

Klinisches Denken, ärztliche Kunst und Artificial Intelligence

Stefano Bassetti, Martin C. Hirsch, Edouard Battegay

Als „klinisches Denken“ bezeichnet man alle Denkprozesse von Ärztinnen und Ärzten zum Stellen einer Diagnose und zur Festlegung eines Therapie- und Betreuungsplans. Artificial Intelligence (AI) wird das menschliche klinisch-diagnostische Denken erweitern, verbessern und beschleunigen, es aber vermutlich nicht ersetzen. Ihre Anwendung in der Medizin hat das Potenzial, medizinische Diagnosefehler drastisch zu reduzieren und den Ärzten mehr Zeit für die Betreuung ihrer Patienten zu verschaffen.

ABKÜRZUNGEN

COPD	chronisch-obstruktive Lungenerkrankung
EbM	evidenzbasierte Medizin
LLM	Large Language Model
NHS	National Health Service

Was ist das klinische Denken?

Das klinische Denken („Clinical Reasoning“) wird von Trowbridge definiert als „kognitiver und nicht kognitiver Prozess, mit dem eine Fachperson im Gesundheitswesen bewusst und unbewusst mit dem Patienten und der Umgebung interagiert, um Patientendaten zu sammeln und zu interpretieren, Nutzen und Risiken von Handlungen abzuwägen und die Präferenzen des Patienten zu verstehen, um schließlich eine Arbeitsdiagnose zu stellen und einen Behandlungsplan zu definieren mit dem Ziel, das Wohlbefinden des Patienten zu verbessern“ [1].

Das klinische Denken besteht somit aus 2×2 Hauptelementen: dem diagnostischen Denken und therapeutischen Denken (Management) sowie dem kognitiven und nicht kognitiven (teilweise intuitiven) Prozess. Das diagnostische Denken führt zur Formulierung einer Arbeits- oder Verdachtsdiagnose. Diese basiert auf der Erwägung und Gegenüberstellung von in einer bestimmten Situation infrage kommenden Diagnosen (differenzialdiagnostisches Denken). Das klinische Denken wird durch Arzt-Faktoren (z. B. Fachwissen, Erfahrung, Persönlichkeit, Emotionen), Patienten-Faktoren (z. B. Multimorbidität), Umfeld (z. B. Team) und durch den Kontext beeinflusst, in dem das Denken stattfindet (z. B. Hausarztpraxis vs. Intensivstation) (► **Abb. 1**).

Warum ist das klinisch-diagnostische Denken wichtig?

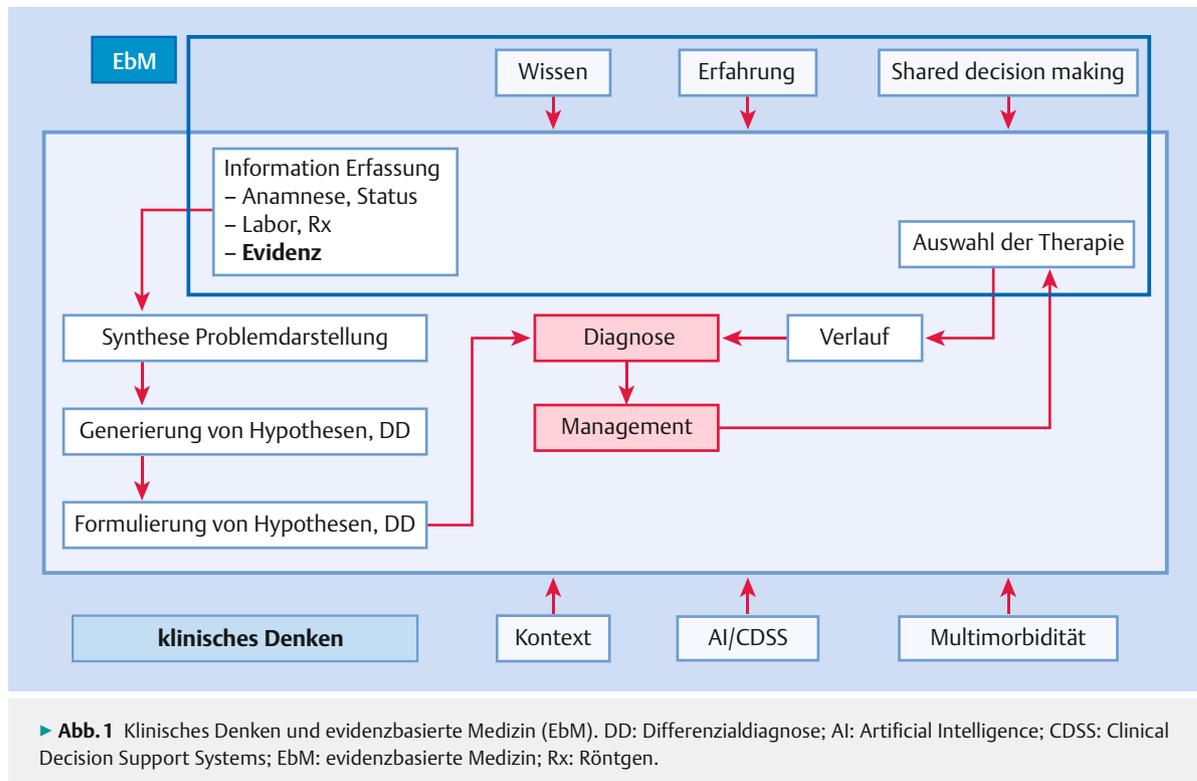
Ärztinnen und Ärzte perfektionieren laufend ihr diagnostisches Denken, um möglichst schnell und effizient richtige Diagnosen zu stellen [2]. Letztere sind die unerlässliche Basis für eine erfolgreiche Behandlung.

Andererseits führt ein defizitäres diagnostisches Denken zu Fehldiagnosen, mit gravierenden Konsequenzen für Patientinnen und Patienten. So sollen medizinische Fehler die dritthäufigste Todesursache in den USA sein, nach Herz-Kreislauf-Erkrankungen und Krebs [3], und Diagnosefehler sind die häufigste Art von Fehlern in der Medizin [4]. Schätzungsweise 795000 Amerikaner erleiden eine permanente Behinderung oder sterben jährlich als Folge eines Diagnosefehlers [5]. Diagnosefehler haben meistens mehrere Ursachen: Kognitive Faktoren, wie zum Beispiel eine fehlerhafte Synthese des Problems, tragen am häufigsten zu einem Diagnosefehler bei (in ca. 75 % der Fälle), gefolgt von systembedingten Faktoren (in 65 %; z. B. Probleme mit Standards und Prozeduren) [6].

Wie läuft das klinisch-diagnostische Denken ab?

