der medizinischen Trainingstherapie (MTT), gleich wirksam oder sogar überlegen ist (primäres Studienziel). Zugelassen wurde die Studie von der Ethik-Kommission der Medizinischen Hochschule Hannover.

Nachdem eine Vorstudie (sozusagen eine Phase-II-Studie) die technische Machbarkeit von AGT-Reha gezeigt hatte, konnte mit dieser vergleichenden Therapiestudie (sozusagen eine Phase-III-Studie) im Februar 2017 begonnen werden. Sie wurde als monozentrische prospektive Nicht-Unterlegenheitsstudie angelegt. Eingeschlossen sind Patienten im Alter von 18 bis 65 Jahren, die an chronischen Schulterschmerzen (gemäß einer vorgegebenen Liste von ICD-Diagnosen) leiden und die in stationärer Rehabilitation waren. Nach stationärer Entlassung erfolgt unmittelbar anschließend eine sechsmonatige Nachsorge entweder mit AGT-Reha oder mit MTT. Nach Einschluss eines Probanden in die Studie während dessen stationärer Rehabilitation erfolgt die Zuteilung des Rehabilitanden zu einer der zwei Behandlungsgruppen. Hierzu wird der Rehabilitand zunächst gefragt, ob er eine der beiden Nachsorgeformen präferiert. Bevorzugt der Rehabilitand keine der beiden Nachsorgeformen, erfolgt eine randomisierte Zuteilung zu einer der beiden Behandlungsgruppen.

Primäres Erfolgskriterium im Hinblick auf die Wiedererlangung der Arbeitsfähigkeit ist die Bestimmung der Schulterfunktion über den Shoulder-Pain-and-Disability-Index (SPADI). Eine Stichprobenumfangsberechnung ergab, dass 42 Patienten pro Therapiegruppe benötigt werden (α : 0,025, β : 0,2, SPADI-Differenz: 10). Als statistischer Hypothesen-Test wird der t-Test für unabhängige Stichproben verwendet sowie ein Intervallinklusionsverfahren für die Untersuchung der Nicht-Unterlegenheit. Diese Studie befindet sich noch in der Durchführung. Bis September 2019 wurden 69 Patienten in die Studie aufgenommen, von denen bis zu diesem Zeitpunkt bei 30 Patienten die Abschlussuntersuchung durchgeführt werden konnte. ◀

¹²Zum Begriff Informatik-Therapeutikum vgl. Haux (2010, S. 606).

Auch wenn die in den Beispielen 2 und 3 verwendeten Informatik-Werkzeuge noch nicht als funktional umfassende, *intelligente* Maschinen angesehen werden können, ist jedoch abzusehen, dass dieser bei assistierenden Gesundheitstechnologien eingeschlagene Weg dazu führen wird, solche Maschinen zu bekommen. In anderen medizinischen Anwendungsgebieten haben sich schon neue rechtlich-ethische Fragen im erweiterten Zusammenwirken bis hin zu der Frage geteilter Autonomie zwischen Menschen und Maschinen gestellt (Steil et al. 2019). Die beiden Studien zeigen zudem auf, dass auch Technikbewertung im erweiterten Zusammenwirken mit der in der klinischen Forschung entwickelten Evaluationsmethodik durchgeführt werden kann, einschließlich eines fairen Vergleichs von verschiedenen Ansätzen durch vergleichende Interventionsstudien.

Digitalisierung – und mit ihr erweitertes Zusammenwirken – durchdringt alle gesellschaftlichen Bereiche. Aufgrund der positiven Ergebnisse in der klinischen Medizin sollte überlegt werden, ob diese Evaluationsansätze auch generell bei dem Zusammenwirken lebender und nicht lebender Entitäten verwendet werden können oder sollten. Die Einmaligkeit eines jeden Menschen (Gaus und Muche 2013, S. 5) ist auch hier gegeben. Die schon dadurch gegebene Komplexität und Variabilität wird noch vergrößert durch das Zusammenwirken von Menschen und funktional umfassenden, ‚intelligenten‘ Maschinen. Eine nur auf Machbarkeit oder Einzelfalluntersuchungen, nur auf Simulation oder Vorhersageberechnung, aber *ohne vergleichende Intervention* basierende Beurteilung erweiterter Zusammenwirkens würde dieser komplexen Problematik nicht gerecht werden. Wie bei den *klassischen* kontrollierten klinischen Studien geht es auch im erweiterten Zusammenwirken darum, trotz komplexer Sachverhalte, Entscheidungen auf wissenschaftlicher Basis herbeizuführen, dies zum Wohl der Menschen. Nicht zuletzt sollte überlegt werden ob sektorübergreifende Institutionen, ähnlich dem BfArM oder der EMA in der Medizin, für die Überwachung der Prüfung und für die Zulassung von erweitertem Zusammenwirken einzurichten sind.

