

Wie belastbar ist Deine Diagnose?

Wahrscheinlichkeitstheoretische Betrachtungen unsicherer Untersuchungsergebnisse und gutachterlicher Äußerungen

Achim D. Gruber, Axel Wehrend, Thomas Steidl¹, Katharina Kramer², Sabine Merbach², Wolfgang Baumgärtner, Thomas Buyle, Niclas-Frederic Weisser, Andreas Beineke³

¹ Vorsitzender des Arbeitskreises Forensische Veterinärmedizin in der Deutschen Veterinärmedizinischen Gesellschaft (DVG)

² Vorsitzende des Arbeitskreises Diagnostische Veterinärpathologie

³ Vorsitzender der Fachgruppe Pathologie in der DVG

Manchmal geht es um folgenreiche Entscheidungen: Ist die kostenintensive Therapie gerechtfertigt? Soll das Bein amputiert werden? Kann nur noch die Euthanasie das Leiden beenden? Und vor Gericht geht es um das Finden eines angemessenen Urteils. Doch Untersuchungsergebnisse und forensisch-gutachterliche Äußerungen können infolge ungenügender Datenlage einen gewissen Grad an Unsicherheit enthalten und zu Fehleinschätzungen führen. So entsteht regelmäßig ein Spannungsfeld zwischen den Erwartungen der Auftraggebenden und der tatsächlichen Leistbarkeit einer Wahrscheinlichkeitsaussage. Die Kenntnisse über verschiedene Arten und Ursachen von Unsicherheiten sowie ihre angemessene Kommunikation im Einzelfall können dazu beitragen, Fehlentscheidungen zu vermeiden.

Unsicherheiten und das Risiko, bei unvollständiger Datenbasis einer Fehleinschätzung zu unterliegen, sind aus vielen tierärztlichen Tätigkeiten nicht wegzudenken. Sie sind jedoch unangenehm und werden daher gern verdrängt, übersehen oder nicht mitgeteilt, was für Mensch und Tier zu einem weiten Spektrum an nachteiligen Konsequenzen führen kann. Bestimmte diagnostische Disziplinen sind dafür anfälliger als andere. Je komplexer die Grundlagen für eine Einschätzung sind, desto stärker muss ihre Fehleranfälligkeit angenommen werden. In der Radiologie etwa gilt die Uneinheitlichkeit zwischen den Diagnosestellenden als ihre wichtigste Achillesferse [1]. Besonders in der pathologischen Biopsie-, Zytologie- und Obduktionsdiagnostik können viele Einflussfaktoren die Urteilsunsicherheit bei uneindeutigen Befunden einschränken. In Anbetracht des zunehmenden Wunsches der Klientel nach maximaler diagnostischer und therapeutischer Betreuung ihrer Tiere bei gleichzeitig verstärkter Neigung zu Rechtsstreitigkeiten bei Therapieversagen kommt der Urteilssicherheit wahrscheinlich eine steigende Bedeutung zu. Eine noch höhere Komplexität und damit Irrtumswahrscheinlichkeiten können für manche forensische Begutachtungsverfahren angenommen werden. Doch kennen wir sie und gehen wir damit immer angemessen um?

Quellen für Unsicherheiten

Die Belastbarkeit einer Einschätzung zu vagen oder komplexen Zusammenhängen basiert auf zahlreichen Faktoren, darunter die persönliche Fachkompetenz (Aus- und Weiterbildung, Erfahrung u. a.), aktiv selbst hinterfragende, kognitive Fähigkeiten sowie die Qualität des strukturierten Vorgehens bei der Analyse [1]. Gleichzeitig wirken oft zahlreiche verfahrens- sowie fallspezifische Störquellen auf das Ergebnis. Diese sollen hier zur systematischen Betrachtung ihrer Ursachen und mit Blick auf ihre Kommunikationswürdigkeit in drei Gruppen eingeteilt werden: 1.) vorhersehbare, vorgangstypische Unsicherheiten, 2.) fallspezifische, erkennbare Unsicherheiten und 3.) nicht zu erwartende oder nicht erkennbare Unsicherheiten.

Eine solche Differenzierung ist für den professionellen Umgang mit Fehleranfälligen im tierärztlichen und forensischen Alltag entscheidend, u. a. indem sie den Blick für die Zuständigkeiten, Verantwortungen und nicht zuletzt Haftbarkeiten aller Beteiligten schärft. In jedem Einzelfall sind alle drei Gruppen in mehr oder weniger großem Ausmaß als Fehlerquellen mit addierender Wirkung auf die Gesamtunsicherheit zu erwarten.

Im Folgenden werden diese Zusammenhänge am Beispiel mikroskopischer Biopsie- und Zytologieuntersuchungen veranschaulicht.

Gruppe 1: Vorhersehbare, vorgangstypische Unsicherheiten

Die histologische Untersuchung gilt gemeinhin als der Goldstandard in der Diagnostik der meisten Tumoren, Entzündungsformen und anderen geweblichen Veränderungen. Die eindeutige Identifikation spezifischer mikroskopischer Muster (engl. *patterns*) erlaubt der Pathologin¹ in den meisten Fällen eine eindeutige Zuordnung zu einer spezifischen Entität, ihrer Ursache, bei Tumoren die Beurteilung einer vollständigen Exzision, dem zu erwartenden Verlauf (Prognose) und vielen anderen, entscheidungsrelevanten Aspekten. Worin liegt hier die Fehleranfälligkeit, eine gute Fachkompetenz vorausgesetzt und jegliche Probenverwechslung ausgeschlossen? Maßgeblich ist die Repräsentativität der Stichprobe für die Gesamtheit der Veränderung. Bereits im tiermedizinischen Grundstudium wird vermittelt, dass ein histologisches Präparat nur wenige µm dick ist. Wenn man sich einen typischen Hauttumor eines Hundes als die Größe einer Turnhalle vorstellt, würde die Pathologin etwa eine daumendicke, etwa zentral gelegene Scheibe daraus unter dem Mikroskop betrachten. Selbst wenn zur Beurteilung der Resektionsränder eine zweite, rechtwinklig dazu gelegte Ebene beurteilt wird, bleiben 99,9 Prozent des Tumors unbesehen. Die Konsequenz für die Belastbarkeit der Einschätzung zum Resektionsrand liegt auf der Hand. Dennoch, für die meisten histologischen Untersuchungen reicht dieser Grad an Repräsentativität völlig aus.

Anders ist die Situation bei inhomogenen oder komplexer strukturierten Läsionen. Zu den häufigen Untersuchungen zählen auch entfernte Hundemilzen mit tumorverdächtigen Umfangsvermehrungen. Oft stellen Hämatome den Löwenanteil des Volumens der Raumforderung dar, wogegen die eigentliche Ursache, etwa ein lebensbedrohliches Hämangiosarkom oder nur eine alterstypisch normale, noduläre Hyperplasie der weißen Pulpa, darin sehr klein sein

¹ Zur einfachen Lesbarkeit wird für tiermedizinische Berufe und Spezialbezeichnungen aufgrund der Überzahl des weiblichen Geschlechts lediglich die weibliche Form genannt, die damit stellvertretend für alle Geschlechter steht.

kann [2]. Folglich besteht ein erhöhtes Risiko einer falsch-negativen Diagnose, wenn die sehr viel kleinere Veränderung, die die Blutung auslöste, nicht zufällig unter das Mikroskop gelangte. Im Bild der Turnhalle könnte die oder der Betrachtende weniger, hoffentlich repräsentativer Anschnitte leicht einen Elefanten übersehen, der in einer Hallenecke steht. In der Routine-Untersuchungspraxis kann dieses Risiko vermindert werden durch eine gründliche Auflamellierung und Probenauswahl sowie eine Erhöhung der Zahl der mikroskopischen Präparate. Hier sind jedoch praktische Grenzen gesetzt, die einen Anspruch auf eine sichere, vollständige Untersuchung quasi ausschließen. Der Anteil an unterdiagnostizierten Milztumoren wird wahrscheinlich unterschätzt [2].

Aus den Beispielen wird klar, dass sowohl die Entnahme der diagnostischen Probe durch die Tierärztin als auch der Zuschnitt in der Pathologie wesentlichen Einfluss nehmen können auf den Grad der Unsicherheit des Ergebnisses. Gleiches ist vielfach dokumentiert für falsch-negative Diagnosen bei Prostatakarzinomen [3] oder anderen, kleinsten oder schwer zugänglichen Veränderungen. Diese Art der Fehleranfälligkeit ist hier verfahrens- und entitätsinhärent, bekannt und damit vorhersehbar. Da all dies auch Gegenstand des tierärztlichen Studiums ist, können Unsicherheiten aus dieser Kategorie bei Tierärztinnen als grundsätzlich bekannt vorausgesetzt werden. Sie müssen in der Regel nicht mehr mitgeteilt werden. Begründete Ausnahmen können jedoch vorliegen, etwa bei bekannten Limitierungen eines neuen Testverfahrens.

Ähnliches gilt in Bezug auf die Repräsentativität der Probe auch für zytologische Untersuchungen von Flüssigkeiten oder geweblichen Feinnadelaspiraten (Abb. 1). Da hier das Verhältnis zwischen Probengröße und tatsächlichem Umfang der Veränderung oft noch

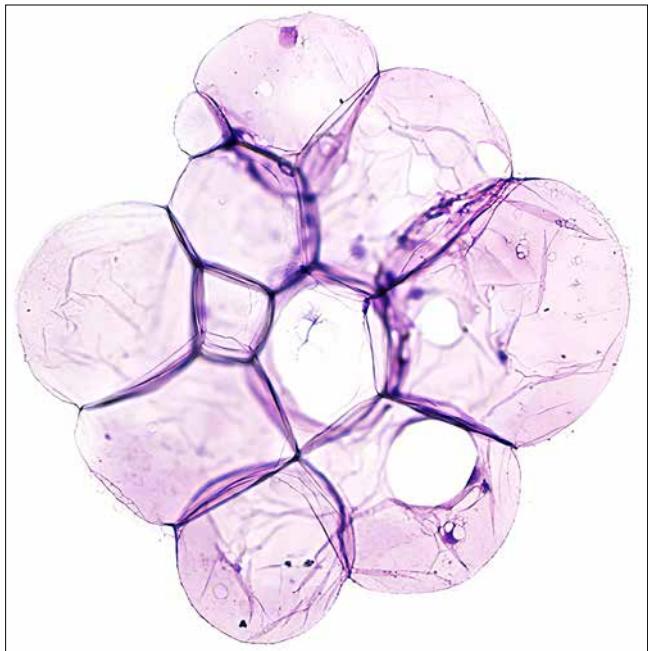


Abb. 1: Die Belastbarkeit einer zytologischen Diagnose hängt auch von der Repräsentativität der Probe und den Begleitinformationen im Untersuchungsauftrag ab. Das hier abgebildete Muster (600-fache Vergrößerung) wird bei einem kurzen Vorbericht wie „Hund, Umfangsvermehrung Haut“ üblicherweise zur Diagnose „Lipom“ führen. Je nach genauer Entnahmelokalisation und klinischem Kontext könnte es sich jedoch auch um unverändertes Unterhautfettgewebe, ein gut differenziertes, aber infiltratives Lipom, Omentumanteile in einer Nabelhernie oder Beckenhöhlenfettgewebe in einer Perinealhernie handeln. Die finale Zuordnung erfordert eine Betrachtung im klinischen Gesamtkontext.

ungünstiger ist und in zytologischen Ausstrichen wichtige Informationen zumeist fehlen, etwa zu strukturellen Gewebezusammenhängen, gelten diese grundsätzlich als anfälliger für diverse Unsicherheiten als histologische Untersuchungen. Auch diese Unsicherheiten sind prinzipiell methodeninhärent, bei Fachpersonal als grundsätzlich bekannt vorauszusetzen und werden daher in der Regel nicht mitgeteilt, sofern keine konkrete Diskrepanz zu vorberichtlichen Angaben auffällt. Anders ist es bei zusätzlichen, jedoch fallspezifischen und erkennbaren Ungewissheiten, die der Risikogruppe Nr. 2 zuzuordnen sind.