Das klinisch-diagnostische Denken ähnelt Prozessen, welche dem Denken und der Entscheidungsfindung im Allgemeinen zugrunde liegen. Ein hilfreicher Erklärungsansatz für diese Prozesse zeigt die **Dual-Prozess-Theorie**, welche die Anwendung zweier paralleler Denksysteme bei unserem Entscheiden beschreibt:

- Das **schnelle Denken** (Typ 1) ist assoziativ, läuft automatisch, unbewusst und ohne Anstrengung ab und basiert auf Mustererkennung, „Bauchgefühl“ und **Heuristiken** (d. h. Faustregeln und mentalen Abkürzungen, die es erlauben, rasch Entscheidungen und



Urteile zu treffen, ohne weitere langwierige Analysen durchführen zu müssen) [7]. Beispiel: Beim Betrachten eines Hautbefunds mit gruppierten Bläschen auf gerötetem Grund, die im Bereich des rechten Hemithorax bandförmig angeordnet sind, wird der erfahrene Arzt sofort, assoziativ und ohne Anstrengung das Muster wiedererkennen und die Blickdiagnose eines Herpes Zoster stellen.

- Das **langsame Denken** (Typ 2) ist analytisch, mühevoll, läuft bewusst ab und basiert auf hypothetisch-deduktiven Denkprozessen [7]. Beispiel: Beim Betrachten des gleichen oben beschriebenen Hautbefunds wird der Student, der noch nie einen Zoster gesehen hat, kein Muster erkennen und keine Assoziationen haben. Er wird deshalb bewusst mittels langsamen Denkens die Situation analysieren müssen und überlegen, welche Krankheit gruppierte kutane Bläschen auf gerötetem Grund mit dermatomaler Verteilung und neuralgiformen Schmerzen bei einem Patienten über 50 Jahre verursachen könnte.

Der Anteil des schnellen und langsamen Denkens am diagnostischen Denken und am Entscheidungsprozess variiert je nach Aufgabe, Situation und Erfahrung. Je „typischer“ und klarer eine Situation ist und je größer die Erfahrung einer Person ist, desto häufiger wird vorwiegend assoziativ und schnell gedacht. Erfahrene Klinikerinnen und Kliniker treffen im beruflichen Alltag ca. 95% der Entscheidungen durch schnelles Denken [8]. In komplexen oder unbekanntem Situationen, in denen man durch schnelles, assoziatives Denken keine Diagno-

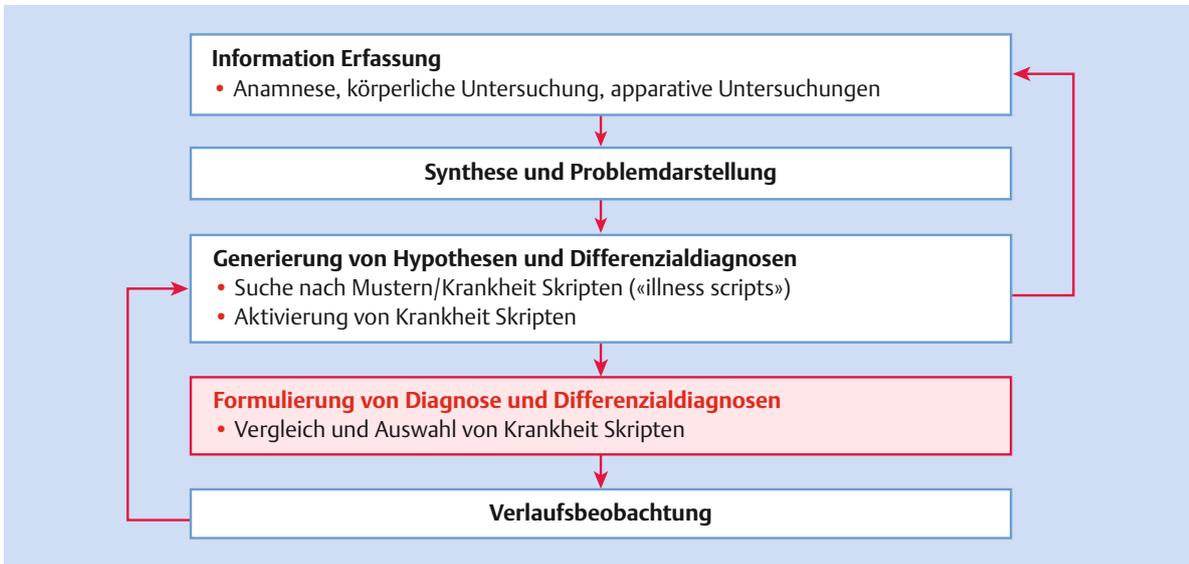
sen oder Differenzialdiagnosen generieren kann, muss man analytisches, langsames Denken anwenden. Erfahrene Kliniker wechseln immer wieder rasch hin und her zwischen assoziativem und analytischem Denken. Sie erkennen außerdem, oft intuitiv, wenn gewisse Aspekte eines Falles nicht zu einem assoziativ suggerierten Muster passen, und schalten dann auf langsames, analytisches Denken um.

Elemente des diagnostischen Denkens

Die Schlüsselemente des diagnostischen Denkens sind in ► **Abb. 2** dargestellt. Zentral sind Anamnese und körperliche Untersuchung, die bereits in der Mehrheit der Fälle zur korrekten Diagnose führen. Bei einem Kollektiv von notfallmäßig hospitalisierten internistischen Patienten war die Anamnese das wichtigste einzelne Instrument zur Diagnosestellung: 20% der Diagnosen konnten allein durch die Anamnese gestellt werden, weitere 40% durch die Kombination von Anamnese und körperlicher Untersuchung. Die Kombination von Anamnese, körperlicher Untersuchung und zusätzlichen apparativen Basisuntersuchungen (Basis-Labor, EKG, Thorax-Röntgenbild) ermöglichte eine Diagnose in weiteren 33% der Fälle [9].

Informationserfassung

Die **Anamnese** soll zunächst 1 bis möglichst 2–4 Leitsymptome semantisch genau definieren (► **Abb. 3**): Die Wahrscheinlichkeit unterschiedlicher Diagnosen ist z. B. bei „Thoraxschmerzen“ eine andere, wenn eine Patientin