Eine auf den Erfolgen der Therapieforchung basierende Weiterentwicklung von Ansätzen und von Methoden zur Evaluation, welche wie in der Therapieforchung auch ethische und rechtliche Aspekte einzuschließen hat, sollte frühzeitig angegangen werden.

5 Fazit und Ausblick (Karafyllis und Haux)

Wir haben hier zwei Bewertungsmethoden vorgestellt. Hintergrund war die Idee: Wenn natürliche und künstliche Systeme zusammenwirken, müssen dies auch deren Beurteilungsmaßstäbe (*Werte*) und Evaluationskriterien tun.

Verdeutlicht wurde dies an etablierten Bewertungspraxen von Ingenieurwesen und klinischer Medizin. Zu diesem und weiterem Zusammenwirken besteht hoher interdisziplinärer Forschungsbedarf und das Desiderat transdisziplinärer Methoden.

Als Vorteile der VDI-Richtlinie 3780 „Technikbewertung“ erwiesen sich, dass acht Werte im Wechselspiel betrachtet und Priorisierungen formuliert werden. Positiv ist auch, dass die berufliche Verantwortung hin zur gesellschaftlichen Verantwortung des Ingenieurs erweitert wird. Dadurch wird das Wertegeflecht, in dem Menschen wirklich leben, modelliert. Nachteilig erscheinen unter empirischen Gesichtspunkten der Mangel an statistisch abgesicherten Bewertungen und der unklare zeitliche Horizont der Bewertung. Die Forderung nach Langzeitstudien wird bislang durch die Richtlinie nicht abgedeckt. Das betrifft auch die überprüfbare Zielerreichung des Wertes *Versorgungssicherheit*. Vielmehr wird auf das tugendethische *Ethos des Ingenieurs* abgehoben und damit auf die individuelle Verantwortung, die gegenüber dem Berufsstand und seiner Werte und Normen wie auch gegenüber der Gesellschaft wahrzunehmen ist.

Als Vorteile der Evaluationsmethodik in klinischen Studien erwiesen sich die statistisch breite Basis und die Langzeitorientierung. Der Mensch wird in seiner/ihrer Variabilität betrachtet, in Form eines Typus. Als nachteilig erwies sich der Fokus auf den singulären Wert *Gesundheit*, der etwa Wechselwirkungen mit der Umwelt (z. B. Plastikmüll in Kliniken) und auch der Versorgungssicherheit (Stichwort: Erhalt von Krankenhäusern im ländlichen Raum) nicht berücksichtigt. Durch die mathematisch-informatisch gestützte Entscheidungsgrundlage werden Ärztin und Arzt zunächst moralisch entlastet, bleiben aber für den Dialog mit Patienten auf ihre eigene Urteilskraft und Verantwortung verwiesen. Für die Schulung, wie das Zusammenwirken menschlicher und nicht-menschlicher Systeme das ärztliche und auch das pflegerische Handeln verändert, empfiehlt sich das interdisziplinäre Gespräch mit den Informatikern. Von diesen wiederum ist Selbstreflexion und gesellschaftliches Folgenbewusstsein gefragt, um denjenigen Expertentyp zu vermeiden, den Joseph Weizenbaum bereits 1977 sorgenvoll den „zwanghaften Programmierer“ nannte: (Weizenbaum 1977, S. 155–179).

Als Ausblick möchten wir eine Grundsatzfrage für die weitere Diskussion stellen: Welches *Menschenbild* wird beim Zusammenwirken lebender und nicht-lebender Entitäten (SYnENZ) implementiert? Denn auf diesem basieren auch die Normen, Methoden und Kriterien der Bewertung. Grundsätzlich wird das Menschenbild bei Debatten um das „Zusammenwirken“ zwischen zwei Polen abgehandelt, zu denen es sich bewusst zu verhalten gilt:

- a) Der eine Pol markiert *den Menschen* (als Gattungsbegriff) als sog. *human factor* oder sogar *human error*. Dieser streitbare Begriff ist v. a. in den Ingenieurwissenschaften und der Informatik erkenntnisleitend. Bei der Gestaltung für die Zukunft streben Wissenschaft und Technik fehlerfreie Technologien an, bei denen die Fehleranfälligkeit primär im menschlich defizitären Handeln und Verhalten gesehen wird (Beispiele: automatisiertes Fahren, OP-Roboter). Dies führt, weitergedacht, zur Frage: Sollte der Mensch bei wichtigen Handlungen am besten weiträumig eliminiert werden, z. B. der manchmal fehldiagnostizierende Arzt? Ist eine algorithmenbasierte Diagnostik grundsätzlich wünschenswert? (*Technik ohne Mensch?*). Dann hätte sich der Wert *Sicherheit* verselbstständigt und wäre von einer ethischen Theorie des guten Lebens ebenso entkoppelt wie vom zwischenmenschlichen Dialog. Diese Sicht führt auch dazu, Probleme im Umgang mit Techniken und Produkten als nur *technisch* zu lösende zu konzipieren. Beispiele: Gegen häufige Verkehrsunfälle kann jedoch nicht nur, wenn überhaupt, automatisiertes Fahren helfen, sondern dies können auch engmaschigere Alkoholkontrollen und regelmäßige Fahrtüchtigkeitsüberprüfungen z. B. von Autofahrern höheren Alters, wie sie in vielen EU-Ländern bereits vorgeschrieben sind (nicht aber in Deutschland). Fehldiagnosen von Ärzten sind sicherlich nicht nur durch mangelndes Wissen, sondern auch im ständig steigenden Zeitdruck im Praxisalltag bzw. klinischen Alltag begründet und damit in ökonomischen Sachzwängen des Gesundheitssystems. In summa: Technik und Evaluation können und sollten nicht von politischen, ökonomischen und gesellschaftlichen Urteilen und Entscheidungen entlasten.
- b) Der zweite Pol markiert den Ansatz, *die Menschen* (als Singularetantum) in ihrer physisch-psychischen Variabilität *und* als Grundlage von Gesellschaft zu betrachten. Das Ziel ist nicht etwa ein *Mensch ohne Technik*, sondern das Erreichen einer humanen Technik bzw. angepassten Technologie, die den Menschen als Mittel zum Leben *ihrer* Werte dient. Maßgaben sind dabei sowohl die Komplexität des Menschen wie auch dessen Variabilität. Interdisziplinär geplante und statistisch abgesicherte Langzeitstudien zur Technikfolgenforschung sind notwendig, um das Ziel einer humanen Technik überhaupt anstreben zu *können*, und zwar in all denjenigen vielen neuen Bereichen, in denen lebende und nicht-lebende Systeme zusammenwirken.

Ende 2018 hat ein unabhängiges europäisches Expertengremium „Ethik-Leitlinien zur verantwortungsvollen KI“, d. h. zur Künstlichen Intelligenz, zur

Diskussion gestellt (HEG-KI 2019). Die Verfasser schreiben: „Wir sind der Überzeugung, dass die KI das Potenzial hat, die Gesellschaft signifikant zu transformieren. Die KI ist kein Selbstzweck, sondern ein vielversprechendes Mittel, um das menschliche Gedeihen und somit das Wohlbefinden von Individuum und Gesellschaft und das Gemeinwohl zu steigern sowie zur Förderung von Fortschritt und Innovation beizutragen“ (HEG-KI 2019, Art. 9). Gleichzeitig betonen sie:

„Um dies zu erreichen, müssen KI-Systeme **auf den Menschen ausgerichtet**¹³ sein und auf der verpflichtenden Grundlage stehen, dass ihre Nutzung im Dienste der Menschheit und des Gemeinwohls steht, mit dem Ziel, menschliches Wohl und menschliche Freiheit zu mehren. Obwohl KI-Systeme großartige Chancen bieten, entstehen durch sie auch bestimmte Risiken, die angemessen und verhältnismäßig behandelt werden müssen. Wir haben jetzt die wichtige und günstige Gelegenheit, auf die Entwicklung dieser Systeme gestalterischen Einfluss zu nehmen. Wir wollen gewährleisten, dass wir den sozio-technischen Umgebungen, in die sie eingebettet sind, vertrauen können, und wir wollen erreichen, dass die Hersteller von KI-Systemen dadurch einen Wettbewerbsvorteil erlangen, dass sie die vertrauenswürdige KI in ihre Produkte und Dienstleistungen integrieren.“ (ebd., Art. 10, S. 5).