Gruppe 2: Fallspezifische, erkennbare Unsicherheiten

Erneut sei am Beispiel der Pathologie erörtert, was auch für viele andere Analyseverfahren zutreffen kann. Selbst bei einer hohen Zahl und starker Repräsentativität der mikroskopierten Anschnitte eines Tumors können diagnostisch relevante Muster fehlen oder vage bleiben. So können bei fortgeschrittenen malignen, entdifferenzierten Tumoren histogenetisch entscheidende Differenzierungskriterien fehlen. In anderen Fällen kann die Unterscheidung zwischen reaktiver Hyperplasie und Neoplasie aufgrund fehlender Entscheidungskriterien schwerfallen (Abb. 2). Dann muss das Urteil zwangsläufig einen gewissen Grad an Unsicherheit aufweisen, die der untersuchenden Person bewusst ist.

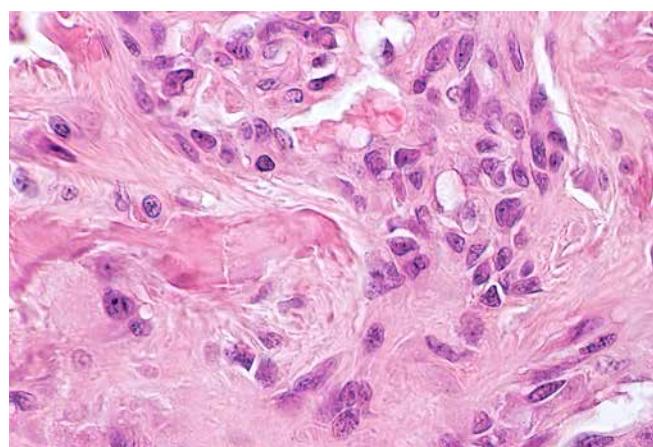


Abb. 2: Beispiel eines unsicheren mikroskopischen Befunds aus einem Osteosarkom eines Hundes (HE-Färbung, 400-fach). Anhand eines solchen Präparates soll oft aus wenigen, winzigen Biopsien von einer tumorverdächtigen Knochenveränderung entschieden werden, ob die Gliedmaße zu amputieren ist. Wie hier dargestellt können manche Osteosarkome in bestimmten Regionen quasi identische Muster aufweisen wie eine aktive, reparative und regenerative Kallusbildung nach einer Fraktur. Solange klare Leitlinien für das Vorgehen fehlen, muss die Pathologin eigenverantwortlich über den Sicherheitsgrad ihrer Diagnose entscheiden. Zweit- und Drittmeinungen sowie zusätzliche Biopsieuntersuchungen können diesen erhöhen, jedoch erfordert die Entscheidung über eine Amputation immer ein komplexes Abwägen im klinischen Gesamtkontext.

Neben der individuellen Fachkompetenz und diversen, als situatives Rauschen (engl. *occision noise*, s. u.) bezeichneten Störfaktoren, etwa der momentanen Konzentrationsfähigkeit, spielen am Ende die Erfahrung und Intuition eine entscheidende Rolle bei der Einschätzung des Grades der Unsicherheit [1]. Da diese Art der Unsicherheit höchst fallspezifisch, der Pathologin bewusst und entscheidend für weitere klinische Weichenstellungen ist, muss sie im Bericht mitgeteilt werden. Im einfachsten Fall erfolgt dies mit dem

Zusatz „Verdacht auf“ vor dem diagnostischen Urteil, womöglich mit Nennung von Differenzialdiagnosen. Wertvoller kann die Angabe eines geschätzten Wahrscheinlichkeitsgrades sein.

Für standardisierte Formulierungen von Unsicherheitsgraden liegen in der Literatur diverse, vielfach uneinheitliche und sich teils widersprechende Vorschläge vor [4–10]. Einige davon setzen spezifische Begriffe mit prozentualen Wahrscheinlichkeiten gleich, etwa eine 95-prozentige Sicherheit bei der Einschätzung „*mit an Sicherheit grenzender Wahrscheinlichkeit*“ [4, 8]. Andere ordnen demselben Prozentsatz den Begriff „*mit hoher Wahrscheinlichkeit*“ zu und geben für „*mit an Sicherheit grenzender Wahrscheinlichkeit*“ einen Sicherheitsgrad von 99,99 Prozent an [7]. Wieder andere Autoren lehnen einen Bezug zu einem numerischen Prozentsatz der Wahrscheinlichkeit als Pseudogenauigkeit grundsätzlich ab und stützen sich allein auf die übliche Bedeutung der Begriffe aus dem deutschen Sprachgebrauch [9]. Dabei ist der Wunsch nach einer maßähnlichen Beurteilung durch das damit vermittelte pseudo-objektive Gefühl zwar verständlich, jedoch ohne belastbare Datenbasis eher von zweifelhafter Natur.

Die Fachgruppe Pathologie in der DVG sah hier Entscheidungsbedarf für eine Vereinheitlichung der benutzten Begriffe. In ihrer Mitgliederversammlung am 10.03.2024 wurde nach Diskussion und Abwägung der Möglichkeiten der in **Tabelle 1** gelistete Standard für die Formulierung von Unsicherheitsgraden vorgeschlagen. Dieser Konsens unterscheidet sich kaum von ähnlichen Vorschlägen früherer tiermedizinischer Standardwerke [4, 5, 8], verzichtet jedoch bewusst auf einen Bezug zu Prozentsätzen.