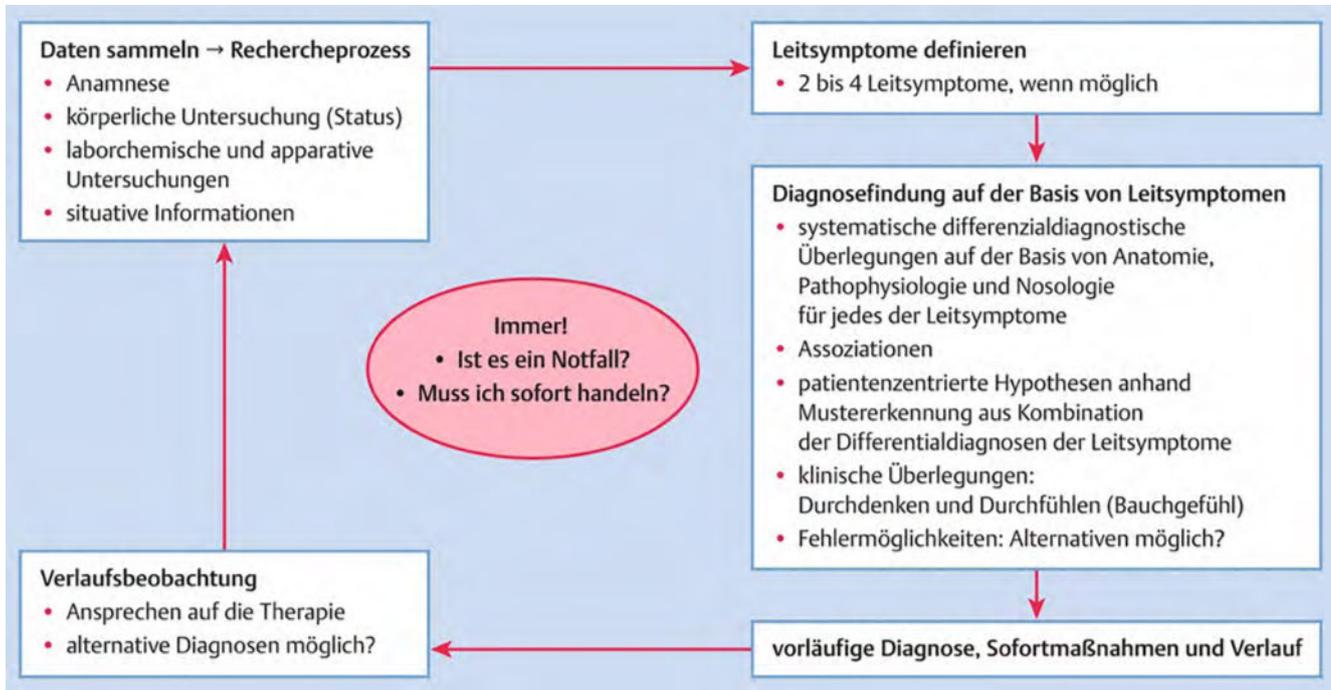
► **Abb.2** Hauptelemente des diagnostischen Denkens [1, 10].

sich mit „belastungsabhängigen retrosternalen Thoraxschmerzen“ vorstellt (bei einem akuten Koronarsyndrom) oder mit „atemabhängigen Thoraxschmerzen rechts“ (bei einer Pleuritis). Oft hilft auch die Kombination von 2 oder gar 3 Leitsymptomen: „atemabhängige Thoraxschmerzen rechts, nach 10 Stunden dauernder Flugreise“ oder „atemabhängige Thoraxschmerzen rechts und Schmerzen in der linken Wade“. Diese Kombination suggeriert in Sekundenschnelle die mögliche Diagnose Lungenembolie oder im zweiten Fall Venenthrombose und Lungenembolie. Die Kombination „atemabhängige Thoraxschmerzen rechts und leichte Hämoptoe“ beinhaltet auch die Differenzialdiagnose Lungenembolie, ist jedoch breiter. Könnte es sich alternativ um ein Lungenkarzinom, eine chronisch-obstruktive Lungenerkrankung (COPD) mit Husten und Rippenfraktur, ein Mesotheliom, eine Tuberkulose, oder gar ein Goodpasture-Syndrom handeln? Danach geht man zur weiteren Einengung der Diagnose mit einer Ausschluss- oder Einschlussheuristik vor („Haben Sie Fieber oder Fiebergefühl? Wie lange dauert das Husten schon? Wie ist der Appetit? Hatten Sie eine lange Flugreise?“). Zudem werden mit der Anamnese Risikofaktoren (z. B. Rauchen, medikamentöse oder hormonelle Behandlungen ...) und Expositionen (z. B. epidemiologischer Kontext, berufliche Exposition, Aufenthalt in den Tropen ...) ermittelt, welche an spezifische Krankheiten denken lassen und die Wahrscheinlichkeit gewisser Erkrankungen erhöhen oder vermindern.

Das Gleiche gilt für die **körperliche Untersuchung**. Die Befunde müssen genau erhoben werden: Es ist wesentlich für die korrekte Diagnosestellung, ob z. B. bei einem schmerzhaften Gelenk klinische Zeichen einer Entzündung (Rötung, Schwellung) festzustellen sind (und somit

eine Arthritis vorliegt) oder eben nicht (Arthralgie). Einige Diagnosen werden allein durch die klinische Untersuchung gestellt, z. B. bei dermatologischen Krankheiten oder wenn mittels Auskultation eines Perikardreibens eine Perikarditis diagnostiziert werden kann. In den meisten Situationen wird aber mit der klinischen Untersuchung wiederum die Wahrscheinlichkeit für das Vorliegen gewisser Krankheiten verändert. Das diagnostische und differenzialdiagnostische Denken begleitet die Erhebung der Anamnese und der klinischen Untersuchung als kontinuierlicher, iterativer Prozess: Je mehr Informationen gesammelt werden, desto mehr Hypothesen werden generiert (assoziativ und aktiv analytisch), die mit immer genaueren anamnestischen Nachfragen und klinischen Untersuchungen eingegrenzt, verifiziert oder ausgeschlossen werden sollen. Wenn das Leitsymptom eine progrediente, belastungsabhängige Atemnot ist, mit beidseits eindrückbaren Unterschenkel-Ödemen, wird man z. B. nach weiteren Symptomen einer Herzinsuffizienz fragen (z. B. nach einer Orthopnoe) und nach weiteren klinischen Zeichen der Herzinsuffizienz suchen (Untersuchung der Halsvenen, hepatojugulärer Reflux usw.).

Anamnese und klinische Untersuchung erfolgen zwar nach einer gewissen Struktur (Leitsymptom, jetziges Leiden, Systemanamnese ...). Auf der Basis dieser Grundstruktur soll aber im Verlauf des Gesprächs und der Untersuchung ein hypothesengesteuertes Vorgehen zu einer vollständigeren und besseren Informationserfassung beitragen. Wenn man z. B. während der Anamnese und Untersuchung die Hypothese einer Herzinsuffizienz in Betracht zieht, ist die Wahrscheinlichkeit viel höher, dass nach allen typischen Symptomen der Herzinsuffizienz gefragt wird und genauer nach den typischen Zeichen gesucht wird, als wenn man nur „checklistenartig“ Anamnese und Status



► **Abb. 3** Der diagnostische Prozess (Quelle: Battagay E, Hrsg. Differenzialdiagnose Innerer Krankheiten. Vom Symptom zur Diagnose. 21. Aufl. Stuttgart: Thieme; 2017).

