



Schaden & Unfall

Validierung – weshalb sie in der Behandlung so wichtig ist wie in der Begutachtung

von Prof. Dr. med. Andreas Stevens, Facharzt für Neurologie, Psychiatrie und Psychotherapie, Medizinisches Begutachtungsinstitut und Universität Tübingen

1. Einleitung

Um eine richtige Entscheidung zu treffen und die richtige (Be-)Handlung auszuführen, muss die Wirklichkeit möglichst gut erfasst sein. Dies betrifft bei Gesundheitsstörungen den behandelnden Arzt ebenso wie den Gutachter und – bei Rechtsstreitigkeiten – den Richter. Nach psychologischen Studien unterliegen Menschen einem sogenannten Wahrheitsbias, d. h., sie unterstellen Aussagen zunächst, dass sie wahr sind (Grice 1981, Vrij and Baxter 1999). Im Alltag wird unser Denken und Handeln durch eine Reihe solcher Biases, also systematische kognitive Verzerrungen, geleitet, die es uns ermöglichen, rasch und mühelos, ohne rationale und damit aufwendige Prüfung, eine Entscheidung zu treffen. Biases bzw. Heuristiken sind deshalb im Alltag wichtig und aus Sicht der Evolution arterhaltend sinnvoll. Wenn es hingegen um Expertise geht, darf das Erkennen der Welt – nennen wir es allgemein Wahrheitsfindung – nicht solchen unbewusst ablaufenden Mechanismen überlassen bleiben. Dies betrifft auch Entscheidungen im Bereich der Gesundheit oder des Rechts.

Mit dem Begriff der Validierung bezeichnet man die planvolle kritische Überprüfung der vorliegenden Informationen. Als Grundlage einer rationalen Entscheidung können nur Annahmen dienen, die nach den unten besprochenen Kriterien wahrscheinlich wahr sind. Validierung bedeutet vor allem, dass eine systematische Falsifizierung der zu prüfenden Annahmen oder Behauptung versucht wird. Wie dies – im Zusammenhang mit Gesundheitsstörungen – praktisch geleistet werden kann, wird hier kurz umrissen.

2. Begriffe

Der Wahrheitsbegriff

Der Wahrheitsbegriff wird hier nur in einem stark eingeschränkten Sinn verwendet. Es geht nur um diejenigen Sachverhalte, die für die Behandlung, die Begutachtung oder die richterliche Urteilsbildung relevant sind – also nicht um eine „Theorie der Welt“. Ferner gilt die Korrespondenztheorie: Wahrheit ist eine solche Beschreibung der Welt, die in möglichst exakter Korrespondenz mit der Wirklichkeit steht. Solche Wahrheiten entsprechen dem Konzept des „justified true belief“

Inhalt

1. Einleitung	1
2. Begriffe	1
3. Methoden	5
4. Richterliche Bewertung	6
5. Zusammenfassung	7

(JTB, Gettier 1963). Es handelt sich um Überzeugungen, welche durch geeignete Methoden (justified) gewonnen wurden, deren Zuverlässigkeit bekannt ist, z. B. Sensitivität und Spezifität, und die nachweislich richtig eingesetzt wurden (Dokumentation). Ferner kann die Überzeugung erst dann als (bisher) wahr gelten, wenn sie ernsthaften Versuchen der Falsifikation standgehalten hat. Dies ist ein wesentlicher Schritt, der erlaubt, wissenschaftliche Erkenntnisse von nicht-wissenschaftlichen zu trennen.

Nicht empirisch falsifizierbare Behauptungen haben beliebige Wahrheitswerte (z. B. Deutungen, psychodynamische Modelle, Narrative). Sie können schon deshalb nicht als JTB gelten – sie sind aber nicht notwendig falsch. Ferner muss die Behauptung vereinbar mit den übrigen für wahr gehaltenen Überzeugungen sein, das sog. doxastische System muss konsistent sein.

Was ist Validierung?

Validierung bedeutet, dass Annahmen bzw. Behauptungen, darin inbegriffen eigene Annahmen, einer kritischen Überprüfung unterzogen werden. Nichtwahre Behauptungen können eine ganze Reihe von Ursachen haben, diese liegen teils bei dem Probanden, teils bei dem Untersucher/Richter.

2.1 Verzerrungen auf Probandenseite

Sowohl bei der Anamnese (also den verbalen oder auf Beschwerdelisten abgegebenen Angaben) als auch bei allen Befunderhebungen, die der Mitarbeit des Probanden bedürfen, kann es zu Antwortverzerrungen kommen.

Antwortverzerrungen als nicht-authentisches Verhalten

Der heute allgemein akzeptierte Begriff für nicht-authentisches Verhalten ist „Antwortverzerrungen (AV)“ (response bias). Dies umfasst sowohl nicht-authentische Bekundungen zum Befinden als auch ein

verzerrtes Leistungsbild. Der Begriff „Antwort“ sollte nicht ausschließlich als verbale Äußerung

verstanden werden, vielmehr geht es insbesondere um Verhaltensantworten, dazu gehören auch die Lautäußerungen.

Zu den Antwortverzerrungen bei psychologischen oder körperlichen Leistungstests zählt die suboptimale Anstrengungsbereitschaft, mangelnde Testkooperation (suboptimal effort). Antwortverzerrungen können negativ sein – der Betroffene stellt sein Befinden schlechter dar, als er selbst es empfindet, bzw. sein Leistungsvermögen geringer, als es ist – oder positiv, dann stellt er sein Befinden/Leistungsvermögen deutlich besser dar (durch Vorbereitung auf Tests oder „Spickzettel“).