Für den praktischen Einsatz der einzelnen Begriffe können keine Regeln oder Kriterien formuliert werden. Hierfür entscheidet vielmehr die urteilende Person rein subjektiv und intuitiv anhand aller verfügbaren Informationen auf Basis der eigenen Kompetenz, Erfahrung, Literatur usw. Es handelt sich quasi um eine professionelle Eigenleistung, die prinzipiell anzweifelbar sein kann.

In der Regel sind nur die in Gruppe 2 genannten Unsicherheiten in Untersuchungsberichten bzw. Gutachten zu erwähnen. Im Einzelfall können auch Gründe vorliegen, andere offensichtliche Unsicherheiten mitzuteilen.

1.	sicher (Diagnosestellung ohne weitere Ergänzungen)
2.	mit an Sicherheit grenzender Wahrscheinlichkeit
3.	mit hoher Wahrscheinlichkeit (= hochwahrscheinlich)
4.	wahrscheinlich
5.	ausgeglichen wahrscheinlich (zwei oder mehr Möglichkeiten erscheinen ähnlich wahrscheinlich)
6.	weniger wahrscheinlich
7.	unwahrscheinlich (= kaum wahrscheinlich)
8.	auszuschließen (= unmöglich, undenkbar)

Tab. 1: Empfohlene Abstufung von Wahrscheinlichkeitsgraden bei der Formulierung von Diagnosen und gutachterlichen Äußerungen. Die Bedeutung der Worte entspricht dem allgemeinen Sprachverständnis. Auf äquivalente Prozentzahlen wird aufgrund fehlender Basis verzichtet.

Für die Einschätzung „*sicher*“ stellt sich die grundsätzliche Frage, wie sicher eine Diagnose überhaupt sein kann. Zweifelsohne unterliegt auch jede als „*sicher*“ formulierte Diagnose – als Ausdruck fehlender, erkennbarer Differenzialdiagnosen aus der aktuellen Perspektive der urteilenden Person – allen anderen Unsicherheiten der hier genannten Gruppen 1 und 3. Obwohl diese als verfahrensähnert und generell bekannt (Gruppe 1) oder nicht zu erkennen (Gruppe 3) nie auszuschließen sind, bedarf es nicht ihrer expliziten Erwähnung. Darin kommen auch die Individualität und prinzipielle situative Abhängigkeit einer jeden Diagnose bzw. gutachterlichen Einschätzung zum Ausdruck [1].

Formulierung von Wahrscheinlichkeiten mit prozentualen Angaben
Die prozentuale Angabe einer Wahrscheinlichkeit kann durchaus gerechtfertigt sein, sofern sie auf einer wissenschaftlichen Basis beruht. So haben etwa zahlreiche Forschungen die Überlebenswahrscheinlichkeiten für bestimmte Zeitintervalle nach einer spezifischen Tumordiagnose ermittelt. Mit Referenz zu solchen Literaturangaben kann es daher zielführend und legitim sein, von mittleren Überlebenswahrscheinlichkeiten von XY Prozent nach 3, 6 oder 9 Monaten nach Diagnosestellung zu sprechen. Die Nennung prozentualer Wahrschein-

lichkeiten kann in ähnlicher Weise für viele andere, wissenschaftlich untermauerte, medizinische Zusammenhänge berechtigt sein.

Formulierung von Unsicherheiten ohne graduelle Abstufung: *Non liquet*-Aussagen

In manchen Fällen ist eine Unsicherheit eher grundlegender Natur ohne einen erkennbaren Bezug zu einem Wahrscheinlichkeitsgrad. Dafür finden als *non liquet* (lat. für: „es ist nicht klar“)-Formulierungen bezeichnete Begriffe Anwendung. Sie sollen zum Ausdruck bringen, dass ein Sachverhalt anhand der vorliegenden Daten nicht einzuschätzen oder eine gutachterliche Frage nicht zu beantworten ist. Mit anderen Worten: Es gibt alternative Erklärungsmöglichkeiten und es müssen weitere Informationen eingeholt werden, zumeist durch ergänzende Untersuchungen, um einen Zusammenhang sicher oder zumindest sicherer beurteilen zu können. Zu dieser Gruppe gehören Begriffe wie „möglich“, „nicht auszuschließen“, „kommt in Betracht“ und „kein Hinweis auf“. Auch in diese Kategorie fallen die Einschätzungen „Bild wie bei“ oder „vereinbar mit“, die den Lesenden zwar eine höhere Wahrscheinlichkeit suggerieren könnten, jedoch prinzipiell als *non liquet*-Aussagen gelten.

Konsequenzen in Form klinischer oder richterlicher Entscheidungen

Alle bislang genannten Unsicherheiten sind für die klinische Entscheidung (etwa Amputation, aufwändige Therapie oder Euthanasie) in voller Komplexität des klinischen Kontextes zu berücksichtigen, zumeist auch unter Einbeziehung der Klientel. Die Plausibilität eines jeden Puzzleteils und die Stimmigkeit des Gesamtbildes sind durch die verantwortliche Klinikerin in jedem Fall kritisch zu hinterfragen, bevor ein womöglich folgenschwerer Entschluss getroffen wird. Im Zweifelsfall kann oft auch ein Dialog mit der Pathologin bei der Entscheidungsfindung weiterhelfen.