„abarbeitet“. Die hypothesengesteuerte, aktive Untersuchung führt auch dazu, dass Befunde genauer beschrieben und erhoben werden, was ihre Aussagekraft erhöht: Sind z. B. die peripheren Unterschenkel-Ödeme tatsächlich eindrückbar und eher kardialer Ursache, oder handelt es sich eher um Lymphödem oder Lipödem? Oft findet und erkennt man zudem Befunde nur, wenn man weiß, was man sucht.

Auch die Anordnung und Analyse sowie die Interpretation der Ergebnisse der **apparativen Diagnostik** (Labor, Bildgebung, anderer apparativen Diagnostik) sollte kontextuell und hypothesengesteuert aufgrund der geschätzten Vortest-Wahrscheinlichkeit für eine bestimmte Krankheit erfolgen. Die Schätzung dieser Vortest-Wahrscheinlichkeit basiert auf der Anamnese, der klinischen Untersuchung und dem klinischen Kontext.

Synthese und Problemdarstellung

Die so erfassten relevanten Daten und Informationen werden als Problemdarstellung meistens in einem Satz zusammengefasst. Dieser Satz soll die Patientensituation mit charakterisierenden Begriffen (semantischen Qualifikatoren) eingrenzen und gegebenenfalls im Kontext der zugrunde liegenden pathophysiologischen Mechanismen „einkapseln“ [1, 10].

Beispiel: „58-jähriger febriler Patient mit bekannter koronarer Herzkrankheit und aktuell akuter, blutiger nosokomialer Diarrhö und peritonitischem Abdomen“.

- Semantische Qualifikatoren: febril, akut, blutig, nosokomial.
- Einkapselung: koronare Herzkrankheit, Peritonismus.

Beim Patienten wurde eine ischämische Kolitis diagnostiziert.

Semantische Qualifikatoren sind oft binäre Begriffe und grenzen die Bedeutung eines Symptoms, eines klinischen Zeichens oder einer Krankheit besser ein. Im oben erwähnten Beispiel: febril versus afebril, akut versus chronisch, blutig versus nicht blutig und nosokomial versus ambulant erworben. Semantische Qualifikatoren helfen, die Patienteninformationen zu organisieren und in Informationspakete zu bündeln. Diese Informationspakete benutzt das diagnostische Denken, um diagnostische Hypothesen (Differenzialdiagnosen) zu vergleichen und gegenüberzustellen (bei unserem Patienten z. B. virale oder bakterielle Gastroenteritis versus *Clostridioides-difficile*-assoziierte Kolitis versus ischämische Kolitis) [1, 10].

Klinikerinnen und Kliniker speichern ihr Wissen über Krankheiten und Symptome als **Krankheitskript** ab. Ein Skript ist ein mentales Schema, das die relevanten Symptome und klinischen Zeichen einer bestimmten Krankheit beinhaltet, typischerweise strukturiert nach prädisponierenden Faktoren, pathophysiologischem Mechanismus und klinischen Folgen. Durch die klinische Tätigkeit und die zunehmende Erfahrung werden die Skripten kontinuierlich entwickelt. Sie organisieren das klinische Wissen und machen es für das diagnostische Denken verfügbar,

indem Problemdarstellung und Skripten in Verbindung gesetzt werden [1, 10]. Wenn ein Arzt einen Patienten erstmals trifft, verbindet er sofort, schnell und unbewusst die Situation und den Kontext, und die verbalen und nicht verbalen, kognitiven und nicht kognitiven Informationen und Eindrücke mit verschiedenen Krankheitsskripten als mögliche Differenzialdiagnosen bei diesem Patienten. Diese Verbindung wird **Skriptaktivierung** genannt. Sie kann auch bewusst durch langsames, analytisches Denken erfolgen.

Die **Einkapselung**[1] der Problemdarstellung mit pathophysiologischen Mechanismen oder Syndromen erlaubt eine weitere Präzisierung und Bündelung der Informationen und erleichtert die Assoziation mit abgespeicherten Skripten. Im oben angeführten Beispiel erhöht die Präsenz einer koronaren Herzkrankheit die Wahrscheinlichkeit einer Atherosklerose der Mesenterialgefäße, und der Peritonismus lässt sofort an eine mögliche Darmperforation denken.

Merke

Je mehr genaue und spezifische Informationen sowie spezifische beschreibende Begriffe (Qualifikatoren) in einer Problemdarstellung beinhaltet sind, desto höher ist die Wahrscheinlichkeit, dass richtige Assoziationen mit abgespeicherten Krankheiten entstehen.

Eine logische Darstellung des Problems (z.B. Vorstellung des Patienten nach dem SOAP-Schema), welche analog einem Krankheitsskript strukturiert ist, erleichtert die Skriptaktivierung.

Die korrekte Erkennung der Schlüsselinformationen und die Verwendung der spezifischen Begriffe führen rasch zu korrekten Assoziationen. Umgekehrt führen eine ungenaue Synthese und Problemdarstellung mit Begriffen, die für andere Situationen spezifischer sind, rasch zu Fehl-assoziationsen und falschen Diagnosen (z.B. „abszedierende Pneumonie“ versus „Kaverne“ in der Lunge).

Generierung von Hypothesen (Arbeitsdiagnosen und Differenzialdiagnosen)

Vom ersten Augenblick der Arzt-Patienten-Begegnung an integriert die Ärztin oder der Arzt erste Eindrücke (manchmal gar Blickdiagnosen), weitere Eindrücke, Eigenschaften, Informationen (z.B. blasses Hautkolorit, Schweiß auf der Stirn, psychomotorische Unruhe, tachykarder Puls, fauliger Stuhlgeruch im Zimmer) laufend automatisch, unbewusst und schnell in Mustern. Diese werden ebenfalls unbewusst und schnell mit abgespeicherten Krankheitsskripten verglichen (z.B. mit dem Skript „anämisierende obere gastrointestinale Blutung mit Meläna“) – es findet also eine Skriptaktivierung statt. Diese nicht analytische Mustererkennung wird kombiniert mit analytischem Denken, indem die verschiedenen aktivierten Skripten verglichen

und gegenübergestellt werden (z.B. Tachykardie, psychomotorische Unruhe und Schwitzen bei hämorrhagischem Schock versus septischem Schock), und indem weitere für die Skripten typischen Eigenschaften aktiv gesucht werden (z.B. Hämatemesis, kalte Extremitäten, Nachweis von Teerstuhl bei der Rektaluntersuchung versus Fieber, warme Extremitäten usw.). Die erfahrene Klinikerin oder der erfahrene Kliniker kann auf viel mehr und genauere Skripten zurückgreifen als die Unerfahrenen. Entsprechend Erfahrene können deshalb mehr Muster erkennen und assoziieren und wissen, welche zusätzlichen spezifischen Clues/Eigenschaften/Informationen zu suchen sind. Sie benutzen spezifischere Qualifikatoren zur Problemdarstellung und sind somit in der Lage, die verschiedenen aktivierten Skripte besser vergleichend in Kontrast zu setzen. Wenn der Fall komplex oder ungewöhnlich ist, oder wenn die Ärztin oder der Arzt wenig Erfahrung hat, müssen Hypothesen durch analytisches, langsames Denken erzeugt werden (z.B. Analyse der möglichen Ursachen einer Tachykardie), weil keine Muster erkannt werden und keine Assoziationen mit Krankheitsskripten entstehen.