Motivationale Aspekte

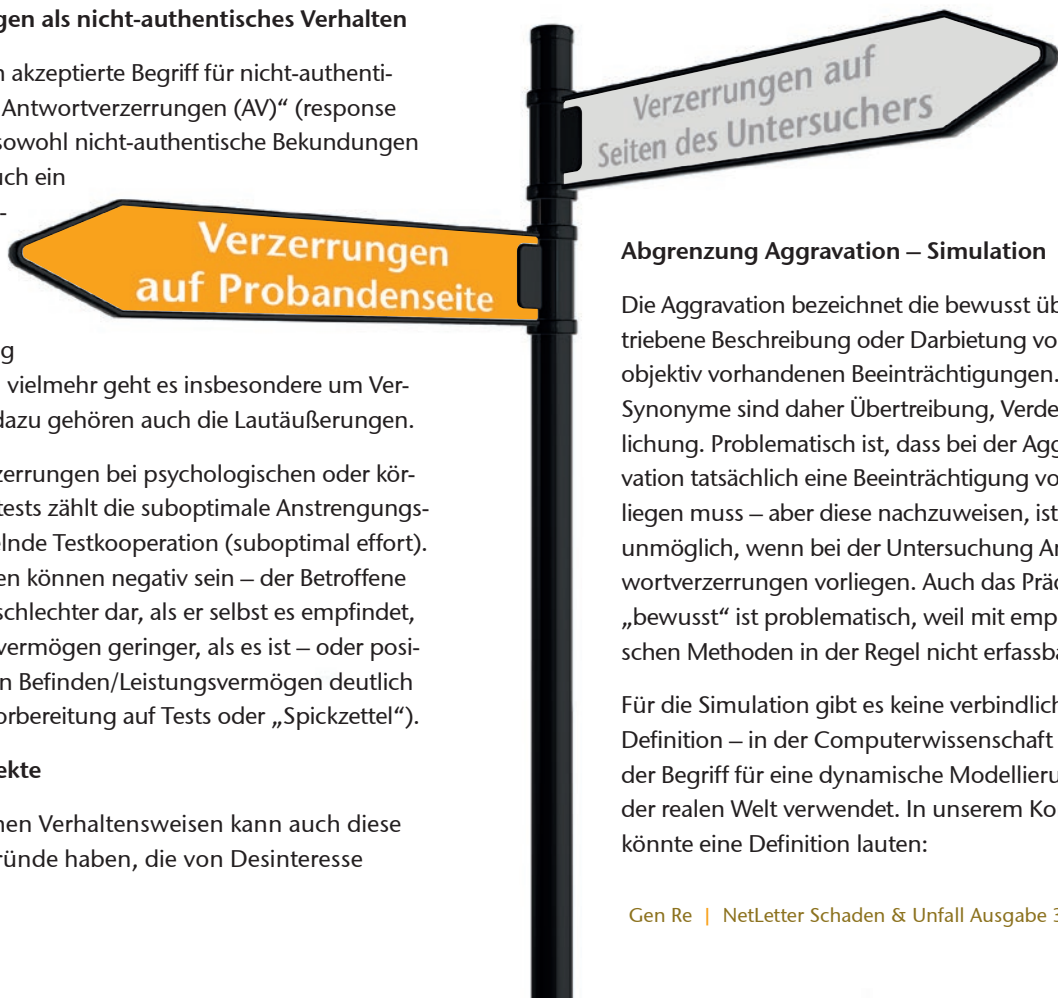
Wie alle menschlichen Verhaltensweisen kann auch diese unterschiedliche Gründe haben, die von Desinteresse

(mangelnder Motivation) an der Untersuchung bis zum hochmotivierten Täuschungsversuch reichen. Ein Schluss auf die zugrunde liegende Motivation ist mit empirischen Methoden – also wissenschaftlich – nicht möglich, da solche innerpsychischen Gegebenheiten der Untersuchung nicht zugänglich sind. Man darf zwar aus dem Kontext Mutmaßungen darüber anstellen, muss aber im Blick behalten, dass solche Mutmaßungen eben nicht in die Kategorie „JTB“ gehören, sondern in die nicht-falsifizierbaren und damit unwissenschaftlichen Hypothesen.

Bei ein und demselben Probanden können im Verlauf einer Untersuchung sowohl positive als auch negative Antwortverzerrungen vorkommen, ebenso selbstverständlich auch authentisches Leistungsverhalten, je nachdem, welches Bild der Proband vermitteln will. Es wäre daher vorschnell, aus dem Kontext der Untersuchung sogleich die Erwartung einer bestimmten Form der Antwortverzerrung zu folgern.

„Impression management“

Ein häufiges Phänomen ist das „Impression management“, d. h., der Untersuchte versucht zunächst durch positive Antwortverzerrung die Sympathie und das Vertrauen des Untersuchers oder Richters zu gewinnen. Er schildert sich als arbeitsam, sozial, um das Wohl der Familie und seiner Freunde bemüht, erscheint zur Untersuchung in Begleitung von Fürsprechern (Angehörige, sein Hausarzt), demonstriert bei der körperlichen Untersuchung Fitness und Compliance, um z. B. in der Testpsychologie Leistungen zu zeigen, die einer Demenz entsprechen.



Abgrenzung Aggravation – Simulation

Die Aggravation bezeichnet die bewusst übertriebene Beschreibung oder Darbietung von objektiv vorhandenen Beeinträchtigungen. Synonyme sind daher Übertreibung, Verdeutschung. Problematisch ist, dass bei der Aggravation tatsächlich eine Beeinträchtigung vorliegen muss – aber diese nachzuweisen, ist oft unmöglich, wenn bei der Untersuchung Antwortverzerrungen vorliegen. Auch das Prädikat „bewusst“ ist problematisch, weil mit empirischen Methoden in der Regel nicht erfassbar.

Für die Simulation gibt es keine verbindliche Definition – in der Computerwissenschaft wird der Begriff für eine dynamische Modellierung der realen Welt verwendet. In unserem Kontext könnte eine Definition lauten:

- a) Es werden unphysiologische (medizinisch nicht mögliche) oder anderweitig als vorgetäuscht erkannte Symptome dargeboten.
- b) Es ist ein externer Anreiz erkennbar, sodass die Vermutung naheliegt, dass die Symptompräsentation mit dem Anreiz in Verbindung steht.

Die Feststellung von Simulation ist noch anspruchsvoller als die der Aggravation, denn sie fordert

- a) den Nachweis der Antwortverzerrung,
- b) den Nachweis, dass die betr. Funktion gar nicht beeinträchtigt ist und
- c) die Feststellung, dass das Verhalten bewusst ist.