Trotz der verdienstvollen Arbeit an diesen Ethik-Leitlinien und ihrer Begrüßungswürdigkeit im Einzelnen (z. B. in der Hervorhebung menschlicher Autonomie) muss festgehalten werden: Auch hier wird die Frage nach dem zugrunde liegenden Menschenbild nicht gestellt. So ergeben sich allein mit Blick auf die obigen Passagen zahlreiche Rückfragen zum Verhältnis der Begriffe *menschlich*, *Individuum*, *Gesellschaft* und *Menschheit*; zudem auch *Gemeinschaft*, der dem Terminus *Gemeinwohl* zugrunde liegt und sozialwissenschaftlich von dem der *Gesellschaft* unterschieden wird. Hinzu kommt die Problematik, dass im Verweis auf *Systeme* der KI die Grenzen des jeweils gemeinten Systembegriffs nicht selbsterklärend sind. Der kleinste gemeinsame normative Nenner der Ethik-Leitlinien sind die Menschenrechte, für die aber aus wissenschafts- und technikethischer Sicht gilt, dass sie primär einen größten gemeinsamen Zähler kennzeichnen, d. h. an *Menschheit* ansetzen. Wie *menschlich* die Menschheit in den verschiedenen Ländern, ihren Gesellschaften und Gemeinschaften, konzipiert werden soll, berührt fundamentale Wertfragen und bleibt ein politischer, gesellschaftlicher und vor allem auch juristischer

¹³Fettdruck im Original.

Aushandlungsprozess, der bislang nationalstaatlich institutionalisiert ist. Entsprechend nennen die Autoren das Recht an erster Stelle der drei zentralen *Komponenten* für eine vertrauenswürdige KI: „rechtmäßig“, gefolgt von „ethisch“ und „robust“ (ebd., S. 2).

Die grundlegende Entscheidung zum Menschenbild ist: Betrachten wir eine menschliche oder natürliche Grenze als Hindernis, d. h. als eine, die überwunden werden muss? Dies führt zur Frage nach den ethischen Grenzen der Optimierung des Lebendigen überhaupt, sei sie durch Digitalisierung oder Biotechnologie oder deren Interaktion im Rahmen der Converging Technologies (Bio-, Nano-, Info-, Kogno-Technologien) verursacht. In den Worten Weizenbaums (1977, S. 337 ff.) sollten wir wachsam sein gegenüber dem „Imperialismus der instrumentellen Vernunft“, der durch Digitalisierung und Computertechnologie in besonderer Weise befördert werde. Dies heißt aber nicht, dass wir gesamtgesellschaftlich davon entbunden wären, die bestmögliche Technologie anzustreben und ihr Zielerreichen kontinuierlich zu überprüfen. Die VDI-Richtlinie zur Technikbewertung und der geschilderte Gold-Standard der Evaluation qua klinischer Studien haben beide gezeigt, wie wichtig es ist, in Alternativen bzw. Alternativ-Hypothesen denken zu können. Die interdisziplinäre Ausbildung der zukünftigen Expertinnen und Experten,¹⁴ wie sie einst fast überall durch ein verpflichtendes *Studium generale* angestrebt wurde, bleibt zentrale Aufgabe im Wissenschaftsbereich, um mögliche Folgen eigenen Handelns beim Zusammenwirken menschlicher und nicht-menschlicher Systeme überhaupt in den Blick nehmen zu können.

Literatur

- Ammicht Quinn, R., & Potthast, T. (Hrsg.). (2015). *Ethik in den Wissenschaften: 1 Konzept, 25 Jahre, 50 Perspektiven*. Tübingen: IZEW.
- Braunschweigische Wissenschaftliche Gesellschaft (BWG), & Kommission Synergie und Intelligenz (SYnENZ). (2020). <https://bwg-nds.de/kommissionen/kommission-syrenz/>. Zugegriffen: 30. Jan. 2020.
- Birnbacher, D. (1995). *Tun und Unterlassen*. Stuttgart: Reclam.
- Ditton, H. (2010). Evaluation und Qualitätssicherung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (3. Aufl., S. 607–623). Wiesbaden: Springer VS.

¹⁴In diesem Sinne haben die Fächer Informatik und Philosophie an der TU Braunschweig seit 2017/2018 die Pflichtveranstaltung „Ethik der Technik, Wirtschaft und Information“ in das Curriculum integriert, die auch von Studierenden der Ingenieurwissenschaften besucht werden kann.

- Gaus, W., & Muche, R. (2013). *Medizinische Statistik*. Stuttgart: Schattauer.
- Groß, D., & Schmidt, M. (2018). E-Health und Gesundheitsapps aus medizinethischer Sicht. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 61(3), 349–357.
- Gross, R., & Löffler, M. (1997). *Prinzipien der Medizin*. Berlin: Springer.
- Haux, R., et al. (2003). Kriterien für gute medizinische Forschung. In W. Eich (Hrsg.), *Wissenschaftlichkeit in der Medizin, Teil IV* (S. 181–201). Frankfurt a. M.: VAS.
- Haux, R. (2010). Medical informatics: Past, present, future. *International Journal of Medical Informatics*, 79, 599–610.
- Haux, R. (2016). My home is my hospital. On recent research on health-enabling technologies. *Studies in Health Technology and Informatics*, 226, 3–8.
- Haux, R., Koch, S., Lovell, N. H., Marschollek, M., Nakashima, N., & Wolf, K.-H. (2016). Health-enabling and ambient assistive technologies: Past, present, future. *IMIA Yearbook of Medical Informatics*, 25(Suppl. 1), 76–91.
- Haux, R., Marschollek, M., & Wolf, K.-H. (2016). Über assistierende Gesundheitstechnologien und neue Formen kooperativer Gesundheitsversorgung durch Menschen und Maschinen. In A. Manzeschke (Hrsg.), *Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen?* (S. 131–143). Baden-Baden: Nomos.
- HEG-AI (Hochrangige Expertengruppe für Künstliche Intelligenz). (2019). Ethik-Leitlinien für eine vertrauenswürdige KI. Brüssel: Europäische Kommission (Entwurfssfassung vorab veröffentlicht 18.12.2018; abgestimmte Fassung 10.4.2019).
- Hubig, C., & Reidel, J. (Hrsg.). (2003). *Ethische Ingenieurverantwortung Handlungsspielräume und Perspektiven der Kodifizierung*. Berlin: Edition Sigma.
- Immhof, H. (1974). *Medizinische Statistik*. Stuttgart: Schattauer.
- Jaspers, K. (1946). *Vom lebendigen Geist der Universität*. Heidelberg: Schneider.
- Jesdinsky, H. J. (Hrsg.). (1978). *Memorandum zur Planung und Durchführung kontrollierter klinischer Studien*. Stuttgart: Schattauer.
- Kammeyer, H.-U. (2014). Grundsätzliches zur Ethik für Ingenieure. In L. Hieber & H.-U. Kammeyer (Hrsg.), *Verantwortung von Ingenieurinnen und Ingenieuren* (S. 33–37). Wiesbaden: Springer VS.
- Kittler, F. (2017). Die Informationsbombe. Gespräch mit Paul Virilio. In T. Baumgärtel (Hrsg.), *Texte zur Theorie des Internets* (S. 41–54). Stuttgart: Reclam.
- König, W. (2013). VDI-Richtlinie zur Technikbewertung. In A. Grunwald (Hrsg.), *Handbuch Technikethik* (S. 406–410). Metzler: Stuttgart.
- Krohn, W. (2007). Realexperimente – Die Modernisierung der ‚offenen Gesellschaft‘ durch experimentelle Forschung. *Erwägen Wissen Ethik*, 18(3), 343–356.
- Leiner, F., Gaus, W., Haux, R., Knaup, P., Pfeiffer, K. P., & Wagner, J. (2012). *Medizinische Dokumentation* (6. Aufl.). Stuttgart: Schattauer.
- McCray, A. T., Gefeller, O., Aronsky, D., Leong, T. Y., Sarkar, I. N., Bergemann, D., et al. (2011). The birth and evolution of a discipline devoted to information in biomedicine and health care. As reflected in its longest running journal. *Methods of Information in Medicine*, 50, 491–507.
- Marschollek, M., Becker, M., Bauer, J., Bente, P., Dasenbrock, L., Elbers, K., Hein, A., Kolb, G., Künemund, H., Lammel-Polchau, C., Meis, M., Meyer zu Schwabedissen, H., Remmers, H., Schulze, M., Steen, E.-E., Thoben, W., Wang, J., & Haux, R. (2014).