In diesen Situationen werden mit **probabilistischem Denken** und einer **Bayes-Analyse** aufgrund von z.B. epidemiologischen Daten (Prävalenz, Vortest-Wahrscheinlichkeit) erste Hypothesen aufgestellt. Bei einem 70-jährigen Raucher mit einem Diabetes mellitus Typ II, der sich wegen neuen belastungsabhängigen retrosternalen Thoraxschmerzen vorstellt, ist bspw. die Wahrscheinlichkeit eines akuten Koronarsyndroms aufgrund des hohen kardiovaskulären Risikos sehr hoch. **Induktiv** werden weitere Hypothesen formuliert – auf der Basis gelernten oder erarbeiteten Wissens (Ätiologien, Anatomie, Pathogenese und Pathophysiologie).

Merke

Bei allen differenzialdiagnostischen Überlegungen ist es im Alltag wichtig, dass man initial aktiv lebensbedrohliche oder gefährliche Krankheiten (Skripten) in Betracht zieht, welche nicht verpasst werden dürfen und/oder sofort behandelt werden müssen (► Abb.3). Die entscheidende erste Frage also ist: Ist es schlimm?

Das fragen sich auch die Patientin und der Patient, das ebenfalls involvierte Gesundheitspersonal sowie Angehörige der Patientin oder des Patienten. Bei Thoraxschmerzen muss z.B. sofort überlegt werden, ob ein Myokardinfarkt vorliegen könnte.

Formulierung von Diagnose und Differenzialdiagnosen

Die assoziativ und analytisch identifizierten Krankheitsskripte sind die hypothetischen Diagnosen, welche in einem Fall infrage kommen (Differenzialdiagnosen). Sie werden nach Wahrscheinlichkeit priorisiert: Je größer die Übereinstimmung zwischen Eigenschaften des Falles

und einem Skript ist, desto sicherer ist die Diagnose. Die Diagnose muss präzise formuliert werden, weil das weitere diagnostische und therapeutische Vorgehen von ihr abhängt. Oft ist die Diagnose initial noch unklar. Trotzdem sollte man sich festlegen und zumindest eine explizite Verdachtsdiagnose formulieren, mit den entsprechenden Differenzialdiagnosen. Ist eine Situation noch absolut unklar, ist es hilfreich, eine beschreibende Diagnose festzuhalten, d. h. die Symptome und Zeichen, die führend erscheinen, mit einer breiten, nach Wahrscheinlichkeit priorisierten Differenzialdiagnose, zu notieren.

Verlaufsbeobachtung

Die Verlaufsbeobachtung ist ein weiterer wichtiger diagnostischer Baustein (► **Abb. 3**). Eine Diagnose muss immer wieder kritisch hinterfragt werden. Manchmal sind initial gestellte Diagnosen schlichtweg falsch.

Merke

Die gestellten Diagnosen und Differenzialdiagnosen sollen deshalb aufgrund des Verlaufs kontinuierlich erneuert, kritisch hinterfragt und angepasst werden.

Klinisches Denken und evidenzbasierte Medizin

Ende des 20. Jahrhunderts prägte die Gruppe von David Sackett u. Kollegen den Begriff „evidenzbasierte Medizin“ (EbM). Sie definierte **EbM** als **Integration der besten verfügbaren wissenschaftlichen Evidenz mit klinischer Expertise und mit den Werten des Patienten** [11]. Somit bildet die EbM die Basis eines patientenzentrierten klinischen Denkens mit den 3 Kernelementen

- „Patientendaten sammeln und interpretieren“,
- „Nutzen und Risiken von Handlungen abwägen“ und
- „Präferenzen des Patienten verstehen“ (► **Abb. 1**).

Mit **klinischer Expertise** ist die Fähigkeit gemeint, die klinischen Fertigkeiten und die klinische Erfahrung zu nutzen, um den individuellen Gesundheitszustand und die Diagnose eines jeden Patienten, seine individuellen Risiken und Vorteile von potenziellen Interventionen sowie seine persönlichen Werte und Erwartungen rasch zu identifizieren.

Die **Werte des Patienten** sind seine einzigartigen Präferenzen, Sorgen und Erwartungen, die jeder Patient in eine klinische Begegnung mitbringt und die in die klinischen Entscheidungen einbezogen werden müssen, wenn sie dem Patienten dienen sollen. Mit der Integration der 3 Elemente der EbM (Evidenz + klinische Expertise + Patientenwerte) verbessern die Beteiligten die klinischen Ergebnisse und die Lebensqualität der Patienten [11].

Seit der Einführung der EbM kam es oft auch in der ärztlichen Weiter- und Fortbildung zu einer einseitigen Fokussierung und Reduktion der EbM auf die Suche und Evaluation der Evidenz aus wissenschaftlichen Studien. Dabei gerieten die 2 anderen essenziellen Bestandteile der EbM (klinische Expertise und Patientenwerte) mehr und mehr in den Hintergrund. Das führte zur Kritik, dass die EbM krankheits-, aber nicht patientenzentriert sei, bspw. weil die Resultate von Studien, die in selektionierten Patientenkollektiven durchgeführt werden, nicht auf individuelle Patienten (z. B. polymorbide oder alte Patienten) übertragen werden können, welche die Einschlusskriterien der Studie nicht erfüllt hätten. Andererseits wurde der Fehlschluss propagiert, dass Therapien oder Maßnahmen, für die es (noch) keine genügende wissenschaftliche Evidenz gibt, grundsätzlich nicht durchgeführt werden dürfen. Diese Problematik manifestierte sich sehr eindrücklich z. B. zu Beginn der COVID-19-Pandemie, als einige Behörden und Experten (meistens ohne klinische Erfahrung auf dem Gebiet der Infektionskrankheiten) den Nutzen einer Maske zum Schutz gegen eine SARS-CoV2-Übertragung bestritten. Sie führten das situativ absurde Argument an, dass es dafür keine genügende Evidenz gebe, weil zu diesem Thema (noch) keine oder keine qualitativ genügenden Studien publiziert worden seien. Die jahrhundertalte klinische Erfahrung mit Masken zum Schutz gegen übertragbare respiratorische Erkrankungen, bei gleichzeitig fehlendem Schadenpotenzial, wurde dabei ignoriert.