Die Wahrheitskonvention

Die Diskussion, inwieweit Verhaltensweisen bewusst oder unbewusst sind, ist insofern müßig, als mit empirischen Methoden keine Aussagen über die Bewusstseinsinhalte, z. B. Motivationen anderer Menschen, getroffen werden können. Die volkstümliche Definition der Lüge ist die einer bewusst falschen Behauptung, wobei der Lügner ihre Wahrheit versichert. Der implizierte Vertrauensbruch ist mit moralischer Entrüstung assoziiert. In unserer Gesellschaft gilt die Wahrheitskonvention: Eine Mitteilung wird ohne Weiteres für wahr aus der Sicht des Sprechers erachtet, es sei denn, sie erfolgt in einem besonderen Kontext, der dafür sorgt, dass sie allgemein nicht für wahr gehalten wird, z. B. im Theater.

Eine Lüge ist nicht notwendigerweise eine falsche Behauptung, sondern lediglich eine Behauptung, von der der Sprecher selbst meint, dass sie falsch

denn Menschen vergessen Äußerungen, Erlebnisse, auch frühere Behandlungen und sagen damit objektiv, aber in bester Absicht die Unwahrheit, ohne dass sie lügen: sie erliegen einem Irrtum. Das menschliche Gedächtnis ist kein Videorekorder, sondern rekonstruiert dynamisch die Vergangenheit, und die Gedächtnisinhalte verändern sich automatisch bei jedem Abruf.

Betrug ist eine vorsätzliche Handlung, deren Ziel ist, in jemandem eine falsche Überzeugung hervorzurufen oder ihn darin verharren zu lassen. Ein Betrug setzt immer eine Handlung voraus, so ist das bloße Verschweigen eines Sachverhaltes, der den anderen zu einer Korrektur seiner falschen Überzeugung veranlassen würde, noch kein Betrug. Antwortverzerrung kann auch Betrug sein, z. B. wenn ein Proband zur Untersuchung einen bereits ausgefüllten Antwortbogen für den MMPI mitbringt (positive AV) oder er eine Lähmung mimt (negative AV). Die Begriffe Lüge, Ausrede und Betrug sind Bezeichnungen, die im Gutachten nicht verwendet werden sollten.

2.2 Verzerrungen auf Seiten des Untersuchers

Zum Zusammenspiel von Befunderhebung und Diagnose

Chronologisch an erster Stelle ist die unzureichende Auswertung von Vorberichten zu nennen. Oft sieht der Behandler oder auch der Sachverständige nur auf die Diagnose. Das Problem liegt darin, dass die Diagnose keine Tatsachen mitteilt, sondern lediglich die Wertung des Verfassers jenes Berichtes: Eine Diagnose ist im Gegensatz zu einer Befundfeststellung keine Tatsache. Die Diagnose

eines Voruntersuchers sollte daher niemals in das doxastische System des Urteilenden als Tatsache aufgenommen werden.

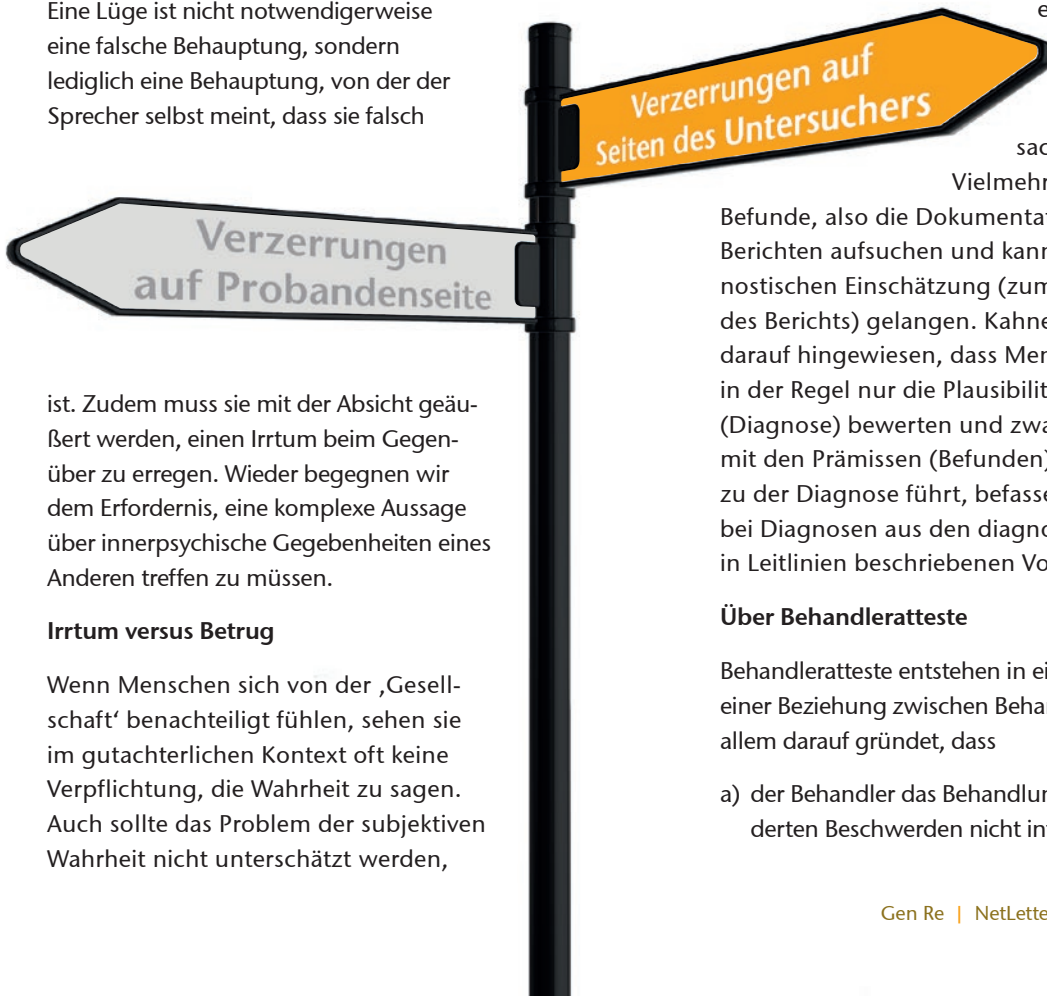
Vielmehr muss der Urteilende die

Befunde, also die Dokumentation der Tatsachen, in den Berichten aufsuchen und kann dann selbst zu einer diagnostischen Einschätzung (zum Zeitpunkt der Abfassung des Berichts) gelangen. Kahneman et al. (1974) haben darauf hingewiesen, dass Menschen aus Bequemlichkeit in der Regel nur die Plausibilität der Schlussfolgerung (Diagnose) bewerten und zwar spontan, sich aber selten mit den Prämissen (Befunden) und dem Syllogismus, der zu der Diagnose führt, befassen. Der Syllogismus besteht bei Diagnosen aus den diagnostischen Kriterien bzw. dem in Leitlinien beschriebenen Vorgehen.

Über Behandleratteste

Behandleratteste entstehen in einem besonderen Kontext, einer Beziehung zwischen Behandlern und Patienten, die vor allem darauf gründet, dass

- a) der Behandler das Behandlungersuchen und die geschilderten Beschwerden nicht infrage stellt und



ist. Zudem muss sie mit der Absicht geäußert werden, einen Irrtum beim Gegenüber zu erregen. Wieder begegnen wir dem Erfordernis, eine komplexe Aussage über innerpsychische Gegebenheiten eines Anderen treffen zu müssen.

Irrtum versus Betrug

Wenn Menschen sich von der ‚Gesellschaft‘ benachteiligt fühlen, sehen sie im gutachterlichen Kontext oft keine Verpflichtung, die Wahrheit zu sagen. Auch sollte das Problem der subjektiven Wahrheit nicht unterschätzt werden,

b) Bedürfnisse des Patienten nach Verständnis, Erklärung, Entlastung und Unterstützung auch durch Krankmeldung befriedigt werden.

Ein Beispiel für die – theorie- und faktenfreie – Befriedigung von Bedürfnissen sind behandlerseitige Erklärungen, dass der Patient sich zu sehr aufgeopfert, nur durch Leistung Anerkennung gefunden habe („Burn-out“), Nein-sagen lernen müsse und andere Weisheiten, die als „Zirkus-Barnum-Psychologie“ (Meehl 1986) bezeichnet wurden, weil sie jeder für sich gerne annimmt.

Untersuchungen zur Stichhaltigkeit von Behandlerberichten

Die Stichhaltigkeit von Behandlerberichten ist Gegenstand zahlreicher Untersuchungen gewesen mit ernüchternden Ergebnissen. Garb hat in seinem Standardwerk „Studying the Clinician“ (1998, Garb und Boyle 2015) auf typische Probleme von Behandlerberichten hingewiesen, die auch in Deutschland nachvollzogen werden konnten (Höfler und Wittchen 2000, Höfler et al. 2002, Stevens et al. 2013). Nach diesen (und anderen) empirischen Daten ist die Wahrscheinlichkeit, dass eine im Kontext der Behandlung gestellte psychiatrische Diagnose unzutreffend ist, hoch, im Bereich von $p = 0,5$. Mitchell et al. (2009, 2011) veröffentlichten die Ergebnisse einer Metaanalyse ($n = 50.371$ Patienten in 41 Studien). Die Behandler klassifizierten nur 47,3% der Fälle richtig, hier in Übereinstimmung mit einem strukturierten diagnostischen Interview.

Richter unterliegen diesbezüglich häufig einem doppelten Bias. Sie meinen, da der Behandler kein Interesse an dem Ausgang des Rechtsstreits hat, sei seine Beurteilung besonders aussagekräftig, weil neutral. Nun korreliert weder die Unbefangenheit eines Behandlers mit der Treffsicherheit seiner Diagnose, noch kann man bei einem Behandler ohne weiteres Neutralität, also Unabhängigkeit, von den Interessen des Patienten unterstellen.

Der Anamnesefehler

Nachfolgend sind Anamnesefehler zu nennen, d.h., die Anamnese wird nur insoweit erhoben, als nach Information gesucht/gefragt wird, die die Verdachtsdiagnose bestätigen und auch dies nur unvollständig (Bias der Repräsentativität). Dies ist sinnlos, da eine einzige Information, die die Diagnose widerlegt, die ganze Arbeit zunichte macht. Ähnlich ist es bei Fragebögen wenig sinnvoll, nur solche vorzulegen, die die Diagnose bestätigen könnten, etwa bei einer Depression nur Depressionsinventare.

Unzureichende Befunderhebung

Die nächste Fehlermöglichkeit ist die unzureichende Befunderhebung. Der häufigste Fehler ist, dass gar kein Befund erhoben wird. Immer wieder wird das Versäumnis dadurch kaschiert, dass als Befund, nun aber in der 3. Person Indikativ anstatt in indirekter Rede, die Beschwerden einfach wiederholt

werden. Damit reduziert man die Menge der Tatsachen, die bei der Begegnung zwischen Untersucher und Proband erhoben wurden, auf die objektive Feststellung, dass der Proband dies und jenes geäußert hat. Dass der Proband dergleichen geäußert hat, ist dann Tatsache, ob aber das, was er geäußert hat, auch zutrifft, also der Wahrheitswert seiner Angaben, bleibt unbekannt. Da Gutachten laut Becker (2008) dazu eingeholt werden, um Beweis über die Richtigkeit der klägerischen Behauptungen zu erheben und nicht etwa darüber, ob der Proband dergleichen behauptet, führen solche Gutachten nicht weiter. Sätze wie „orientierende neurologische Untersuchung unauffällig“ teilen auch keine Tatsachen mit, sondern nur, dass der Untersucher meint, er habe nichts auffälliges festgestellt. Da man aber nicht weiß, auf welche Untersuchungsschritte sich diese Wertung bezieht, kann man mit der Angabe nichts anfangen.

Die Mitarbeit des Probanden und Urteilsfehler

Zu bedenken ist ferner, dass viele Untersuchungen von der Mitarbeit abhängen, z. B. die Hörprüfung, Sehtests oder Kraftmessungen. Bei solchen Untersuchungen muss gewährleistet werden, dass hinreichende Anstrengungsbereitschaft besteht, s. unten. Bei anderen, wie der Messung des Körpergewichts, meist nicht (Ausnahme: Anorexie-Patientinnen können durch Wasseraufnahme vor dem Wiegen „tricksen“).

Der Untersucher muss ferner methodisch versiert und gerüstet sein. Er muss z. B. (als Neurologe) wissen, wie man einen Tremor oder geltend gemachten Schwindel untersucht. Ferner drohen Urteilsfehler, z. B. wenn die Krallenhand bei n. ulnarisparese mit der Schwurhand der n. medianusparese verwechselt wird oder die Innervationsgebiete von Hirnnerven nicht bekannt sind.

Formale Fehler – z. B. fehlende Begründung der Diagnose

Auch formale Fehler sind häufig, z. B. in Behandlerberichten die fehlende Begründung der Diagnose. In der Begründung sollte erörtert werden, welche Tatsachen für und welche gegen die Diagnose sprechen. Dies gewährleistet einerseits, dass der Untersucher selbst noch einmal eine Bilanz seiner Erkenntnisse zieht und ferner, dass der Leser des Berichts nachvollziehen kann (im Sinn eines Syllogismus) weshalb die Diagnose „X“ gestellt wurde.

Versäumte Validierung und ihre möglichen Folgen

Versäumte Validierung, etwa wenn der Untersucher vorschnell zu einer – falschen – Diagnose gelangt, kann üble Folgen haben. Ausgehend von nicht-validen Beschwerden oder Befunden werden Krankheiten behandelt, die gar nicht vorliegen. Kognitives Training, wenn keine echten kognitiven Störungen bestehen, mag unschädlich sein (jedoch eine Verschwendung von Ressourcen), aber diagnostische Arthroskopien, wenn keine Erkrankung vorliegt, können zu schweren Komplikationen führen. Antidepressiva können bei Nicht-Depressiven per Antriebssteigerung zu suizidalen Handlungen führen. Bei

Patienten, die über therapierefraktäre Schmerzen klagen, kommt es häufig zu einer Eskalation der Medikation mit schließlich Mehrfachverordnung von Opioiden – ohne dass die auf der Hand liegende Frage der Antwortverzerrung (malingered pain-related disability, MPRD) angegangen oder wenigstens ein Medikamentenspiegel erhoben wurde.

3. Methoden

Die Vielzahl ärztlicher Validierungstests

Es gibt viele ärztliche Validierungstests. Der Untersucher muss während seiner Befunderhebung fortlaufend, sozusagen beiläufig, auf Anzeichen der Manipulation ebenso wie unzureichender Kooperation achten. Dies setzt klinische Erfahrung und Wissen über Anatomie und Physiologie voraus. Wer die Versorgungsgebiete sensibler und die Kennmuskeln motorischer Nerven nicht kennt, wer nicht weiß, welche Schwindel- und Tremorformen es gibt und wie man sie untersucht, wird nur eine unzulängliche Untersuchung und eine unsichere Bewertung der Befunde abgeben können.

Diejenigen Verfahren, die es z. B. erlauben, nicht-authentische Hör- oder Gleichgewichtsstörungen zu erkennen, sind in den ersten Kapiteln des Standardwerkes zur HNO-ärztlichen Begutachtung von Feldmann und Brusis (2012) beschrieben, die der augenärztlichen Diagnostik in den entsprechenden Kapiteln bei Lachenmayr (2012, Kap. 3). Für das neurologische Fachgebiet wurde das erste Werk zu validierender Diagnostik 1912 von Erben verfasst, eine aktuelle Behandlung findet man bei Hirsch (2018).

Es würde den Rahmen dieses Aufsatzes sprengen, eine systematische und fachübergreifende Abhandlung der ärztlichen Validitätsdiagnostik zu liefern. Hier nur zwei Beispiele aus der Neurologie: Eine echte Ptosis (Lidheberschwäche) ändert sich nicht in Abhängigkeit von der Körperposition, auch ist die Hebung der Augenbrauen nicht beeinträchtigt. Echte Sensibilitätsstörungen sind nicht – abgesehen vom Thalamusschlaganfall – in der Mittellinie begrenzt. Der Untersucher muss die kritischen Tests

- a) kennen und
- b) einsetzen.

Die Bedeutung psychologischer Verfahren

Psychologische Verfahren können Verzerrungen der Beschwerdeschilderung erkennen. Dies ist eine wichtige Leistung. Ärzte können zwar bei der Untersuchung vorgetäuschte körperliche Symptome erkennen, aber kaum einschätzen, ob geschilderte Beschwerden (z. B. Symptome, die

außerhalb der Untersuchung bestanden haben sollen oder psychische Beschwerden) echt sind – nach empirischen Arbeiten bestenfalls mit Ratewahrscheinlichkeit. Häufig liegt ihre Erkennungsrate sogar darunter, weil die Untersucher systematischen kognitiven Verzerrungen unterliegen (Hirsch 2018, Vrij 1999, 2000, Garb 1998, Merten et al. 2019).

Wenn Beschwerden und Befundtatsachen nicht übereinstimmen, muss zuerst die Validität der Beschwerden geprüft werden, jedenfalls bevor man versucht, die Diskrepanz mit unwissenschaftlichen (nicht falsifizierbaren) Konzepten wie „Somatisierung, Hilfsuchen durch Beschwerdebetonung“ zu bereinigen. Ferner können bei der psychologischen Untersuchung Antwortverzerrungen bei der kognitiven Leistungsprüfung erkannt werden.

Sowohl die Beschwerden- als auch die Leistungsvalidierungstests funktionieren mit Präzision (ausgezeichnete Sensitivität und Spezifität [Keppler et al. 2017, Rogers und Bender 2018]), sie sind an Gutachtenpopulationen getestet und sind dem ärztlichen Urteil in aller Regel eindeutig überlegen.

Eine grundsätzliche, aber nicht allgemein beachtete Überlegung ist, dass auch etwas vorliegen muss, auf das sich die Aussage „authentisch oder nicht“ beziehen kann, bevor die Authentizität geprüft wird. Vielfach sieht man Berichte, in denen weder eine systematische Erhebung der Beschwerden noch irgendwelche Leistungsdiagnostik erfolgte, dann aber irgendein Validierungstest gegeben wurde und mitgeteilt wird „bei unauffälliger Validierung ist Aggravation ausgeschlossen“. Keinesfalls kann aus einem unauffälligen Validitätstest geschlossen werden, dass alle Aussagen bzw. Leistungen authentisch sind – und ebenso wenig umgekehrt. Es kann durchaus sein, dass die gezeigten Gedächtnisstörungen valide sind, allerdings der angegebene Sehverlust auf dem rechten Auge nicht. Es muss also eine Korrespondenz zwischen gezeigter Störung und Validierungsverfahren bestehen. Eine rückblickende Validierung, d. h., eine Prüfung, ob früher erhobene Tests (oder ärztliche Befunde) valide sind, ist meist nicht möglich. Mitunter erlauben die Befunde aber doch, Inkonsistenzen festzustellen, die Zweifel an der Validität begründen.

Validierungsverfahren grundsätzlich einsetzen

Ein Verdacht des Untersuchers ist keine notwendige Voraussetzung für den Einsatz ärztlicher oder

psychologischer Validierungsverfahren. Vielmehr sollten die Verfahren grundsätzlich eingesetzt werden, die Einschätzung des Untersuchers mag durch die Validitätsdiagnostik eine Korrektur erfahren und ihm Lernen ermöglichen. Die meisten Behandler kennen die Basisraten für Antwortverzerrungen nicht und unterschätzen deren Häufigkeit. Hinzu treten systematische Verzerrungen, wie der „self-efficiency bias“ („aufgrund meiner langjährigen Erfahrung...“ Garb 1998, Garb und Boyle 2015). Nach zahlreichen empirischen Studien korreliert die Dauer der Berufsausübung nicht mit der diagnostischen Treffsicherheit, vielmehr nehmen – ohne kritisches Feedback – die Fehler zu.

Auch die Vermutung, außerhalb von Begutachtung seien Leistungen und Beschwerden stets authentisch, weil kein Motiv für eine Antwortverzerrung bekannt ist, ist nicht haltbar. Zunächst heißt „kein Motiv bekannt“ nicht, dass auch keines vorhanden ist. Ferner können ganz unterschiedliche Gründe zu Antwortverzerrungen führen, z. B. Desinteresse an der Untersuchung. Van Egmond, Kummeling und Balkom (2005) fanden bei 42% von Patienten klare Erwartungen in Bezug auf zusätzliche positive Auswirkungen der Behandlung, die außerhalb rein therapeutischer Aspekte lagen, wobei nur 9.5% ihre Behandler darüber informierten. Mehrfachvorstellungen und stationäre Aufenthalte werden von Rentenberatern ausdrücklich empfohlen, um durch eine Fülle von wohlmeinenden Attesten dem Leistungsantrag Nachdruck zu verleihen (Fachausdruck: „stealth referral“). Merten und Giger (2018) konnten zeigen, dass 69% einer Bevölkerungsstichprobe es für völlig adäquat halten, in bestimmten Situationen eine Krankheit vorzutäuschen, wobei subjektive Gerechtigkeitsvorstellungen eine Rolle spielen. So waren 31% bereit, für eine Krankmeldung eine Krankheit vorzutäuschen, aber 59% hielten dies für angemessen, um eine Opferentschädigung zu erhalten. Erklärung: In vielen Ländern, auch in Deutschland, wird nicht das Erleiden eines Geschehens entschädigt, sondern nur eine dadurch verursachte Krankheit. Dies widerspricht dem Gerechtigkeitsempfinden.

Zur Prävalenz von Antwortverzerrungen

Zahlreiche Untersuchungen haben gezeigt, dass gerade im Kontext der Behandlung Antwortverzerrungen häufig sind, die Prävalenz liegt z. B. nach Untersuchungen an ca. 10.000 Personen in Rehakliniken der DRV zwischen 40–70% (Merten et al., 2019). Dennoch hört man im Gerichtssaal immer wieder den Refrain, es müsse eine authentische Erkrankung vorliegen, weil der Kläger doch deswegen bei mehreren Ärzten in Behandlung stand.

Behandler sollten angesichts dieser häufigen und wichtigen Differentialdiagnose viel öfter validieren, als sie es tun. Viele Behandler vermeiden Validierung, weil sie keine Vergütung dafür erhalten und weil sie den Fall nicht verlieren wollen. Natürlich ist Validierung nicht nötig, wenn ein eindeutiger, nicht falsifizierbarer Befund, wie ein radiologisch gesicherter Speichenbruch, vorliegt oder eine Lungenentzündung, auch

bei Bagatellen wie Schnupfen mag man darauf verzichten. Auf der anderen Seite sah der Verfasser grafisch manipulierte Röntgenbilder und gefälschte Medikamentenpläne. Bei anhaltenden Kopfschmerzen ohne wegweisenden Befund, bei Schwindel ohne erkennbare Ursache, bei geltend gemachten Depressionen und Posttraumatischer Belastungsstörung sollten aber grundsätzlich Validierungen erfolgen, vor allem, wenn Arbeitsunfähigkeit oder Psychotherapiebedarf attestiert werden soll. Geschieht dies nicht, ergibt sich oft nach drei bis vier Jahren ein scheinbar konsistentes Bild: Alle Behandler, die jeweils voneinander und ihren diagnostischen Einschätzungen wussten, attestierten diagnostische Gemische aus ängstlich-depressiven Syndromen, PTBS und somatoformen Schmerzen. Die Perspektive „aus der Luft“, also die oberflächliche Betrachtung, lässt alles stimmig erscheinen, und weil z. B. der Richter glaubt, bei so vielen Diagnosen, so vielen Behandlern und sogar stationären Behandlungen (bei denen der Proband vier Wochen lang „gesehen“ wurde), müsse alles seine Richtigkeit haben. Die kritische Prüfung der einzelnen Berichte („von unten“) ergibt aber, dass praktisch keine Befunde erhoben wurden, schon gar keine, die irgendeine definierte Diagnose zuließen, dass Arbeitsunfähigkeit attestiert wurde, ohne dass die Behandler wussten, was der Patient eigentlich beruflich tut, ohne dass das Leistungsvermögen geprüft wurde – und ohne Validierung. Etliche Berichte erweisen sich als „copy & paste“-Dokumente ohne individuellen Inhalt. Liest man mehrere Berichte derselben Klinik, fällt auf, dass psychische Befunde und psychodynamische Betrachtungen keineswegs individuell für den Patienten verfasst wurden, sondern Versatzstücke sind. Das Arbeiten mit Textbausteinen, die eine umfängliche Anamnese- und Befunderhebung „simulieren“, geht mitunter soweit, dass (versehentlich) in einem Befund über eine Frau neben einer sorgfältigen thermischen Untersuchung der Vestibularorgane (die nie stattgefunden hat, wie die Befragung ergab) ausführlich Befunde am männlichen Genitale inkl. Hodensackreflex beschrieben wurden. Dies in einem Gutachten, welches das Gericht für besonders ausführlich und überzeugend hielt. Bei Befragung wurde dann eingeräumt, dass man „natürlich Befunde aus dem Internet verwendet“, angeblich sei das so üblich, und hier sei versehentlich der männliche Baustein reingerutscht. (Man darf schmunzeln.) Die rückblickend geltend gemachte Berufsunfähigkeit gründet nicht selten auf ein solches System von Pseudoattesten, die einer Prüfung nicht standhalten.

Der Sachverständige X, der darauf hinweist und Behandlerdokumente kritisch auswertet, tut sich schwer. Viele Richter entscheiden nach einem Mehrheitsprinzip: „Da allein der Sachverständige X das Vorliegen einer Depression verneint, wobei der Kläger sich deswegen sogar in stationärer Behandlung befunden hat, kann das Gericht seiner Bewertung nicht folgen.“

4. Richterliche Bewertung

„Es obliegt dem Gericht, Gutachten, und zwar das selbst beauftragte wie auch vorgelegte, auf ihre Aussagekraft

und Schlüssigkeit zu prüfen“ (BGH 8.3.1955, 5 StR 49/55, BGHSt 7, 239). Dieser Anspruch wird nicht immer umgesetzt. Der Sachverständige soll Tatsachen liefern, und der Richter wendet dann das Recht an – aber zuerst muss der Richter prüfen, ob die als Tatsachen abgegebenen Behauptungen auch wahr sind. Freilich vereinfacht der Richter sich seine Aufgabe, wenn er einfach das als prozessuale Wahrheit übernimmt, was der Sachverständige schreibt oder sagt. Die Regel, dass die Überzeugung des Richters maßgeblich ist, führt häufig dazu, dass der Richter lediglich die seiner Lebenserfahrung entsprechende Plausibilität der Schlussfolgerung betrachtet, z. B. für wie wahrscheinlich er selbst es hält, dass der Kläger depressiv ist. Dabei wendet er einfache Heuristiken an, die rationale Prüfung und Validierung der einzelnen Argumente bleibt auf der Strecke.

So äußerte ein Senat des OLG Stuttgart 2018 in einem Verfahren, das Gericht benötige keine diagnostischen Kriterien, man wisse aus alltagspraktischer Erfahrung, was eine Depression sei und was nicht. Auch bei Sachverständigen kommen die oben erwähnten Fehler bei Auswertung der Akten, Anamneseerhebung, Untersuchungsmethodik und Bewertung vor. Eine personengebundene Wahrheitsfunktion gibt es nicht, d. h., die epistemische Position einer sachverständigen Aussage hängt nicht von der Person ab, die sie ausspricht, sondern von der Methode, mit der sie erreicht wurde. Also muss auch das Gerichtsgutachten sorgsam geprüft werden.

Ferner besteht das doxastische System eines Urteils nicht nur aus der Äußerung oder Sicht des Sachverständigen. Vielmehr müssen auch die von den Parteien vorgelegten Dokumente nach JTB-Standard geprüft werden. Wie gesagt genügt es nicht, festzustellen, dass acht Behandler eine Depression o. Ä. diagnostiziert und Arbeitsunfähigkeit bescheinigt haben. Vielmehr muss jedes Stück für sich validiert werden, bevor es als Argument für

irgendeine Behauptung taugt. Eine „Mehrheitsregel“ existiert in der Erkenntnistheorie nicht und der Kohärenzschluss ist aus logischen und statistischen Gründen nicht zulässig (Olsson 2005). Der Richter muss zusammen mit dem Sachverständigen das Material durcharbeiten, bis er den epistemischen Wert jedes Dokumentes erkennen kann, einschließlich der Beweiserhebung durch den Sachverständigen selbst. In der Praxis bedeutet dies, dass der Richter sich erläutern lässt, welche der Dokumente Befunde enthalten und welche Schlussfolgerungen daraus zu ziehen sind – auch der Richter muss, anstatt zu glauben, „der behandelnde Arzt wird schon alles richtig gemacht haben“, vorrangig um eine Falsifizierung bemüht sein.

5. Zusammenfassung

In der Medizin, gleich ob es um Behandlung geht oder um rechtliche Bewertung einer Gesundheitsstörung, ist das kritische Hinterfragen des Sachverhaltes wesentlich. Die eigene Bewertung gründet meist nicht auf Erkenntnissen, die ausschließlich vom Untersucher oder Urteilenden selbst erhoben wurden, sondern zudem auf der Interpretation von Feststellungen anderer. Dabei muss beachtet werden, ob es sich um übermittelte Tatsachen oder „nur“ um Bewertungen (Diagnosen, Berufsunfähigkeit) anderer handelt.

Durch diagnostische Kriterien, wissenschaftliche Leitlinien, die methodische Vorgehensweisen für einzelne Krankheiten beschreiben, und aussagekräftige Methoden steht ein Arsenal von Erkenntnis- und Prüfungsmöglichkeiten zur Verfügung, das auch genutzt werden sollte. Vor allem muss die Falsifikation der Arbeitsdiagnose unternommen werden. Validierungsverfahren, gleich ob ärztlich oder psychologisch, dienen weder der Feststellung noch dem Ausschluss einer Erkrankung. Sie erlauben lediglich, diejenigen Bestandteile des Systems von Annahmen (doxastisches System) zu überprüfen, die durch den Probanden beeinflusst werden können: Anamnese und Leistungstests. Dabei können Antwortverzerrungen natürlich auch dann auftreten, wenn „echte“ Gesundheitsstörungen vorliegen: Z. B. kann ein Proband wirklich an einem Diabetes und einer Neuropathie leiden, aber nicht-authentische Angaben zu seinem psychischen Befinden und Leistungsvermögen machen. Eine differenzierte und planvolle Vorgehensweise ist wesentlich – von der Behandlung bis zur sachverständigen und richterlichen Bewertung.

Literatur

Erben, S., 1912. *Diagnose der Simulation nervöser Symptome*. Urban & Schwarzenberg, 1912.

Feldmann, H., Brusis, T., 2012. *Das Gutachten des HNO-Arztes*. 7. Aufl., Stuttgart, New York: Thieme.



Garb, H. N. 1998. *Studying the clinician: judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Garb, H. N., Boyle, P. A., 2015. *Understanding why some clinicians use pseudoscientific methods: findings from research on clinical judgement*. In: Lilienfeld SO, Lynn SJ, Lohr JM (eds.): *Science and pseudoscience in clinical psychology*. 2nd ed. New York: The Guilford Press, 17–38.

Gettier, E. L. 1963. "Is justified true belief knowledge?" *Analysis* 23: 121–3.

Grice, P., 1981. *Presupposition and Conversational Implicature*. In: *Radical Pragmatics*, P. Cole (ed.), New York: Academic Press, 183–97.

Hirsch, A. R., 2018. *Neurological Malingering*. Taylor & Francis Inc; Abingdon.

Höfler, M., Wittchen, H.-U., 2000. *Why do primary care doctors diagnose depression when diagnostic criteria are not met?* *International Journal of Methods of Psychiatric Research*; 9, 110–120.

Jacobi, F., Höfler, M., Meister, W., Wittchen, H.-U., 2002. *Prävalenz, Erkennens- und Verschreibungsverhalten bei depressiven Symptomen. Eine bundesdeutsche Hausarztstudie*. *Nervenarzt*; 73, 651–658.

Keppler, C., Plohmann, A. M., Pflueger, M., et al. 2017. *Symptom Validation in Independent Medical Evaluations*. *Fortschritte der Neurologie-Psychiatrie*. 85 (1):17–33. DOI: 10.1055/s-0042-116681.

Lachenmayr, B., 2012. *Begutachtung in der Augenheilkunde*. 2. Aufl. Berlin, Heidelberg: Springer.

Meehl, P. E., 1986. *Diagnostic taxa as open concepts: metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision*. In: Millon T., Klerman GL (eds.): *Contemporary directions in psychopathology: Toward the DSM-IV*. New York: Guilford Press, 215–231.

Merten, T., Giger, P., 2018. *Wie häufig treten Simulation und Aggravation in der Begutachtung auf? Schätzungen von Laien*. *Forens Psychiatr Psychol Kriminol* 12, 164–171. <https://doi.org/10.1007/s11757-018-0465-4>.

Merten, T., Kaminski, A., Pfeiffer, W., 2019. *Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients*, *The Clinical Neuropsychologist*, DOI: 10.1080/13854046.2019.1694073.

Mitchell, A. J., Rao, S., Vaze, A., 2011. *International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors*. *Br J Gen Pract*, 61, 583.

Mitchell, A. J., Vaze, A., Rao, S., 2009. *Clinical diagnosis of depression in primary care: a meta-analysis*. *Lancet*, 374, 609–619.

Olsson, E. J., 2005. *Against Coherence*; Oxford University Press, Oxford, New York.

Rogers, R., Bender, S. D. (Eds.), 2018. *Clinical assessment of malingering and deception* (4th ed.). The Guilford Press.

Stevens, A., Fabra, M., Thies, E., 2013. *Self-Report vs. clinical interview for posttraumatic stress disorder in medicolegal assessment*. *German Journal of Psychiatry*; 16: 87–94.

Tversky, A., Kahneman, D., 1974. *Judgment under Uncertainty: Heuristics and Biases* Amos Tversky; Daniel Kahneman Science, New Series, Vol. 185, No. 4157, 1124–1131.

van Egmond, J., Kummeling, I., Balkom, T., 2005. *Secondary gain as hidden motive for getting psychiatric treatment*. *European Psychiatry*, 20 (5–6), 416–421.

Vrij, A., 2000. *Detecting Lies and Deceit*. Wiley Chichester UK, NY, NY.

Vrij, A., Baxter, M., 1999. *Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences*. *Expert Evidence* 7, 25–36. <https://doi.org/10.1023/A:1008932402565>.

Über den Autor

Prof. Dr. med Andreas Stevens ist Facharzt für Neurologie, Psychiatrie und Psychotherapie. Seit 1996 ist er als Gutachter tätig in Tübingen und Berlin. Er ist Autor von über 100 Veröffentlichungen in internationalen Zeitschriften. Er unterrichtet an der Universität Tübingen und bei Ärztekammern (Curriculum Ärztliche Begutachtung); E-Mail: stevens@med-begutachtung.de.



Herausgeber

General Reinsurance AG
Theodor-Heuss-Ring 11
50668 Köln
Tel. +49 221 9738 0
Fax +49 221 9738 494

Bildnachweis

Foto Titel (li): © smolaw11 - gettyimages.com
Foto Titel (re): © marchmeena29 - gettyimages.com

Redaktion

Mirko von Haxthausen (verantwortlich),
Markus Burbach, Gerhard Riedel
Tel. +49 221 9738 156
Fax +49 221 9738 824
mirko.vonhaxthausen@genre.com
www.genre.com/business-school

Layout

gläser projekte GmbH, Köln

The Cologne Re.
Oldest Reinsurance Company in the World – Established 1846

Seite 5: © voyata - gettyimages.com