Ähnliches gilt für das richterliche Urteil, wobei Richterinnen und Richter sowohl im Strafverfahren (§ 261 Strafprozeßordnung – StPO) als auch im Zivilverfahren (§ 286 Zivilprozeßordnung – ZPO) grundsätzlich frei in ihrer fallspezifischen Beweiswürdigung sind. Allerdings muss das Gericht regelmäßig auf Sachverständige zurückgreifen, um den entscheidenden Richterinnen und Richtern Sachkunde auf einem bestimmten Wissensgebiet, etwa der Pathologie oder Forensik, zu vermitteln, die bei ihnen nicht oder nicht vollständig vorliegt [11].

Fallspezifische, erkennbare Unsicherheiten sollten in der Epikrise eines Untersuchungsberichts bzw. Gutachtens erörtert werden, wobei der auftraggebenden Person bzw. dem Gericht auch Lösungsvorschläge zur Erhöhung der Sicherheit unterbreitet werden können.

In allen genannten Bereichen lautet deshalb die zentrale Frage: Sind die vorliegenden Befunde bzw. Beweise für eine klinische Entscheidung bzw. das Urteil ausreichend? Falls daran Zweifel bestehen, müssen vor einer möglicherweise folgenschweren Fehlentscheidung – sofern möglich – weitere Untersuchungen durchgeführt bzw. Beweise eingeholt werden, um den Sicherheitsgrad auf das erforderliche Maß zu erhöhen.

Aber wie viel Evidenz ist genug? In der Klinik handelt es sich meist um eine komplexe, subjektive, teils intuitive Entscheidung, die jedoch stets die Haftbarkeit der entscheidenden Person miteinschließt. Daraüber sollte man sich immer bewusst sein! Wurde eine Fehlentscheidung getroffen, die justiziable Konsequenzen nach sich zog, ist die Entscheidung anhand aller dokumentierten Evidenzen im Nachhinein zu rechtfertigen, um eine Haftung möglichst abzuwenden. Dafür ausschlaggebend ist die Nachvollziehbarkeit der Entscheidungsgrundlagen sowie der Rückschlüsse, die aus diesen Grundlagen gezogen wurden. Von zentraler Bedeutung in den Konsequenzen einer Fehl-

entscheidung kann auch die Unterscheidung zwischen Dienst- und Werkvertrag sein, die stark vom Einzelfall abhängt. Der Vertragstyp ändert jedoch nichts an den Anforderungen an den Evidenzstatus.

Bei gutachterlichen Bewertungen hat das Gericht anhand des Gutachtens im Wege der freien Beweiswürdigung zu einer eigenständigen Überzeugungsbildung zu gelangen [12]. Bereits der Terminus „Überzeugung“ bringt zum Ausdruck, dass vor Gericht nicht Erkenntnisse gewonnen werden können, die mit einem naturwissenschaftlichen oder gar mathematischen Beweis vergleichbar wären. Mithin kann das Gericht unter bestimmten Voraussetzungen auch von der sachverständigen Einschätzung abweichen, sofern in seinen Augen dafür hinreichende Gründe vorliegen. Aufgrund der mangelnden eigenen Sachkunde benötigt das Gericht im zu erstattenden Gutachten regelmäßig eine möglichst eindeutige Aussage auf alle gestellten Fragen. Aus diesem Grund wird sowohl in zivil- als auch in strafrechtlichen Verfahren zumeist eine „sichere“ oder eine „mit an Sicherheit grenzender Wahrscheinlichkeit“ getroffene Aussage als ausreichend angesehen (vgl. Tab. 1).

Die Sicherheit, mit der richterliche Fragen beantwortet werden können, hängt in hohem Maß von der konkreten Formulierung der Fragestellung ab. Insofern kann es hilfreich sein, die Fragen des Gerichts mit der zuständigen Richterin oder dem Richter zu erörtern und ggf. gemeinsam umzuformulieren. So können optimal formulierte Fragen die richterliche Überzeugungsfindung erleichtern, was stark vom Sachverstand der begutachtenden Person abhängen kann. Hier kann und soll die oder der Sachverständige beratend tätig sein.

Im Spannungsfeld zwischen Erwartungen und Sicherheit

Mehrere Umstände könnten Pathologinnen dazu veranlassen, sich in ihrem Urteil häufiger festzulegen, als es gerechtfertigt wäre. Auftraggebende einer Untersuchung erwarten vom Ergebnis schließlich eine möglichst eindeutige Festlegung, die eine klare klinische Entscheidung rechtfertigt. Jede dabei formulierte Unsicherheit führt entweder zu einem erhöhten Risiko einer klinischen Fehlentscheidung oder weiterem diagnostischen Aufwand, und beides ist unerwünscht. Stellt die untersuchende Person wiederholt vage Diagnosen, wird womöglich ihre Kompetenz angezweifelt. Auf der anderen Seite kann ein großes Interesse der untersuchenden Person bestehen, sich gegen spätere Haftungsansprüche abzusichern, was am besten durch verbale Einschränkungen der Diagnosesicherheit gelingt. Je nach Mentalität und Erfahrung kennen wir daher auch Kolleginnen und Kollegen, die sich entweder auffällig oft mit „Verdacht auf...“-Diagnosen diesbezüglich absichern oder sich zu häufig festlegen und damit die Folgen eigener Fehldiagnosen in Kauf nehmen müssen.

Ähnlichen Spannungen sehen sich forensische Gutachtende ausgesetzt, wenn das Gericht von ihnen erwartet, sich in ihren Schlussfolgerungen mit „sicher“ oder „mit an Sicherheit grenzender Wahrscheinlichkeit“ möglichst festzulegen (vgl. Tab. 1). Jedoch führen die hohe Komplexität vieler Verfahren mit zahlreichen unbekannten Einflussfaktoren bei seriösen und kritischen Gutachtenden nicht selten zu stärkeren Einschränkungen ihrer Bewertungen, als es in den Augen des Gerichts für einen klaren Urteilsspruch wünschenswert wäre. Die Last der Entscheidung liegt am Ende immer bei der gutachtenden Person, ob sie bzw. er pragmatisch zu einem schnellen Verfahren beiträgt und dabei womöglich das Risiko einer Fehlentscheidung eingeht, oder ob dem Prinzip „in dubio pro reo“ folgend – und nicht zuletzt zur eigenen Absicherung gegenüber späteren Haftungsansprüchen – dem begründeten Zweifel ein angemessener Raum gewährt wird. Sowohl im klinischen als auch im forensischen Spannungsfeld können Fachkompetenz, Erfahrung, strukturierte Verfahrenshygiene und konsequente Sachlichkeit helfen, folgenschwere Fehler für alle Beteiligten zu vermeiden [1].