Aber ärztliche Tätigkeit, die auf der EbM basiert, darf und kann sich nicht auf die wissenschaftliche Analyse allein und die klinische Anwendung von Studienergebnissen beschränken. Sackett hatte schon 1996 klargestellt, dass die wissenschaftliche Evidenz die Grundlage für Entscheidungen liefern soll, aber nie die klinische Expertise ersetzen kann [12]. Nur die Ärztin und der Arzt mit ihrer klinischen Expertise können entscheiden, ob die externe Evidenz auf ihren individuellen Patienten überhaupt übertragbar ist, und wenn ja, ob das den Erwartungen und Wünschen ihres Patienten entspricht („Shared Decision Making“). Zudem gibt es oft Situationen, für die es keine klare Evidenz gibt. So stellt ein strategisches Papier des National Health Service (NHS) Großbritanniens fest, dass nur ca. 15% der Faktoren in NHS-Datenbanken gespeichert werden, die als Basis der Entscheidungen des Gesundheitspersonals dienen. Also beeinflussen sehr viele weitere Faktoren und Umstände ärztliche Entscheidungen [13]. Ärztinnen und Ärzte werden einerseits die verfügbare Evidenz berücksichtigen, aber vorwiegend aufgrund ihrer Erfahrung und Expertise Entscheidungen treffen (müssen). Oder Maßnahmen, die nie in randomisierten Studien belegt wurden, müssen trotzdem durchgeführt werden (z. B. Antibiotikatherapie zur Behandlung einer bakteriellen Meningitis) [12].

Merke

Es ist ein entscheidender Unterschied, ob eine Maßnahme in Studien nicht wirksam war, oder ob es keine Evidenz für diese Maßnahme gibt, einfach weil sie nie in Studien untersucht wurde: „Absence of Evidence is not Evidence of Absence“ [14].

Cave

Alle 3 Elemente der EbM müssen also zwingend kombiniert und zusammen angewandt werden, um die Patienten optimal zu behandeln und zu betreuen.

Ist die Medizin eine Wissenschaft oder eine Kunst? Was ist die „ärztliche Kunst“?

Heute assoziieren wir „Kunst“ mit Kreativität und Inspiration, und „Wissenschaft“ mit Messungen und Logik [15]. Die moderne Medizin wird aufgrund ihrer gewaltigen Fortschritte meistens als Wissenschaft verstanden, und wenn man in der Umgangssprache von „ärztlicher Kunst“ spricht, meint man oft Aspekte der ärztlichen Tätigkeit, die man nicht messen kann, wie die Fähigkeit, Emotionen der Patientinnen und Patienten zu erkennen und auf diese einzugehen, oder das „Bauchgefühl“, das einem z. B. hilft, eine komplexe Situation besser zu verstehen und einzuordnen [15].

Die alten Griechen unterschieden Kunst und Wissenschaft nicht nach dem Verhältnis von genialer Intuition zu kühler Logik, sondern vielmehr durch das Ziel der Disziplin. Der Wissenschaftler arbeitete einfach, um Grundprinzipien und Gesetzmäßigkeiten zu verstehen, der Künstler, um diese Prinzipien für einen konstruktiven Zweck anzuwenden. In diesem Sinn ist die Medizin also eine Kunst und keine Wissenschaft, weil Ärztinnen und Ärzte die Natur zu einem praktischen Zweck erforschen und nutzen [15].

Die ärztliche Tätigkeit könnte als die Kunst definiert werden, welche Wissen, klinische Fertigkeiten und klinische Erfahrung einsetzt, um Patienten optimal zu betreuen. Dabei muss eine einfühlsame menschliche Beziehung zum Patienten aufgebaut werden („to take care“) und das klinische Denken mit den 3 Säulen der evidenzbasierten Medizin verbunden werden (wissenschaftliche Erkenntnisse, klinische Kompetenz und Erfahrung, Werte und Erwartungen der Patienten); (s. Infobox Elemente der „ärztlichen Kunst“).

Merke

Das Geheimnis der ärztlichen Kunst wurde 1927 von Francis W. Peabody mit dem berühmten Satz treffend zusammengefasst: „The Secret of the Care of the Patient is in Caring for the Patient“ [16].

ZUSATZINFO

Elemente der „ärztlichen Kunst“

- Aufbau einer empathischen Beziehung mit dem Patienten
- Identifikation des Problems
- Berücksichtigung und Einbezug des Kontextes (z. B. soziale Situation)
- Wünsche, Erwartungen und Sorgen des Patienten erkennen
- Diagnose stellen
- verfügbare wissenschaftliche Evidenz suchen, evaluieren und individualisiert anwenden
- partizipative Entscheidungsfindung („Shared Decision Making“)
- Behandeln
- Umgang mit Komplexität (z. B. Multimorbidität)
- Umgang mit Unsicherheit

Artificial Intelligence (AI) und klinisch-diagnostisches Denken

Als AI bezeichnet man letztlich statistik- und entropie-basierte Systeme, die auf der Basis von Feststellungen aus Datenbanken, Bildern, Klängen, Sensoren, dem Internet usw. Netzwerkstrukturen für Entscheidungen entwickeln. Diese Netzwerkstrukturen bilden sich aufgrund der Eingaben und der rückgekoppelten Qualität der Schlüsse eigenständig heraus. Solche Systeme werden daher auch als selbstlernende Systeme bezeichnet, und der Prozess als Machine Learning. AI-Systeme können die Qualität und Effizienz der ärztlichen Tätigkeit nachgewiesenermaßen und signifikant verbessern. Hier einige Beispiele:

- **Verbesserung der bildgestützten Diagnostik:** Ein AI-unterstütztes Mammografie-Screening ist genauer und reduziert die Arbeitsbelastung für die Beurteilung der Bilder um 44% [17]. Die AI-unterstützte Koloskopie reduziert die Rate von verpassten Polypen und Adenomen um 50% [18]. Spezialisierte Anwendungen zur Bilderkennung für die Analyse von Gesichtsformen zur Differenzialdiagnose genetischer Syndrome sind bereits verfügbar (<https://www.face2gene.com>).
- **AI-Systeme können medizinische Informationen adäquat synthetisieren:** ChatGPT konnte 140 Abstracts von wissenschaftlichen Artikeln mit hoher Qualität und Genauigkeit zusammenfassen und um 70% kürzen [19].
- **AI generiert akkuratere Differenzialdiagnosen als Ärztinnen und Ärzte:** Ein Large Language Model (LLM) lieferte für 302 Fallbeschreibungen mögliche Differenzialdiagnosen: mit einer deutlich höheren diagnostischen Richtigkeit als Klinikerinnen und Kliniker (Accuracy: 59% vs. 34%) [20].