- Multimodal activity monitoring for home rehabilitation of geriatric fracture patients – Feasibility and acceptance of sensor systems in the GAL-NATARS-Study. *Informatics for Health and Social Care*, 39, 262–271.
- Martini, P. (1932). *Methodenlehre der Therapeutischen Untersuchung*. Berlin: Springer.
- Martini, P. (1962). Grundsätzliches zur therapeutisch-klinischen Versuchsplanung. *Methods of Information in Medicine*, 1, 1–5.
- Martin-Sanchez, F. J., & Lopez-Campos, G. H. (2016). The new role of biomedical informatics in the age of digital medicine. *Methods of Information in Medicine*, 55, 392–402.
- Mielke, C., Voss, T., & Haux, R. (2017). Residence as a diagnostic and therapeutic area: A smart home approach. *Studies in Health Technology and Informatics*, 238, 92–95.
- Rapp, F. (Hrsg.). (1999). *Normative Technikbewertung. Wertprobleme der Technik und die Erfahrungen mit der VDI-Richtlinie 3780*. Berlin: Edition Sigma.
- Rödel, C., Graeven, U., Fietkau, R., Hohenberger, W., Hothorn, T., Arnold, D., Hofheinz, R. D., Ghadimi, M., Wolff, H. A., Lang-Welzenbach, M., Raab, H. R., Wittekind, C., Ströbel, P., Staib, L., Wilhelm, M., Grabenbauer, G. G., Hoffmanns, H., Lindemann, F., Schlenska-Lange, A., Folprecht, G., Sauer, R., Liersch, T., & German Rectal Cancer Study Group. (2015). Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): Final results of the multicentre, open-label, randomised, phase 3 trial. *Lancet Oncology*, 16, 979–989.
- Ropohl, G. (2009). *Allgemeine Technologie* (3. Aufl.). Karlsruhe: KIT.
- Schumacher, M. (2016). Entwicklung klinischer Studien von Paul Martini bis heute. *Drug Research*, 66(S01), 5–7.
- Steil, J., Finas, D., Beck, S., Manzeschke, A., & Haux, R. (2019). Robotic systems in operating theatres: New forms of team-machine interaction in health care. On challenges for health information systems on adequately considering hybrid action of humans and machines. *Methods of Information in Medicine*, 58, e14–e25.
- Überla, K. K., et al. (1981). Therapiestudien: Indikation, Erkenntniswert und Herausforderung. In N. Victor (Hrsg.), *Therapiestudien* (S. 7–21). Berlin: Springer.
- VDI – Verein Deutscher Ingenieure. (1991). *VDI-Richtlinie 3780: Technikbewertung – Begriffe und Grundlagen*. Düsseldorf: VDI.
- VDI – Verein Deutscher Ingenieure. (2000). *VDI-Richtlinie 3780: Technikbewertung – Begriffe und Grundlagen*. Berlin: Beuth (textidentisch mit Ausg. 1991).
- VDI – Verein Deutscher Ingenieure. (2002). *Ethische Grundsätze des Ingenieurberufs*. Düsseldorf: VDI. <https://www.vdi.de/ueber-uns/presse/publikationen/details/ethische-grundsätze-des-ingenieurberufs>.
- Victor, N., Dudeck, J., & Broszio, E. P. (Hrsg.). (1981). *Therapiestudien*. Berlin: Springer.
- Weizenbaum, J. (1977). *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt a. M.: Suhrkamp.
- Wang, J., Bauer, J., Becker, M., Bente, P., Dasenbrock, L., Elbers, K., Hein, A., Kohlmann, M., Kolb, G., Lammel-Polchau, C., Marscholke, M., Meis, M., Remmers H., Meyer zu Schwabedissen, H., Schulze, M., Steen, E.-E., Haux, R., & Wolf, K.-H. (2014). A novel approach for discovering human behavior patterns using unsupervised methods. *Zeitschrift für Gerontologie und Geriatrie*, 47, 648–660.

- Wolf, K.-H., et al., Studiengruppe AGT-Reha. (2016). Evaluation der Wirksamkeit und Kosten der poststationären häuslichen Tele-Rehabilitation mit AGT-Reha im Vergleich zur Medizinischen Trainingstherapie. Bericht mit Studienplan.
- Wolf, K.-H., Dehling, T., Haux, R., Sick, B., Sunyaev, A., & Tomforde, S. (2017). On methodological and technological challenges for proactive health management in smart homes. *Studies in Health Technology and Informatics*, 238, 209–212.
- World Medical Association (WMA). (1964). Declaration of Helsinki – Ethical principles for medical research involving human subjects. Adopted 1964, last amendment 2013. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>. Zugegriffen: 30. Jan. 2020.
- Wunderlich, C. A. (1851). Ein Plan zur festeren Begründung der therapeutischen Erfahrungen. *Jahrbücher der gesamten Medicin*, 69, 106–111.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) 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 Kapitel 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.