Gruppe 3: Nicht erwartete und nicht erkennbare Unsicherheiten

Für die Abweichungen bei der Einschätzung von vagen, also mit gewissen Freiheitsgraden versehenen Einschätzungen, sind weitere Ursachen bekannt. Diese wurden bereits systematisch für medizinische und andere Lebensbereiche untersucht und können in zwei Untergruppen eingeteilt werden: Verzerrungen und Rauschen [1]. In der Tiermedizin sind den Autoren dazu keine systematischen Untersuchungen bekannt, jedoch muss im Analogieschluss zu anderen Tätigkeitsfeldern mit vergleichbaren Entscheidungsprozessen und ähnlichen Zusammenhängen gerechnet werden.

Verzerrungen (engl. *bias*) führen zu einer meist gerichteten Abweichung durch einen systematischen Fehler, vergleichbar mit einer Falschjustierung eines Messgeräts. So können Verzerrungen in verschiedenen tiermedizinischen Fachdisziplinen, etwa der Pathologie und Forensik, u. a. durch mangelnde Fachkenntnis entstehen, etwa wenn der untersuchenden Person bestimmte Muster bei der Einschätzung bösartiger Tumoren unbekannt sind. Dadurch können maligne Neoplasien regelmäßig unterdiagnostiziert werden. Andererseits können persönliche, prägende Erlebnisse womöglich dazu führen, dass bestimmte Entitäten überdiagnostiziert werden, etwa wenn eine frühere Fehldiagnose belastend in Erinnerung blieb.

Rauschen (engl. *noise*) bezeichnet zufällige, ungerichtete und unerwünschte Abweichungen vom Idealurteil infolge von zumeist nicht angemessen wahrgenommenen, allgemeinen Lebenseinflüssen. Stress, Übermüdung, temporär herabgesetzte Konzentrationsfähigkeit durch Unwohlsein oder privaten Kummer sind bekannte Ursachen, die auch in der Tiermedizin eine nicht unerhebliche Rolle spielen dürften. Dabei werden abhängig von Ursache und Auswirkung verschiedene Arten von noise unterschieden [1], auf die hier nicht näher eingegangen werden soll.

Bias und *noise* wirken grundsätzlich unabhängig voneinander und additiv [1]. Ihre Folgen sind in zahlreichen Feldern der Medizin gut untersucht, etwa in Bezug auf Intra-Observer-Variabilität und Inter-

Gruppe 1	Unsicherheiten, die in der Methode begründet und bei Fachpersonal allgemein bekannt sind. Sie sind prinzipiell voraussehbar und sollten daher bei der Auswahl der eingesetzten Methode berücksichtigt werden. Sowohl bei der Probennahme als auch bei der Durchführung des Untersuchungsverfahrens können solche Unsicherheiten begrenzt beeinflusst werden.
Gruppe 2	Falltypische oder fallspezifische Unsicherheiten, die aus ungenügenden, diagnostisch relevanten Informationen aus der Probe bzw. aus den dem Gutachten zugrunde liegenden Informationen resultieren. Die untersuchende Person erkennt diese Unsicherheit und formuliert ein vages Urteil. Im Untersuchungsbericht werden <i>non liquet</i> -Aussagen eingesetzt oder eine intuitive, subjektive und grobe Einschätzung einer Wahrscheinlichkeit anhand der in Tabelle 1 genannten Begriffe. Darin kommen die Kompetenz und die Einzigartigkeit der untersuchenden Person zum Ausdruck [1]. Die Konsequenz solcher Unsicherheiten kann in der Empfehlung für weitergehende Untersuchungen bzw. Informationsbeschaffungen bei Gerichtsverfahren liegen.
Gruppe 3	Unsicherheiten durch zumeist weder der auftraggebenden noch der untersuchenden Person offensichtlichen Verzerrungen (engl. <i>bias</i>) und Rauschen (engl. <i>noise</i>). Diese können das Untersuchungsergebnis systematisch und gerichtet (<i>bias</i>) oder zufällig und ungerichtet (<i>noise</i>) verfälschen [1]. Beide bleiben oft unerkannt, können jedoch durch Maßnahmen in der Verfahrenshygiene reduziert werden.

Tab. 2: Prinzipiell können drei verschiedene Arten von Unsicherheiten betrachtet werden, die unterschiedliche Konsequenzen für ihre Erkenntnung und Kommunikation haben

Observer-Variabilität bei der Formulierung von vagen oder schwierigen Diagnosen oder dem Einschätzen des Anteils teilungsaktiver Zellen in einem Tumor [13, 14]. Hier können wiederholte Bewertungen durch dieselbe Person in größeren zeitlichen Abständen oder auch zeitgleiche Einschätzungen durch verschiedene Untersuchende zu teils deutlich abweichenden Ergebnissen führen. *Bias* und *noise* gelten jedoch als „unsichtbare Feinde“, die in der spezifischen Situation den Beteiligten zumeist zwar nicht bekannt sind, deren systematische Bekämpfung sich jedoch in jedem Fall lohnt [1]. Aber wie?

Maßnahmen zur Reduktion von Unsicherheit

Bereits ein solides Verständnis der hier beschriebenen Unsicherheitsquellen sowie das ständige Bewusstsein ihrer möglichen Präsenz sollten bei allen Beteiligten zu einem adäquaten Umgang mit ihnen führen. Dazu soll dieser Artikel beitragen. Darauf hinaus existieren zahlreiche, gezielt einsetzbare Möglichkeiten, um bestimmte Fehlerquellen zu reduzieren [1].

Unter dem Begriff der „Verfahrenshygiene“ können eine Vielzahl von prozessorientierten, strukturierten Qualitätssicherungsmaßnahmen getroffen werden. Dazu zählen etwa „Standard Operating Procedures (SOP)“ für alle Schritte der Probennahme, -verarbeitung und -untersuchung, sowohl für dabei eingesetzte maschinelle Vorgänge als auch solche durch Personen. So kann bei der chirurgischen Tumorentfernung die Aussagekraft der histologischen Einschätzung bezüglich seiner vollständigen Entfernung dadurch erhöht werden, indem die der Chirurgin fraglichen Regionen mit farbigem Nahtmaterial markiert werden. Im pathologischen Labor kann dem Risiko einer falsch-negativen Befundung von kleinen Tumoren in einem größeren Milzhämatom durch präzise Anweisung zur „vollständigen Auflammlierung in 0,5 cm dicken Scheiben und Einbettung aller verdächtigen Veränderungen zur mikroskopischen Untersuchung“ begegnet werden. Auch für die mikroskopische Befundung wurden in einigen Fachdisziplinen Leitlinien für eine möglichst vollständige Erhebung und Kommunikation diagnostisch relevanter Kriterien formuliert [15–17]. Solche Leitlinien können auch vorsehen, die Qualität und Repräsentativität des erhaltenen Probenmaterials mittels Qualitäts-Feedback an die probennehmende Person zu kommentieren [16].

Doch auch SOP können gewisse Restrisiken nie völlig ausschließen, da dem zu betreibenden Aufwand unter praktischen Bedingungen oft Grenzen gesetzt sind. Gleichermaßen gilt für die Zahl der untersuchbaren Tumorbettbiopsien bei Umfangsvermehrungen in der Haut, der Zahl untersuchbarer zytologischer Präparate und viele ähnliche Situationen. Hier lässt sich oft der Grad der Repräsentativität der Proben erhöhen, jedoch nie absolute Sicherheit erreichen. Besonders in der Tiermedizin sind oft enge Grenzen des zu betreibenden Aufwands und der damit zu erreichenden Gewissheit durch die dabei entstehenden Kosten für die Patientenbesitzenden gesetzt.

Eine übliche probate Maßnahme bei diagnostischen Unsicherheiten ist das Einholen von Zweit- oder sogar Drittmeinungen durch kompetente Fachleute. Der Einsatz digitalisierter Präparate und ihr Versand oder ihre Betrachtung über das Internet (Telepathologie) erleichtern dieses Vorgehen erheblich. Zweitmeinungen sind stets zunächst unvoreingenommen zu erzielen, d. h. ohne vorherige Mitteilung der fraglichen Ersteinschätzung, bevor sich eventuell eine gemeinsame Falldiskussion anschließt. Hierzu werden in der Humanmedizin zunehmend Leitlinien und standardisierte diagnostische Algorithmen eingesetzt, die bei unsicheren Befundsituationen auf die zu treffenden Maßnahmen zur Diagnosesicherung abzielen. Dazu zählen auch immunhistochemische oder molekulare Verfahren sowie das gezielte Nachoptieren. Solche Leitlinien können den individuellen Ermessensspielraum und damit Fehlerquellen einschränken. Obwohl Zweitmeinungen und die beispielhaft genannten Folge-

untersuchungen in der Tiermedizin bereits gängige Praxis sind, fehlen strukturierte Leitlinien für ihren Einsatz noch weitgehend.

Die mit Unsicherheiten verbundenen Risiken können u. a. reduziert werden durch

- die Kenntnis ihrer Formen und Ursachen
- eine strukturierte Verfahrenshygiene
- einen aktiv selbst hinterfragenden Arbeitsstil und
- offene Kommunikation

Quantensprünge in der Reduktion von Unsicherheiten werden in Zukunft durch den Einsatz von Computern, hier besonders der künstlichen Intelligenz (KI), erwartet, v. a. bei der Auswertung komplexer Bilder wie in der Radiologie, Pathologie und Zytologie. Die meisten Ansätze befinden sich in der Tiermedizin jedoch noch in der Erprobungsphase [14]. In komplexeren Entscheidungsprozessen wird KI jedoch – zumindest in naher Zukunft – das menschliche Urteilsvermögen nicht ersetzen können, darin sind sich viele Fachleute einig [1].

Bei allen verfahrenstechnischen Möglichkeiten bleiben jedoch das strukturierte Vorgehen und der selbstkritische Geist einer möglichst gut ausgebildeten und erfahrenen Person die zentralen Elemente bei der Reduktion von Fehlern. Auch die zunehmende fachliche Spezialisierung kann von großem Vorteil sein, solange die eigenen Kompetenzgrenzen berücksichtigt werden. Hinzu kommen die solide Kenntnis der in diesem Beitrag geschilderten Fehlerquellen, ihre fallspezifische Abwägung und die offene Kommunikation erkennbarer Unsicherheiten mit den Auftraggebenden.

Zitierte und weiterführende Literatur

(weitere Quellen beim Erstautor)

- [1] Kahneman D, Sibony O, Sunstein CR, Schmidt T (2023): Noise: Was unsere Entscheidungen verzerrt – und wie wir sie verbessern können. Pantheon-Verlag.
- [2] Herman EJ, Stern AW, Fox RJ, Dark MJ (2019): Understanding the efficiency of splenic hemangiosarcoma diagnosis using monte carlo simulations. *Vet Pathol.* 56(6): 856–859.
- [3] Yang C, Humphrey PA (2020): False-negative histopathologic diagnosis of prostatic adenocarcinoma. *Arch Pathol Lab Med.* 144(3): 326–334.
- [4] Köhler H, Kraft H (1984): Gerichtliche Veterinärmedizin. Enke-Verlag.
- [5] Eikmeier H, Fellmer E, Moegle H (1990): Lehrbuch der Gerichtlichen Tierheilkunde. Paul Parey-Verlag.
- [6] Steidl T, Buyle T, Bostedt H, Wehrend A (2020): Rechtssicherheit in der Tierarztpfaxis - Gerichtliche Veterinärmedizin für den Praxisalltag. Thieme-Verlag.
- [7] Kölner N, Nissen K, Rieß M, Sadorf E (2004): Probabilistische Schlussfolgerungen in Schriftgutachten. Luchterhand-Fachverlag.
- [8] Baumgärtner W, Gruber A (2020): Allgemeine Pathologie für die Tiermedizin. 3. Auflage, Thieme-Verlag.
- [9] Forster B (1986): Praxis der Rechtsmedizin für Mediziner und Juristen. Thieme-Verlag.
- [10] <https://www.colliseum.eu/wiki/Wahrscheinlichkeit>, letztes Abrufdatum 21.05.2024.
- [11] Trück T (2023): Münchener Kommentar zur StPO, 2. Aufl. 2023, § 72, Rn. 2. C.H.Beck-Verlag sowie BGH, Urt. v. 7.6.1956 – 3 StR 136/56.
- [12] BGH, Beschl. v. 25.05.2011 – 2 StR 585/10.
- [13] Czerwak A, Schmidt MJ, Olszewska A, Hinz S, Büttner K, Farke D (2023): Reliability and interobserver variability of a grading system of ventricular distension in dogs. *Front Vet Sci.* 10: 1271545.
- [14] Bertram CA, Gurtner C, Dettwiler M, Kershaw O, Dietert K, Pieper L, Pischon H, Gruber AD, Klopflisch R (2018): Validation of digital microscopy compared with light microscopy for the diagnosis of canine cutaneous tumors. *Vet Pathol.* 55(4):490–500.
- [15] Kamstock DA, Ehrhart EJ, Getzy DM, Bacon NJ, Rassnick KM, Moroff SD, Liu SM, Straw RC, McKnight CA, Amorim RL, Bienville D, Cassali GD, Cullen JM, Dennis MM, Esplin DG, Foster RA, Goldschmidt MH, Gruber AD, Hellmén E, Howerth EW, Labelle P, Lenz SD, Lipscomb TP, Locke E, McGill LD, Miller MA, Mouser PJ, O'Toole D, Pool RR, Powers BE, Ramos-Vara JA, Rocabianca P, Ross AD, Sailasuta A, Sarli G, Scase TJ, Schulman FY, Shoieb AM, Singh K, Sledge D, Smedley RC, Smith KC, Spangler WL, Steficek B, Stromberg PC, Valli VE, Yager J, Kupel M (2011) Recommended guidelines for submission, trimming, margin evaluation, and reporting of tumor biopsy specimens in veterinary surgical pathology. *Vet Pathol.* 48(1): 19–31.
- [16] Day MJ, Bilzer T, Mansell J, Wilcock B, Hall EJ, Jergens A, Minami T, Willard M, Washabau R (2008): Histopathological standards for the diagnosis of gastrointestinal inflammation in endoscopic biopsy samples from the dog and cat: a report from the World Small Animal Veterinary Association Gastrointestinal Standardization Group. *J Comp Pathol.* 138 Suppl 1: S1–43.
- [17] Washabau RJ, Day MJ, Willard MD, Hall EJ, Jergens AE, Mansell J, Minami T, Bilzer TW (2010): Endoscopic, biopsy, and histopathologic guidelines for the evaluation of gastrointestinal inflammation in companion animals. WSAVA International Gastrointestinal Standardization Group. *J Vet Intern Med.* 24(1): 10–26.

Eine englische Version des Beitrags ist abrufbar unter: www.dvg.de/wp-content/uploads/2024/06/24-06-11-How-reliable-is-your-diagnosis_-Gruber-et-al_-2024_A4.pdf



Korrespondenz



Prof. Dr. Achim D. Gruber, Ph.D.

Institut für Tierpathologie,
Freie Universität Berlin,
Robert von Oستertag Str 15, 14163 Berlin,
Tel. +49 30 838 62440, achim.gruber@fu-berlin.de

Weitere Autorenschaft:

- Prof. Dr. Dr. h. c. Axel Wehrend, Tierklinik für Reproduktionsmedizin und Neugeborenenkunde, Frankfurter Str. 106, 35392 Gießen
 Dr. Thomas Steidl, Brühlstr. 60, 72147 Nehren
 Dr. Katharina Kramer, Landeslabor Schleswig-Holstein, Max-Eyth-Str. 5, 24537 Neumünster
 Dr. Sabine Merbach, Chemisches und Veterinäruntersuchungsamt Westfalen, Zur Taubeneiche 10-12, 59821 Arnsberg
 Prof. Dr. Dr. h. c. Wolfgang Baumgärtner, Ph.D., Institut für Pathologie, Stiftung Tierärztliche Hochschule Hannover, Bünteweg 17 30559 Hannover
 Dr. Thomas Buyle, Tierärztliche Praxis für Pferde, Frauenländerstr. 71, 71394 Kernen im Remstal
 Prof. Dr. Niclas-Frederic Weisser, Hochschule für Öffentliche Verwaltung Bremen, Doventorscontrescarpe 172 C, 28195 Bremen
 Prof. Dr. Andreas Beineke, Institut für Pathologie, Stiftung Tierärztliche Hochschule Hannover, Bünteweg 17 30559 Hannover