- **AI kann medizinische Daten besser verarbeiten und besser „diagnostisch denken“ als Ärztinnen und Ärzte:** In der Analyse von 20 klinischen Vignetten zeigte GPT-4 eine bessere Verarbeitung der medizinischen Daten sowie bessere klinisch-diagnostische Überlegungen (Synthese, Darstellung des Problems, Differenzialdiagnosen) als Klinikerinnen und Kliniker [21].
- **Ein AI-System wird als „empathischer“ als Ärztinnen und Ärzte wahrgenommen:** 20 Personen stellten 149 klinische Szenarien dar und konsultierten schriftlich online Ärztinnen und Ärzte oder ein Chatbot (AMIE). Die diagnostische Genauigkeit der AI war ähnlich oder besser als diejenige der Ärzte. Aber die Qualität der schriftlichen Konversation mit dem Chatbot wurde in 24 von 26 Kriterien als besser beurteilt, unter anderem bezüglich der Höflichkeit, dem „ehrlich Rüberkommen“ und in puncto „Zuwendung und Engagement ausdrücken“ [22].

Diese Resultate sind eindrücklich.

Allerdings müssen für eine genauere Beurteilung dieser AI-Anwendungen mehrere relevante Limitationen dieser Studien berücksichtigt werden [23]:

- Symptome sind subjektiv und oft wechselnd und die Beschreibung und Gewichtung der Symptome bei der Eingabe in eine AI ist oft schwierig. Sie wird aber vom Kliniker beinahe automatisch berücksichtigt. Um welchen „Brustschmerz“ handelt es sich: Retrosternal? Stechend? Drückend? Oder doch genau im Bereich einer bestimmten Rippe? Und was meint dieser doch etwas aufgeregte Patient genau mit einer Schmerzintensität von 10/10? Und mit seiner „ganz schlimmen Atemnot“, bei nicht veränderter Atemfrequenz, und dem zusätzlichen heftigsten „Schwindel“...? Je präziser Beschreibung und Gewichtung der Symptome eingegeben werden, um so hochwertiger wird die Antwort der AI sein.
- Es werden anamnestisch oft sehr viele Symptome berichtet, die priorisiert werden müssen. Wie das gelingen kann, muss im Alltag für jedes AI-System separat erlernt werden.
- Das Wort „Intelligenz“ ist irreführend. Aufgrund ihrer statistischen Natur und aufgrund ihrer fehlenden Wahrnehmung eigenen Nichtwissens liefern heutige LLM auch schlichtweg „erfundene“ Informationen, sogenannte Halluzinationen. Bei einem „Generative LLM“ wie zum Beispiel ChatGPT sind zwar unstrukturierte Eingaben möglich. Die Eingabe definiert aber die statistische Evaluation durch miteinander verzahnte Netzwerkstrukturen (neuronale Netzwerke). Je nach Eingabe (dazu gehören auch die IP-Adresse des Absenders und dessen bei einer Firma wie Google ausgiebig hinterlegten persönlichen Daten) werden völlig unterschiedliche Outputs generiert. Medizinisches Personal muss daher detaillierte Patienteninformationen und Symptome als „Clues“ genau eingeben.
- Es ist absolut notwendig, ein LLM wie ChatGPT aufzufordern, **explizit Halluzinationen zu vermeiden** und **explizit** auf das Internet (Google, Bing, Yahoo), die National Library of Medicine der USA (pubmed.gov) und im Bedarfsfall auf andere, dedizierte Datenbanken (z. B. orpha.net und findzebra.com für seltene Erkrankungen) zuzugreifen. Nur so ist die Aktualität der Information jenseits von erfundenen Daten möglich. Aber selbst dann entstehen oft kritisch-falsche Antworten, die man nur mit Expertise ausgleichen kann. Die Outputs kann man verbessern, indem man zum Beispiel dedizierte GPTs (Generative Pretrained Transformers) im Rahmen von ChatGPT 4 verwendet, wie „Consensus“, „Medical Diagnosis Analysis“ usw.
- Die Spracheingabe, also die Semantik, erfordert im Falle von LLM-Tools wie ChatGPT die explizite Aufforderung zur Differenzialdiagnose und spezifische, klare und kontextbezogene Fragen, um sicherzustellen, dass die Antworten relevant und nützlich sind.
- Ironie oder Zwischen-den-Zeilen-Töne werden derzeit nicht als solche erkannt und sollten vermieden werden.
- Suggestive Fragen müssen vermieden werden, da LLM fast immer der Suggestion folgen.

Um die vollständige Anamnese und um wirklich alle Informationen zu erhalten, ist eine vertrauensvolle, menschliche und empathische Beziehung zwischen Arzt/Ärztin und Patient/Patientin nötig, die sich nicht in einer nur schriftlichen online-Kommunikation abbilden kann. Daher kann AI bei der Diagnosefindung allenfalls unterstützend, nicht aber substituierend zum Einsatz kommen.

Der Einsatz einer patientenbezogenen AI wirft wichtige nutzen-/schadenbezogene, haftpflichtrelevante, qualitätsbezogene, regulatorische, rechtliche, ethische und datenschutzrechtliche Fragen auf. Solche Systeme müssen reproduzierbare Standards und eine externe Kontrollierbarkeit sowie Schutzsysteme gegen Hacker einhalten. Und last but not least müssen Werkzeuge mit AI durch Programmelemente (Perzeptronen) so gesichert sein, dass diese nicht über statistische Lernmethoden zu Voroder Fehltritten führen oder gar über schlecht konzipierte Lernelemente zu rassistischen Elementen.

Der Einsatz von AI-Systemen in der Klinik hat aber ein großes Potenzial, vor allem, wenn AI in eine akute Situation hinein eine kontextsensitive Entscheidungshilfe liefert. Das klinisch-diagnostische Denken der Ärztinnen und Ärzte kann so in der Situation verbessert und erweitert, also auch geschult werden. Insbesondere könnten so seltene Diagnosen oder ungewöhnliche Präsentationen schneller erkannt werden, alle häufigen Differenzialdiagnosen vollständiger in Betracht gezogen werden und eigentlich offensichtliche Diagnosen nicht übersehen werden. AI kann dem Arzt-Patient-Gespräch zuhören, dieses codieren und kompakt zusammenfassen – und so wertvolle Zeit sparen. Weiterhin könnte ein in der elektronischen Patientendoku-

mentation integriertes AI-System alle relevanten medizinischen Informationen rasch finden, zusammenfassen und verfügbar machen: mit einem riesigen Zeitgewinn und der Reduktion sinnentleerer administrativer Arbeiten [23].

KERNAUSSAGEN

- Das klinisch-diagnostische Denken ist eine der Kernkompetenzen aller klinisch tätigen Ärztinnen und Ärzte.
- Ein besseres Verständnis des klinisch-diagnostischen Denkens hilft, Diagnosefehler zu vermeiden und AI-Systeme adäquat anzuwenden.
- Die AI wird das menschliche klinisch-diagnostische Denken erweitern, verbessern und beschleunigen, aber sehr wahrscheinlich nicht ersetzen.
- Die AI hat das Potenzial, medizinische diagnostische Fehler drastisch zu reduzieren und mehr Zeit für Ärztinnen und Ärzte verfügbar zu machen, die diese für die Betreuung ihrer Patienten brauchen.
- Die ärztliche Tätigkeit ist die Kunst, eine einfühlsame menschliche Beziehung zum Patienten aufzubauen und das klinische Denken mit den 3 Säulen der evidenzbasierten Medizin zu verbinden (wissenschaftliche Erkenntnisse, klinische Kompetenz und Erfahrung, Werte und Erwartungen der Patienten).

Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

Autorinnen/Autoren



Prof. Dr. med. Stefano Bassetti

Studium der Humanmedizin, Universität Zürich. Facharzt Innere Medizin und Infektiologie. 1998–2000 Research Fellowship, Infectious Diseases, Wake Forest University, Winston-Salem, NC, USA. 2006–2014 Leiter des Departement Innere Medizin, Kantonsspital Olten. Seit 1/2015 Leiter

der Klinik für Innere Medizin am Universitätsspital Basel und ordentlicher Professor für Innere Medizin an der Universität Basel. Schwerpunkte: ärztliche Aus-, Weiter- und Fortbildung, Optimierung des klinischen Denkens im klinischen Alltag.



Prof. Dr. Martin Hirsch

Studium der Humanbiologie in Marburg, Promotion in Neurowissenschaften. Gründung mehrerer Unternehmen, u. a. 2010 Ada Health, ein Gesundheits- und Technologieunternehmen. 2020 Berufung auf den neu eingerichteten Lehrstuhl für „Künstliche Intelligenz in der Medizin“

an der medizinischen Fakultät der Philipps-Universität Marburg. Martin Hirsch ist der Enkel des Nobelpreisträgers Werner Heisenberg. Spezielle Interessengebiete: kognitive Neurowissenschaft, Wissensdarstellung durch semantische Modelle und Technologien zur Unterstützung der menschlichen Entscheidungsfindung.



Prof. Dr. med. Edouard Battegay, FACP

Medizinstudium in Basel. Leiter der Medizinischen Poliklinik am Universitätsspital Basel und dann von 2007–2020 der Klinik und Poliklinik für Innere Medizin am Universitätsspital Zürich. Gründer und Präsident der Zurich Academy of Internal Medicine. 2021 Gründung des International

Center for Multimorbidity and Complexity (ICMC, <https://multimorbidity.org>). Autor vieler wissenschaftlicher Artikel sowie Herausgeber verschiedener Bücher, u. a. des Standardwerks „Differenzialdiagnose Innerer Krankheiten – vom Symptom zur Diagnose“. Berufliche Hauptinteressen: Differenzialdiagnose, Hypertonie Multimorbidität, Artificial Intelligence in der Medizin.

Korrespondenzadresse

Prof. Dr. med. Stefano Bassetti

Klinik für Innere Medizin, Universitätsspital Basel
Petersgraben 4
4031 Basel
Schweiz
stefano.bassetti@usb.ch

Zitierweise für diesen Artikel

Dtsch Med Wochenschrift 2024; 2024: DOI 10.1055/a-2201-5412

Es handelt sich um eine angepasste und verkürzte Version des Kapitels: Bassetti S, Hirsch M, Battegay E. Grundlagen des differenzialdiagnostischen Denkens. In: Battegay E, Hrsg. Differenzialdiagnose Innerer Krankheiten. Vom Symptom zur Diagnose. 22. Aufl. Stuttgart: Thieme; 2025

Literatur

- [1] Trowbridge RL, Rencic JJ, Durning SJ. Teaching Clinical Reasoning. Philadelphia: American College of Physicians; 2015
- [2] Hausmann D, Zulian C, Battegay E et al. Tracing the decision-making process of physicians with a Decision Process Matrix. BMC Med Inform Decis Mak 2016; 16: 133
- [3] Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016; 353: i2139. doi:10.1136/bmj.i2139
- [4] Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013; 22 (Suppl. 2): ii21–ii27. doi:10.1136/bmjqs-2012-001615
- [5] Newman-Toker DE, Nassery N, Schaffer AC et al. Burden of serious harms from diagnostic error in the USA. BMJ Qual Saf 2024; 33: 109–120

- [6] Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005; 165: 1493–1499. doi:10.1001/archinte.165.13.1493
- [7] Kahnemann D. *Thinking, fast and slow*. New York: Farrar, Strauss and Giroux; 2011
- [8] Croskerry P. Bias: a normal operating characteristic of the diagnosing brain. *Diagnosis (Berl)* 2014; 1: 23–27. doi:10.1515/dx-2013-0028
- [9] Paley L, Zornitzki T, Cohen J et al. Utility of clinical examination in the diagnosis of emergency department patients admitted to the department of medicine of an academic hospital. *Arch Intern Med* 2011; 171: 1394–1396. doi:10.1001/archinternmed.2011.340
- [10] Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 2006; 355: 2217–2225. doi:10.1056/NEJMr054782
- [11] Sackett DL, Straus SE, Richardson WS et al. *Evidence-Based Medicine. How to Practice and Teach EBM*, 2nd ed. London: Churchill Livingstone; 2000
- [12] Sackett DL, Rosenberg WM, Gray JA et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71–72
- [13] Buck D, Baylis A, Dougall D et al. A vision for population health: towards a healthier future. 2018. Zugriff am 09.08.2024 unter <https://www.kingsfund.org.uk/insight-and-analysis/reports/vision-population-health>
- [14] Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485. doi:10.1136/bmj.311.7003.485
- [15] Greco FA. What is the art of medicine? *Am J Med* 1985; 79: 279. doi:10.1016/0002-9343(85)90303-1
- [16] Peabody FW. The care of the patient. *JAMA* 1927; 88: 877–882. doi:10.2196/31042
- [17] Lang K, Josefsson V, Larsson AM et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol* 2023; 24: 936–944
- [18] Lou S, Du F, Song W et al. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials. *EclinicalMedicine* 2023; 66: 102341
- [19] Hake J, Crowley M, Coy A et al. Quality, Accuracy, and Bias in ChatGPT-Based Summarization of Medical Abstracts. *Ann Fam Med* 2024; 22: 113–120. doi:10.1370/afm.3075
- [20] McDuff D, Schaekermann M, Tu T et al. Towards Accurate Differential Diagnosis with Large Language Models. *arXiv* 2023. doi:10.48550/arXiv.2312.00164
- [21] Cabral S, Restrepo D, Kanjee Z et al. Clinical Reasoning of a Generative Artificial Intelligence Model Compared With Physicians. *JAMA Intern Med* 2024; 184: 581–583
- [22] Tu T, Palepu A, Schaekermann M et al. Towards Conversational Diagnostic AI. *arXiv* 2024. doi:10.48550/arXiv.2401.05654
- [23] Kulkarni PA, Singh H. Artificial Intelligence in Clinical Diagnosis: Opportunities, Challenges, and Hype. *JAMA* 2023; 330: 317–318. doi:10.1001/jama.2023.11440

Bibliografie

Dtsch Med Wochenschr 2024; 149: 1401–1410

DOI 10.1055/a-2201-5412

ISSN 0012-0472

© 2024. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany