

Aus der Medizinischen Klinik und Poliklinik IV  
Klinik der Ludwig-Maximilians-Universität München



***Einfluss von Bayesian Reasoning auf die Diagnoserichtigkeit und  
-schnelligkeit bei Medizinstudierenden***

Dissertation

zum Erwerb des Doktorgrades der Medizin  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität München

vorgelegt von

Alexandra Katharina Kunzelmann

aus

München

Jahr

2023

Aus der Medizinischen Klinik und Poliklinik IV  
Klinik der Ludwig-Maximilians-Universität München



***Einfluss von Bayesian Reasoning auf die Diagnoserichtigkeit und  
-schnelligkeit bei Medizinstudierenden***

Dissertation

zum Erwerb des Doktorgrades der Medizin  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität München

vorgelegt von

Alexandra Katharina Kunzelmann

aus

München

Jahr

2023

Mit Genehmigung der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

Erster Gutachter: Prof. Dr. med. Ralf Schmidmaier

Zweiter Gutachter: Prof. Dr. Martin Fischer

Dritter Gutachter: Priv. Doz. Dr. Michael Lauseker

Weitere Gutachter: \_\_\_\_\_  
\_\_\_\_\_

Mitbetreuung durch den

promovierten Mitarbeiter: Dr. med. Leah Braun

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 12.09.2023

Diese Arbeit wurde publiziert in:

Kunzelmann AK, Binder K, Fischer MR, Reincke M, Braun LT, Schmidmaier R. Improving Diagnostic Efficiency with Frequency Double-Trees and Frequency Nets in Bayesian Reasoning. *MDM policy & practice*. 2022;7(1):23814683221086623.

Impact Factor 2022: 1,54

Teile dieser Arbeit wurden beim gemeinsamen Jahreskongress der Deutschen Gesellschaft für Endokrinologie (DGE) und des Dachverbandes Osteologie (DVO) 2022 vom 18.09.2022 bis 20.09.2022 in Baden-Baden vorgestellt.

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis .....</b>	<b>5</b>
<b>Zusammenfassung: .....</b>	<b>7</b>
<b>Abstract (English):.....</b>	<b>8</b>
<b>Abbildungsverzeichnis.....</b>	<b>9</b>
<b>Tabellenverzeichnis .....</b>	<b>9</b>
<b>1. Einleitung .....</b>	<b>10</b>
1.1 Bayesian Reasoning in der medizinischen Entscheidungsfindung.....	10
1.1.1 Klassische Statistik vs. Bayes Statistik.....	10
1.1.2 Bayesianische Statistik.....	11
1.1.3 Bayesianische Statistik in der Medizin.....	11
1.2 Effizienz bei bayesianischen Aufgaben.....	14
1.3 Verringerung von Fehlern beim bayesianischen Schlussfolgern: Informationsformat und Visualisierung.....	15
1.3.1 Informationsformat.....	15
1.3.2 Visualisierung.....	17
1.3.3 Doppelbaumdiagramme und Netzdiagramme .....	18
1.3.4 Ziel der Studie .....	21
<b>2. Material und Methoden .....</b>	<b>22</b>
2.1 Teilnehmer .....	22
2.2 Studiendesign .....	22
2.3 Medizinischer Kontext.....	24
2.4 Kodierung.....	26
<b>3. Ergebnisse .....</b>	<b>27</b>
3.1 Leistung der Teilnehmer (Genauigkeit) .....	27
3.2 Zeit für die Lösung der Aufgabe und Effizienz.....	31
<b>4. Diskussion .....</b>	<b>34</b>
4.1 Häufigkeits-Doppelbäume und Häufigkeitsnetze erhöhen die Effizienz im Vergleich zu den entsprechenden Wahrscheinlichkeitsvisualisierungen.....	34
4.2 Auch Wahrscheinlichkeits-Doppelbäume und Wahrscheinlichkeitsnetz-Diagramme führen bei Medizinstudenten zu hohen Erfolgsquoten .....	34
4.3 Vergleich von Doppelbäumen und Netzdiagrammen mit einfachen Baumdiagrammen.....	35
4.4 Daten sollten bereits mit Visualisierungen veröffentlicht werden - Algorithmus zur Erstellung der Darstellungen .....	35
4.5 Einflüsse auf Lösung bayesianischer Problemstellungen.....	36
4.6 Psychologische Aspekte von Diagrammen und Persönlichkeitseigenschaften / Optimierung der Präsentationsform.....	39
4.7 Vorteile für die medizinische Ausbildung durch Häufigkeits-Doppelbäume und Häufigkeitsnetze.....	40
<b>5. Limitationen und Ausblick .....</b>	<b>42</b>

<b>Literaturverzeichnis .....</b>	<b>45</b>
<b>Anhang: .....</b>	<b>49</b>
<b>Danksagung .....</b>	<b>56</b>
<b>Affidavit.....</b>	<b>57</b>

## **Zusammenfassung:**

Ärzte haben oft Probleme mit Situationen, in denen sie bayesianische Schlüsse ziehen müssen. Die Darstellung statistischer Informationen als natürliche Häufigkeiten, anstelle von Wahrscheinlichkeiten, und deren Visualisierung führt in verschiedenen Anwendungsbereichen zu einer höheren Genauigkeit bei der Lösung dieser Aufgaben. Der Einfluss von Häufigkeitsdoppelbäumen und -netzen wurde bisher für medizinische Fragestellungen noch nicht untersucht. Diese Studie untersuchte den Einfluss des Informationsformats und dessen Visualisierung auf die *Genauigkeit* und *Geschwindigkeit* von bayesianischen Entscheidungen.

Als Faktoren wurden die Visualisierung (Doppelbaum, Netzdiagramm), das Informationsformat (Häufigkeiten, Wahrscheinlichkeiten) und der Inhalt der Fragestellung (Schilddrüsenkrebs, primärer Hyperaldosteronismus, das Cushing-Syndrom und die familiäre hypokalziurische Hyperkalzämie) untersucht, was zu einem drei-faktoriellen  $2 \times 2 \times 4$ -Design führte. Insgesamt haben 142 Medizinstudierende verschiedener medizinischer Fakultäten in Deutschland Posterior-Wahrscheinlichkeiten in den genannten vier verschiedenen medizinischen bayesianischen Schlussfolgerungsaufgaben vorhergesagt. Neben der Zahl der richtigen Antworten wurde zusätzlich die *diagnostische Effizienz* für die verschiedenen Versionen als Zielkriterium verwendet. Die diagnostische Effizienz ist der Median der Antwortzeit geteilt durch den Prozentsatz der richtigen Schlussfolgerungen.

Häufigkeitsvisualisierungen führten zu einer signifikant höheren *Genauigkeit* und *schnelleren Beurteilungen* als Wahrscheinlichkeitsvisualisierungen. Die Teilnehmer lösten 80 % der Aufgaben im *Häufigkeits-Doppelbaum* und im *Häufigkeits-Netzdiagramm* richtig. Visualisierungen mit Wahrscheinlichkeiten führten ebenfalls zu relativ hohen Leistungsdaten: 73 % im *Wahrscheinlichkeits-Doppelbaum* und 70 % im *Wahrscheinlichkeits-Netzdiagramm*. Die mediane Zeit für eine korrekte Schlussfolgerung war am schnellsten beim *Häufigkeits-Doppelbaum* (2:08 min), gefolgt vom *Häufigkeits-Netz-Diagramm* und dem *Wahrscheinlichkeits-Doppelbaum* (beide 2:26 min) und dem *Wahrscheinlichkeits-Netz-Diagramm* (2:33 min). Die Art der Visualisierung führte zu keinem signifikanten Unterschied.

Häufigkeits-Doppelbäume und Häufigkeitsnetzdiagramme bewirken bei der Beantwortung von Bayes-Aufgaben häufiger korrekte und gleichzeitig auch schnellere Antworten als die entsprechenden Wahrscheinlichkeitsdarstellungen.

## **Abstract (English):**

Physicians often have problems with situations in which they have to draw Bayesian conclusions. The representation of statistical information as natural frequencies, instead of probabilities, and their visualization leads to higher accuracy in solving these tasks in various application areas. The influence of frequency double-trees and net-diagrams has not yet been investigated for medical questions. This study investigated the influence of the information format and its visualization on the *accuracy* and *speed* of Bayesian decisions.

Factors examined were visualization (double-tree, net-diagram), information format (frequencies, probabilities) and context (thyroid cancer, primary aldosteronism, Cushing's syndrome and familial hypocalciuric hypercalcemia), resulting in a three-factorial  $2 \times 2 \times 4$  design. In total, 142 medical students from different medical schools in Germany predicted posterior probabilities in the aforementioned four different Bayesian medical reasoning tasks. In addition to the number of correct answers, the *diagnostic efficiency* for the different versions was also used as a target criterion. The diagnostic efficiency is the median response time divided by the percentage of correct conclusions.

Frequency visualizations resulted in significantly higher *accuracy* and *faster judgements* than probability visualizations. Participants correctly solved 80 % of the tasks in the *frequency double-tree* and the *frequency net-diagram*. Visualizations with probabilities also resulted in relatively high performance rates: 73 % in the *probability double-tree* and 70 % in the *probability net-diagram*. The median time to reach a correct conclusion was fastest for the *frequency double-tree* (2:08 min), followed by the *frequency net-diagram* and the *probability double-tree* (both 2:26 min) and the *probability net-diagram* (2:33 min). The type of visualization did not lead to a significant difference.

Frequency double-trees and frequency net-diagrams produce correct answers more often and at the same time also faster answers than the corresponding probability plots when answering Bayes tasks.



## Abbildungsverzeichnis

<b>Abbildung 1</b> - Netzdiagramme- und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Schilddrüsenbeispiel. Die richtigen Lösungen sind hier bereits fettgedruckt gekennzeichnet. Grafik adaptiert nach Kunzelmann et.al. (2022). .....	20
<b>Abbildung 2</b> - Studiendesign: 2 (Informationsformat: Wahrscheinlichkeiten vs. Häufigkeiten) x 2 (Visualisierung: Doppelbaum vs. Netzdiagramm) x 4 (medizinische Kontexte: Nebenziel der Studie). Zu den 4 medizinischen Kontexten gehörten Schilddrüsenkrebs, primärer Hyperaldosteronismus, das Cushing-Syndrom und die familiäre hypokalziurische Hyperkalzämie. Die Reihenfolge der Aufgaben wurde für jeden Teilnehmer randomisiert. Grafik adaptiert nach Kunzelmann et.al. (2022) .....	23
<b>Abbildung 3</b> - (A) Leistung der Teilnehmer bei den Aufgaben zum bayesianischen Schlussfolgern (kontextübergreifend). (B) Mittlere Zeit für das Lösen einer bayesianischen Aufgabe richtig oder falsch zu lösen (kontextübergreifend). Grafik adaptiert nach Kunzelmann et.al. (2022).	28
<b>Abbildung 4</b> - (A) Leistung der Teilnehmer bei der ersten Aufgabe zum bayesianischen Schlussfolgern (kontextübergreifend). (B) Mittlere Zeit für das richtige oder falsche Lösen der ersten bayesianischen Schlussfolgerungsaufgabe (kontextübergreifend). Grafik adaptiert nach Kunzelmann et.al. (2022).....	30
<b>Abbildung 5</b> - Bildschirmfoto zur Recherche mit den Begriffen <i>bayesian reasoning medical</i> in der Datenbank PubMed. Das Histogramm zeigt, wie sich die Häufigkeit der zu dieser Begriffskombination gefundenen Artikel im Zeitraum von 1981 bis 2022 entwickelt hat. Allein in den letzten <i>fünf</i> Jahren wurden 44 % (367 von 833) der seit 1987 erschienen Artikel publiziert. Stand der Recherche: 5.9.2022. ....	37

## Tabellenverzeichnis

<b>Tabelle 1</b> - Fallbeschreibung des Schilddrüsenbeispiels. Grafik adaptiert nach Kunzelmann et.al. (2022). .....	25
<b>Tabelle 2</b> - Kontextübergreifende Gesamtergebnisse. Prozentsatz der richtigen Schlussfolgerungen, Zeit für die Diagnose mit Median, 1. (Q1) und 3. (Q3) Quartil, Punktzahl für die diagnostische Effizienz. Der Wert für die diagnostische Effizienz wurde berechnet, indem die mediane Zeit durch den Prozentsatz der richtigen Schlussfolgerungen geteilt wurde. N gibt die Anzahl der Teilnehmer an. Grafik adaptiert nach Kunzelmann et.al. (2022). .....	27
<b>Tabelle 3</b> - Ergebnisse des GLMM zur Vorhersage der Genauigkeit der Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).....	29
<b>Tabelle 4</b> - Ergebnisse des LMM zur Vorhersage der Geschwindigkeit der richtigen Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).....	32
<b>Tabelle 5</b> - Ergebnisse des LMM zur Vorhersage der Geschwindigkeit Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).....	33

In dieser Arbeit wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Weibliche und anderweitige Geschlechteridentitäten werden dabei ausdrücklich mitgemeint, soweit es für die Aussage erforderlich ist.

# 1. Einleitung

## 1.1 Bayesian Reasoning in der medizinischen Entscheidungsfindung

In der täglichen klinischen Praxis werden Ärzte häufig mit Bayes Situationen konfrontiert, in welchen sie bedingte Wahrscheinlichkeiten berechnen müssen. Hierzu ist wichtig, den Unterschied zwischen frequentistischer, klassischer Statistik und nichtfrequentistischer Bayes Statistik zu verstehen.<sup>1</sup>

### 1.1.1 Klassische Statistik vs. Bayes Statistik

In klassischer Statistik wird die Wahrscheinlichkeit, auch *objektive Wahrscheinlichkeit* genannt, als Häufigkeit definiert, mit der ein bestimmter Test als positiv gewertet werden kann, wenn man ihn häufig wiederholt. Wichtige Elemente hierzu sind der p-Wert, der Hypothesentest, das Signifikanzniveau und der alpha/beta-Fehler.<sup>1</sup>

Im Gegensatz hierzu wird bei Bayes Statistik die Wahrscheinlichkeit als *Grad der persönlichen Überzeugung bzw. Erwartung* interpretiert. Die *persönliche Überzeugung* wird dabei aus bisher gemachter *Erfahrung*, z.B. Informationen über Inzidenzen aus Fachliteratur oder Vorträgen bzw. auch plausiblen Abschätzungen, gewonnen. Im englischen Sprachgebrauch wird diese „Erwartung“ im Sinne von Bayes als *educated guess* bezeichnet. Statistisch spricht man hier von der a-priori-Wahrscheinlichkeit.<sup>2</sup>

Die a-posteriori-Wahrscheinlichkeit ist die aktualisierte a-priori Wahrscheinlichkeit unter Berücksichtigung neuer Informationen, z. B. diagnostischer Tests. Dieser Prozess, die Korrektur der a-posteriori Wahrscheinlichkeit aufgrund von neuen Erkenntnissen, wird als *bayesian updating* bezeichnet.<sup>2</sup>

Zu Bayes Situationen gehört somit die Beurteilung, wie sich eine neue Information, etwa ein medizinisches Testergebnis, auf die Wahrscheinlichkeit auswirkt, dass der Patient tatsächlich eine bestimmte Krankheit hat. Für den Patienten ist es dann wichtig zu wissen, was genau diese Testergebnisse bedeuten, d. h. wie wahrscheinlich es ist, dass ein positives Ergebnis eines Tests auf eine Krankheit hinweist.<sup>3</sup>

In der klassischen Statistik wird mittels Hypothesentests untersucht, ob ein Patient mit auffälligem Testergebnis tatsächlich an einem bösartigen Tumor erkranken wird. Diese Annahme wird als Nullhypothese formuliert. Auf Basis eines Signifikanzniveaus von 95 % ist beispielweise der entsprechende p Wert des Chi-Quadrat Tests wichtig. Eine Signifikanz von 95 % bedeutet: die Wahrscheinlichkeit, dass man sich irrt, wenn die Nullhypothese verworfen wird, liegt bei maximal 5 %.<sup>4</sup>

Die Schlussfolgerung ist in beiden statistischen Arten ähnlich. Was sich aber grundlegend unterscheidet, ist der Informationsgehalt: Die Bayes Statistik hat zusätzlich noch die a-posteriori Wahrscheinlichkeit berechnet. Damit lässt sich die wohl begründete Vermutung äußern, wie hoch die tatsächliche Wahrscheinlichkeit für die Krankheit bei diesem Testergebnis ist. Dies können klassische Statistik und p Wert nicht.<sup>1</sup>

### 1.1.2 Bayesianische Statistik

Der Satz von Bayes, auch Bayes Theorem, wurde 1763 vom Mathematiker Thomas Bayes entwickelt. Mit dem Satz von Bayes wird die Wahrscheinlichkeit, dass A eintritt unter der Bedingung, dass B bereits eingetreten ist, gemäß der Formel (1) berechnet.<sup>5</sup>

Formel (1) 
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- Das mathematische Zeichen | bedeutet: *unter der Bedingung von*.
- $P(A)$  ist die a-priori Wahrscheinlichkeit des Ereignisses A.
- $P(B)$  ist die a-priori-Wahrscheinlichkeit des Ereignisses B, also die *marginale Wahrscheinlichkeit*, zu verstehen als eine Art Skalierungsfaktor.<sup>6</sup>
- $P(A|B)$  ist die *bedingte* Wahrscheinlichkeit des Ereignisses A unter der Bedingung, dass B eingetreten ist, also die *a-posteriori Wahrscheinlichkeit*.
- $P(B|A)$  ist die *Plausibilitätsfunktion*, also die *bedingte* Wahrscheinlichkeit, inwiefern B zutrifft unter der Bedingung, dass A bereits eingetreten ist.<sup>7</sup>

Die Bayes Statistik nutzt diese Formel, um Schlussfolgerungen über die a-posteriori Wahrscheinlichkeit zu ziehen, indem die a-priori Wahrscheinlichkeit durch Multiplikation mit der Plausibilitätsfunktion aktualisiert wird, dies bezeichnet man als *Bayesian Updating*.<sup>8</sup>

Die Bayes Statistik erlaubt es bereits mit sehr geringen Fallzahlen, gut begründete Schlüsse zu ziehen. Außerdem ist die Interpretation der a-posteriori Wahrscheinlichkeit sehr intuitiv und gut verständlich. Das Ergebnis der Berechnungen nach dem Satz von Bayes liefert eine mathematisch korrekte Schätzung für die Wahrscheinlichkeit, mit der bestimmte Ereignisse eintreten.<sup>1</sup>

Eine Schwäche der bayesianischen Statistik ist, dass die a-priori-Wahrscheinlichkeit nicht immer eindeutig bestimmbar ist.<sup>1</sup>

### 1.1.3 Bayesianische Statistik in der Medizin

Gerade im Bereich der Medizin, wenn nur eine begrenzte Anzahl an Testobjekten zur Verfügung steht, liefert die Bayes Statistik zuverlässige Ergebnisse und eine höhere Aussagekraft als der p Wert der klassischen Statistik. Das Bayes Theorem sagt dem Mediziner, wie stark neue Informationen die Wahrscheinlichkeit für das Vorliegen einer bestimmten Erkrankung verändern sollten.

Somit können Ärzte eine ursprüngliche Diagnose mit jeder neuen Information verändern und irgendwann zu dem Schluss kommen, dass die Wahrscheinlichkeit gering genug ist, um die Diagnose zu verwerfen.<sup>9</sup>

Das bisher bekannteste Beispiel von Bayesian Reasoning im klinischen Alltag ist die Brustkrebsvorsorge, welches Eddy (1982) präsentierte:<sup>10</sup>

Es werden folgende mathematischen Zeichen verwendet:  $\neg$  gelesen als: logisches *Nicht*, beispielsweise  $P(B)$  (Brustkrebs) und  $P(\neg B)$  (kein Brustkrebs).

M+ bedeutet im Folgenden, dass ein positives Mammographie-Ergebnis und somit ein auffälliger Befund, vorliegt. Analog heißt M- im Folgenden, dass ein negatives Mammographie-Ergebnis vorliegt, also kein auffälliger Befund.

*Die Wahrscheinlichkeit, an Brustkrebs zu erkranken, beträgt 1 % für eine Frau einer bestimmten Altersgruppe, die an einer Routineuntersuchung teilnimmt (a-priori-Wahrscheinlichkeit, auch Prävalenz  $P(B)$  genannt). Wenn eine Frau, die an einem Routine-Screening teilnimmt, an Brustkrebs erkrankt, liegt die Wahrscheinlichkeit bei 80 %, dass sie eine positive Mammographie erhält (Sensitivität  $P(M+|B)$ , also die Wahrscheinlichkeit eines positiven Mammographie-Ergebnisses, unter der Bedingung, dass Brustkrebs vorliegt). Wenn eine Frau, die an einem Routine-Screening teilnimmt, keinen Brustkrebs hat, liegt die Wahrscheinlichkeit bei 9,6 %, dass sie ein falsch-positives Mammographie-Ergebnis erhält (Falsch-Positiv-Rate  $P(M+|\neg B)$ , also die Wahrscheinlichkeit eines positiven Mammographie-Befundes, wenn kein Brustkrebs vorliegt).<sup>3</sup>*

Die Frage ist hierbei, wie hoch die Wahrscheinlichkeit ist, dass eine Frau, die an einem Routine-Screening teilnimmt und bei der ein positiver Mammographiebefund vorliegt, tatsächlich Brustkrebs hat. In früheren Studien gingen die meisten Ärzte davon aus, dass diese Wahrscheinlichkeit zwischen 70 und 80 % liegt, was weit vom korrekten positiven prädiktiven Wert entfernt ist.<sup>10,11</sup> Aufgrund dieser Fehleinschätzung wird das Mammographie-Screening in vielen Ländern als notwendig für jede Frau einer bestimmten Altersgruppe propagiert, obwohl es teuer ist<sup>12</sup> und sein medizinischer Nutzen ernsthaft in Frage gestellt werden kann.<sup>13</sup>

In der aktuellen Studie wurden die Fallbeispiele aus dem Fachbereich der Endokrinologie ausgewählt. In diesem Bereich der Inneren Medizin gibt es Erkrankungen, die in der Bevölkerung sehr häufig auftreten, wie zum Beispiel Schilddrüsenerkrankungen, und solche, die sehr seltenen vorliegen, wie das Cushing-Syndrom. Durch Beispiele mit hoher und geringer Häufigkeit kann geklärt werden, ob die *Bekanntheit* des Inhalts der Fragen Einfluss auf das Ergebnis hat.

Veranschaulichung einer bayesianischen Situation anhand eines Schilddrüsenkrebs-Problems:<sup>14</sup>

Folgende Abkürzungen werden verwendet:

C: Schilddrüsenkrebs

$\neg C$ : kein Schilddrüsenkrebs

T+: Positives Testergebnis, in diesem Fall eine auffällige Schilddrüsen-Sonographie

T-: Negatives Testergebnis, also keine auffällige Schilddrüsen-Sonographie.

*Schilddrüsenkrebs-Problem-Wahrscheinlichkeitsformat:*

*Die Wahrscheinlichkeit eines Schilddrüsenkrebses bei einem Patienten in einer endokrinologischen Ambulanz, der einen Schilddrüsenknoten hat, beträgt 5 % (Prävalenz der Krankheit  $P(C)$ ). Wenn ein Patient Schilddrüsenkrebs hat, beträgt die Wahrscheinlichkeit 40 %, dass die Sonographie einen Schilddrüsenknoten mit Anzeichen von Bösartigkeit aufgrund des Vorhandenseins von Mikroverkalkungen zeigt (Sensitivität  $P(T+|C)$ , also die Wahrscheinlichkeit eines positiven Tests, unter der Bedingung, dass Schilddrüsenkrebs vorliegt). Wenn ein Patient keinen Schilddrüsenkrebs hat, beträgt die Wahrscheinlichkeit 12 %, dass die Sonographie einen Schilddrüsenknoten mit Hinweis auf Bösartigkeit aufgrund des Vorhandenseins von Mikroverkalkungen zeigt (falsch-positive Rate  $P(T+|\neg C)$ , also die Wahrscheinlichkeit eines positiven Tests, unter der Bedingung, dass kein Schilddrüsenkrebs vorliegt).*

*Wie groß ist die Wahrscheinlichkeit, dass diese Person mit einem Schilddrüsenknoten mit Mikroverkalkungen tatsächlich Schilddrüsenkrebs hat?*

Hier wird nach dem positiven Vorhersagewert  $P(C|T+)$  gefragt.

Die richtige Lösung ist 14,9 % und kann mit der Formel von Bayes berechnet werden:

$$\begin{aligned} P(C|T+) &= \frac{P(T+|C)P(C)}{P(T+|C)P(C) + P(T+|\neg C)P(\neg C)} \\ &= \frac{40\% \cdot 5\%}{40\% \cdot 5\% + 12\% \cdot 95\%} \approx 14,9 \end{aligned}$$

Eine einfache Herleitung der Formel von Bayes kann beispielsweise in <https://www.sofatutor.com/mathematik/videos/satz-von-bayes> gefunden werden.<sup>15</sup>

Im Hinblick auf die medizinische Entscheidungsfindung ist das Verständnis medizinischer Testergebnisse im Sinne von Befunden sowohl für Medizinstudierende, Ärzte, als auch für Patienten von entscheidender Bedeutung. Diskussionen über Patienten setzen voraus, dass die Ärzte die medizinische Statistik verstehen und Daten zum Nutzen und vermeiden von Schaden von Diagnose-, und Behandlungsoptionen richtig interpretieren können.<sup>16</sup> Durch das Beherrschen dieser

bayesianischen Situationen können mögliche Probleme durch Überdiagnostik, Überbehandlung, aber auch das Übersehen von Krankheiten verhindert werden.<sup>17</sup> Als Überdiagnostik wird die Erkennung von Krebsarten, die nicht zum Tod oder zu Symptomen führen, verstanden.<sup>18</sup> Überbehandlung bedeutet die Durchführung von Maßnahmen zur Behandlung von nicht lebensbedrohlichen Zuständen oder Zuständen, die keine Symptome verursacht hätten, wie beispielsweise eine Mastektomie bei überdiagnostiziertem Brustkrebs.<sup>18</sup> Viele Behandlungen werden durchgeführt, da Mediziner diagnostische Früherkennungsverfahren vermehrt einsetzen, um beispielsweise Tumorerkrankungen frühzeitiger zu erkennen, da sie von unverifizierten Annahmen zur statistischen Wahrscheinlichkeiten ausgehen.<sup>18</sup> Hierzu wurde beispielsweise gezeigt, dass 95 von 100 Medizinern die a-posteriori Wahrscheinlichkeit  $P(\text{Brustkrebs} \mid \text{auffälliges Mammographie Ergebnis})$  auf 70 - 80 % schätzten, statt der korrekten 7,8 %. Aufgrund dieser Fehleinschätzungen werden viele Patientinnen übertherapiert.<sup>10</sup> Im Rahmen des Anstiegs der Inzidenz von Brustkrebs wurden vermehrt Mammographie-Screenings durchgeführt. Jedoch konnte gezeigt werden, dass einer von drei Brustkrebsfällen, durch Vorsorgeuntersuchungen im Rahmen eines Mammographie-Screenings entdeckt wird, überdiagnostiziert ist.<sup>18</sup>

Wenn medizinisches Personal jedoch wüsste, wie wahrscheinlich es ist, aufgrund eines auffälligen Untersuchungsbefundes eine Tumorerkrankung zu diagnostizieren und die Berechnungen der Wahrscheinlichkeiten beherrschen würde, könnten Patienten viele zusätzliche Maßnahmen im Rahmen von Überdiagnostik erspart bleiben und in diesem Rahmen eine Übertherapie vermieden werden.<sup>13</sup> Zudem könnten hierdurch erhebliche Kosten durch Überdiagnostik im Gesundheitssystem vermieden werden.<sup>19</sup> Außerdem könnte durch das präzise Einsetzen von Tests und das Beherrschen der richtigen Interpretation bayesianischer Kontexte bei betroffenen Patienten verhindert werden, Krankheiten zu übersehen.<sup>17</sup>

Aufgrund des Zeitdrucks im klinischen Alltag ist es wichtig, bei der klinischen Entscheidungsfindung sowohl die *Genauigkeit* als auch die *Geschwindigkeit* zu berücksichtigen, um die richtige Diagnose zu stellen.<sup>20</sup>

## 1.2 Effizienz bei bayesianischen Aufgaben

Die Arbeitsbedingungen eines Arztes in der täglichen Praxis entsprechen nicht dem Idealszenario, jedem Patienten die individuell benötigte Zeit widmen zu können, weshalb Ärzte wichtige Entscheidungen unter Zeitdruck treffen müssen.<sup>21</sup> Daher ist es wichtig, dass Medizinstudenten lernen, Wahrscheinlichkeiten in bayesianischen Schlussfolgerungen so schnell wie möglich zu interpretieren und dass angehende Mediziner und Ärzte über eine gute diagnostische Effizienz verfügen. Im wissenschaftlichen Kontext von Studien wird die diagnostische Effizienz wie folgt definiert: Die diagnostische Effizienz ist die für die Diagnosen benötigte Zeit geteilt durch die Anzahl der korrekt gelösten klinischen Fälle.<sup>20</sup> Die Diagnosegeschwindigkeit scheint höher zu sein,

wenn die statistischen Kennzahlen als natürliche Häufigkeiten in Baumdiagrammen angegeben werden.<sup>3,22</sup> Je mehr Denkschritte erforderlich sind, desto länger dauert es, bis eine Lösung für bayesianische Fragen gefunden wird.<sup>23</sup> Die Diagnosegeschwindigkeit wurde jedoch noch nicht für Doppelbäume und Netzdiagramme untersucht, obwohl dies interessant wäre, da die richtige Lösung bereits im Diagramm dargestellt wird.

### **1.3 Verringerung von Fehlern beim bayesianischen Schlussfolgern: Informationsformat und Visualisierung**

#### **1.3.1 Informationsformat**

In Fragen, bei denen bayesianisches Denken angewandt wird, können die Informationen als Wahrscheinlichkeiten oder als natürliche Häufigkeiten angegeben werden. Im mathematischen Formalismus der Wahrscheinlichkeitsrechnung werden Wahrscheinlichkeiten anstelle von Häufigkeiten verwendet. Wahrscheinlichkeiten in Form von Dezimalbrüchen oder Prozentsätzen sind jedoch nicht intuitiv verständlich.<sup>24</sup> Die Darstellung natürlicher Häufigkeiten in bayesianischen Aufgaben führt zu wesentlich korrekteren Ergebnissen.<sup>17,25,26</sup>

In vielen Studien wurde gezeigt, dass in solchen Situationen die Anzahl korrekter Antworten in der Entscheidungsfindung je nach Informationsformat der Daten unterschiedlich ist: Wahrscheinlichkeiten (z.B. 40 %) gegenüber natürlichen Häufigkeiten (z. B. 40 von 100 Patienten).<sup>11,24,27-29</sup>

Leider gelingt es den vielen Ärzten nicht so leicht, bayesianische Aufgaben, bei denen nur Wahrscheinlichkeiten angegeben werden, korrekt zu lösen, was schwerwiegende Folgen haben kann.<sup>11,24,27,28,30</sup>

Gerd Gigerenzer und Ulrich Hoffrage haben Ende der 1990er Jahre den Begriff *natürliche Häufigkeiten* eingeführt und durch ihre Studien gezeigt, dass *natürliche Häufigkeiten* das Verständnis für den positiven Vorhersagewert bei Ärzten und Medizinstudenten fördern.<sup>24</sup> Der Terminus *technicus natürliche Häufigkeit* steht dabei für die Umschreibung von Wahrscheinlichkeiten. Anstatt beispielsweise von einer Wahrscheinlichkeit von 15 % zu sprechen, sagt man „15 von 100 Patienten“, wenn die natürliche Häufigkeit zugrunde gelegt wird. Mathematisch sind beide Aussagen gleichwertig. Für die Interpretation scheint die natürliche Häufigkeit jedoch leichter verständlich zu sein.

Gigerenzer und Hoffrage gingen davon aus, dass das *natürliche* Format im Laufe der menschlichen Entwicklung Häufigkeiten als tatsächlich erlebte Ereignisse in einer Reihe von Ereignissen, anstatt Wahrscheinlichkeiten oder Prozentsätze waren und natürliche Häufigkeiten daher leichter für das menschliche Gehirn zu begreifen waren.<sup>24</sup> Menschen, Tiere und auch neuronale Netze

aktualisieren ihre Kenntnisse durch Beobachtung von Häufigkeiten und ihrer Erfahrung zu verschiedenen Ereignissen. Tiere beispielsweise beobachten die Frequenz verschiedener Gegebenheiten in ihrem Umfeld und passen sich daran an.<sup>24</sup> Natürliche Häufigkeiten sind eine gemeinsame Häufigkeit zweier Ereignisse, wie die Anzahl der Patienten mit Krankheit und einem positiven Testergebnis.<sup>12</sup> Die Darstellung durch *natürliche Häufigkeiten* ist eine Alternative zur kognitiv komplexeren Darstellung der gleichen Informationen in Form von bedingten Wahrscheinlichkeiten, wie Sensitivitäten und Spezifitäten.<sup>31</sup> Bedingte Wahrscheinlichkeiten neigen dazu, zu mehr Fehlern zu führen.<sup>12</sup>

Zur Verdeutlichung dieser Aussage wird das später für die Untersuchung verwendete Beispiel *Schilddrüsenkrebs* als Beispiel mit natürlichen Häufigkeiten erläutert:

*Schilddrüsenkrebs Problem-Häufigkeits-Format:*

*Insgesamt 5.000 von 100.000 Patienten in einer endokrinologischen Ambulanz, die einen Schilddrüsenknoten haben, haben Schilddrüsenkrebs  $P(C)$ . Bei 2.000 der 5.000 Patienten mit Schilddrüsenkrebs wurde in der Sonographie ein Schilddrüsenknoten mit Anzeichen für Bösartigkeit durch das Vorhandensein von Mikroverkalkungen festgestellt (Sensitivität  $P(T+|C)$ ). Von den 95.000 Patienten ohne Schilddrüsenkrebs wurden 11.400 fälschlicherweise einer Sonographie unterzogen, die aufgrund des Vorhandenseins von Mikroverkalkungen einen Schilddrüsenknoten mit Hinweis auf Bösartigkeit zeigte (falsch-positive Rate  $P(T+|\neg C)$ ).*

*Wie viele Patienten, die einen Schilddrüsenknoten mit Mikroverkalkungen haben, haben tatsächlich Schilddrüsenkrebs?*

Hervorzuheben ist, dass die Wahl der natürlichen Häufigkeiten didaktisch optimiert wurde. Statt eine Wahrscheinlichkeit von 5 % mit der natürlichen Häufigkeit „5 von 100“ Patienten auszudrücken, wurde beispielsweise die natürliche Häufigkeit „5.000 von 10.000“ gewählt, da somit im nächsten Satz sofort offensichtlich ist, dass sich die Aussage „2.000 von 5.000“ Patienten, auf die im ersten Satz genannten 5.000 Patienten bezieht. Die großen Zahlen wurden gewählt, um Dezimalbrüche in der weiteren Argumentationsfolge (11.400) zu vermeiden.

Auf diese Weise ist es auch deutlicher zu sehen, dass in der Summe (2.000 + 11.400 =) 13.400 Patienten eine auffällige Sonographie erhalten und dass nur 2.000 von diesen 13.400 tatsächlich Schilddrüsenkrebs haben (das sind ca. 14,9 %).<sup>11,24,27</sup> Die positive Wirkung natürlicher Häufigkeiten hat sich nicht nur im medizinischen, sondern auch im wirtschaftlichen und rechtlichen Kontext gezeigt. Natürliche Häufigkeiten helfen Medizinem bei wichtigen diagnostischen Entscheidungen, sowie den Patienten beim Verstehen der Diagnosen. Außerdem können sie dazu beitragen, dass Universitätsstudenten<sup>32</sup> und Schüler bei bayesianischen Problemstellungen,<sup>33</sup> sowie Jurastudenten bei der Bewertung rechtlicher Fallstellungen<sup>34</sup> und Führungskräfte bei wichti-



gen Managemententscheidungen zu richtigen Schlussfolgerungen gelangen.<sup>35-38</sup> Dieser Erleichterungseffekt durch die Darstellung der Informationen einer bayesianischen Situation unter Verwendung von natürlichen Häufigkeiten und Visualisierungsarten zur Darstellung der Beziehungen zwischen den Zahlen wird mit dem *ökologischen Rationalitätsmodell* oder „*Nested-Sets*“ erklärt.<sup>27,39,40</sup> Das bedeutet, dass die Informationen auf eine Weise dargestellt werden, die für den Menschen verständlich sind, weil sie nicht nur aus der natürlichen Umwelt, sondern auch aus der Evolutionsgeschichte bekannt sind.<sup>41,42</sup> Visuelle Darstellungen tragen dazu bei, kognitive Prozesse der Beobachtung von Häufigkeiten zu aktivieren, die durch die Ökologie im Laufe der Evolution geprägt wurden. Über viele Generationen hinweg haben Menschen die Häufigkeit von Ereignissen, Objekten und Orten analysiert. Ähnlich strukturierte Darstellungen können daher bei der Lösung von Problemen helfen.<sup>41</sup> In psychologischen Studien konnte außerdem gezeigt werden, dass Häufigkeiten automatisch registriert werden können, ohne bewusste Intention und ohne Interferenz mit anderen Denkprozessen.<sup>43</sup> Es wird deutlich, dass die Darstellung, die Zwischenschritte am einfachsten verständlich präsentiert, zu den schnellsten Antworten mit dem höchsten Anteil richtiger Antworten führt.

Die Wirkung der natürlichen Häufigkeiten ist beim bayesianischen Schlussfolgern recht konstant. Laut einer Metaanalyse liegt die Anzahl korrekter Lösungen bei den Versionen mit natürlichen Häufigkeiten etwa bei 24 %, bei den Wahrscheinlichkeitsversionen allerdings bei nur 4 %.<sup>27</sup> Überraschenderweise konvertieren einige Teilnehmer die Daten aus den Häufigkeitsversionen in Wahrscheinlichkeiten, weshalb einige der Häufigkeitsversionen nicht korrekt gelöst werden.<sup>44</sup> Dies konnte in vorangegangenen Studien gezeigt werden, in denen Rechenwege nachvollzogen werden konnten, da Teil der Aufgabe war, dass Teilnehmer alle Gedanken und Rechenwege notieren sollten. Auf diese Weise konnte der Fehler durch Umrechnen der gegebenen Häufigkeiten in Wahrscheinlichkeiten identifiziert werden. Teilnehmer rechneten vermutlich die Häufigkeiten in Wahrscheinlichkeiten um, da die Angaben in Wahrscheinlichkeitsformat in der Lehre in Deutschland vermehrt beigebracht wird.<sup>44</sup>

Trotz des Wissens um die positive Auswirkung von natürlichen Häufigkeiten auf die Leistung werden Medizinstudenten häufig bayesianische Inferenzaufgaben mit Wahrscheinlichkeiten vorgelegt und beigebracht.<sup>45</sup>

### 1.3.2 Visualisierung

Neben dem positiven Effekt der Erleichterung des Lösens bayesianischer Probleme durch natürliche Häufigkeiten, kann auch die Darstellung in Form bestimmter Visualisierungen dazu beitragen, die bayesianische Schlussfolgerung zu verbessern: (Doppel-)Baumdiagramme, 2×2-Tabellen, Icon-Arrays, Netzdiagramme, Euler-Diagramme, Roulette-Rad-Diagramme, Häufigkeitsraster, Einheitsquadrate und vieles mehr.<sup>17,25,33,35,37,46-57</sup> Übungen mit natürlichen Häufigkeiten und

Visualisierungen natürlicher Häufigkeiten oder erfahrungsbasierte Formate vorgeschlagen, um die Fähigkeiten von Medizinstudenten zum bayesianischen Schlussfolgern zu trainieren.<sup>32,58-60</sup>

Einige der genannten Darstellungsvarianten bieten die Möglichkeit, Informationen mit oder ohne Zahlen darzustellen. In den Visualisierungsformen ohne numerischen Wert, werden Größenverhältnisse oftmals graphisch dargestellt.<sup>33,56</sup> Einzelne Visualisierungen, wie beispielsweise (Doppel-)Baumdiagramme, können mit einfachen Mitteln, wie Stift und Papier, schnell gezeichnet werden, während andere Visualisierungen, wie beispielsweise Icon-Arrays oder Häufigkeitsraster, mehr Zeit für die Vorbereitung erfordern, da sie nicht einfach von Hand dargestellt werden können.<sup>25,57</sup> Außerdem konnte gezeigt werden, dass *Häufigkeits*-(Doppel-)Bäume zu genaueren Lösungen von bayesianischen Aufgaben führen, während *Wahrscheinlichkeits*-(Doppel-)Bäume weniger hilfreich sind.<sup>3,17</sup>

Bislang wurden Medizinstudenten bayesianische Aufgaben vor allem mit  $2 \times 2$ -Tabellen oder mit einfachen Baumdiagrammen präsentiert.<sup>25</sup> Doppelbäume und vor allem die relativ neue Art der Visualisierung als Netzdiagramme sind wenig untersucht worden. Binder et al. 2020 verglichen den Doppelbaum und das Netzdiagramm mit der allgemein verwendeten  $2 \times 2$ -Tabelle und der reinen Textversion. Es konnte in diese Studie gezeigt werden, dass  $2 \times 2$ -Tabellen, Doppelbäume und Netzdiagramme die korrekte Lösung von bayesianischen Schlussfolgerungsaufgaben unterstützen, wenn sie natürliche Häufigkeiten enthalten. Dahingegen waren die entsprechenden Wahrscheinlichkeitsdarstellungen weniger hilfreich.<sup>25</sup>

### 1.3.3 Doppelbaumdiagramme und Netzdiagramme

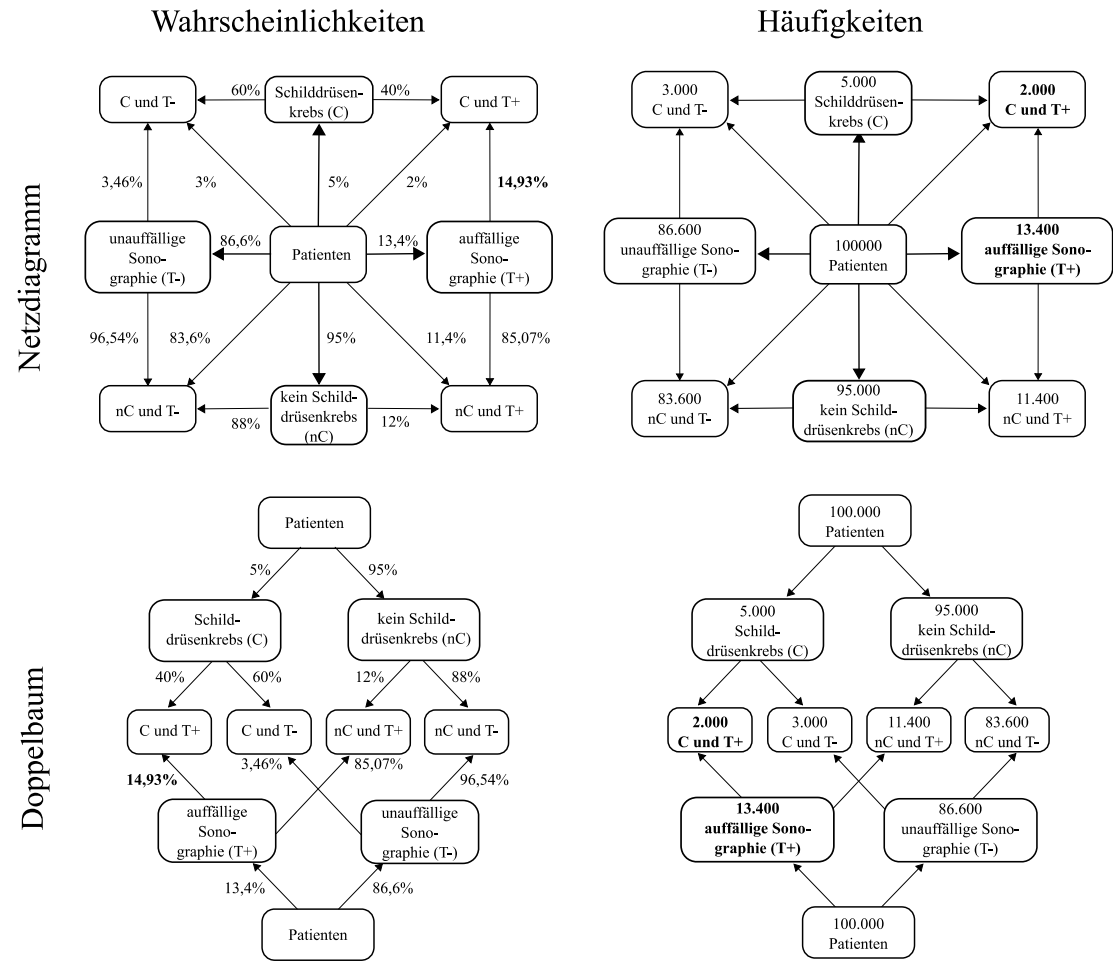
Eine weiterentwickelte Version der oben genannten Baumdiagramme sind Doppelbaumdiagramme und Netzdiagramme, die sich besonders für die Ausbildung von Medizinstudenten und Ärzten eignen, siehe Abbildung 1.<sup>3,25</sup> Doppelbaumdiagramme und Netzdiagramme enthalten bereits die richtige Lösung der Aufgabe. Im Vergleich zu einfachen Baumdiagrammen sind keine weiteren Berechnungen notwendig. Daher könnten Doppelbaumdiagramme und Netzdiagramme für die Verbesserung der Genauigkeit beim bayesianischen Schlussfolgern sogar vielversprechender sein.

Häufigkeitsdoppelbäume stellen ein wichtiges didaktisches Instrument für bayesianische Sachverhalte dar, da sie das Verwechseln von bedingten Wahrscheinlichkeiten vorbeugen, indem man sie von oben und von unten ablesen kann. Auf diese Weise ist der Unterschied zwischen dem positive Vorhersagewert  $P(C|T+)$  mit der Sensitivität  $P(T+|C)$ , welche bei der Verwendung der Formel von Bayes häufig verwechselt werden, offensichtlicher.<sup>61</sup>

In Abbildung 1 lässt sich dieser Vorteil gut erkennen. Betrachtet man den Häufigkeitsdoppelbaum von oben, kann man die Sensitivität des Test  $P(T+|C)$  ablesen: „Wie viele der Patienten mit Schilddrüsenkrebs erhalten eine auffällige Sonographie?“. Die richtige Antwort ist „2000 von

5000 Patienten“. Wählt man hingegen die umgekehrte Leserichtung und betrachtet den Häufigkeitsdoppelbaum von unten, so lässt sich die Frage nach dem positiven Vorhersagewert  $P(C|T+)$  beantworten: „Wie viele der Patienten mit auffälliger Sonographie haben tatsächlich Schilddrüsenkrebs?“, wobei die richtige Antwort nun „2000 von 13400 Patienten“ lautet. Diese Doppelbäume ermöglichen somit, dass sämtliche bedingte Wahrscheinlichkeiten direkt abgelesen werden können.<sup>61</sup>

Das Häufigkeitsnetz, siehe Abbildung 1, bestehend aus einer Knotenverzweigungsstruktur, kann als eine Erweiterung des Doppelbaums gesehen werden. Wie im Doppelbaum können alle vier Randwahrscheinlichkeiten und alle acht bedingten Wahrscheinlichkeiten abgebildet werden. Darüber hinaus und im Gegensatz zum Doppelbaum hat das Netzdiagramm vier Verzweigungen für die gemeinsamen Wahrscheinlichkeiten.<sup>25</sup> Die vier gemeinsamen Wahrscheinlichkeiten können vom mittleren Knoten aus diagonal abgelesen werden, wobei das Symbol  $\cap$  *Schnittmenge* bedeutet:  $P(C \cap T+)$ ,  $P(C \cap T-)$ ,  $P(nC \cap T-)$ ,  $P(nC \cap T+)$ . Die vier Randwahrscheinlichkeiten können vom mittleren Knoten aus horizontal und vertikal abgelesen werden:  $P(C)$ ,  $P(nC)$ ,  $P(T+)$ ,  $P(T-)$ . Die acht bedingten Wahrscheinlichkeiten können auf den äußeren Begrenzungsästen des Netzdiagramms abgelesen werden:  $P(C|T+)$ ,  $P(C|T-)$ ,  $P(nC|T+)$ ,  $P(nC|T-)$ ,  $P(T+|C)$ ,  $P(T+|nC)$ ,  $P(T-|C)$ ,  $P(T-|nC)$ .



**Abbildung 1** - Netzdiagramme- und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Schilddrüsenbeispiel. Die richtigen Lösungen sind hier bereits fettgedruckt gekennzeichnet. Grafik adaptiert nach Kunzelmann et.al. (2022).

### 1.3.4 Ziel der Studie

Ziel der vorliegenden Studie ist es, die Auswirkungen des Informationsformats (Wahrscheinlichkeiten vs. natürliche Häufigkeiten) und der Visualisierung (Doppelbaum vs. Netzdiagramm) auf die Genauigkeit und Geschwindigkeit, und damit die Effizienz des Entscheidungsprozesses zu vergleichen.

In vorangegangenen Studien konnte bereits der positive Effekt der Verwendung von Häufigkeiten statt Wahrscheinlichkeiten, sowie von Visualisierungen statt reinem Text gezeigt werden. Wir wollten daher den Effekt von Doppelbäumen und Netzdiagrammen auf das bayesianische Schlussfolgern untersuchen.

Die folgende Forschungsfrage soll beantwortet werden: Haben natürliche Häufigkeiten und Visualisierungen einen positiven Effekt auf die *Genauigkeit* und die *Diagnosegeschwindigkeit* bei bayesianischen Schlussfolgerungen? Dabei war der inhaltliche Kontext nur Nebenziel der Studie, da wir keinen Effekt auf die Lösung der bayesianischen Fragestellungen erwarteten. Unsere Hypothese war, dass Häufigkeitsvisualisierungen die diagnostische Effizienz am besten fördern. Da sich Häufigkeits-Doppelbäume und Häufigkeitsnetz-Diagramme beide als empfehlenswerte Visualisierungen für bayesianisches Schlussfolgern erwiesen haben, gab es im Vorfeld keine Hypothese darüber, welches dieser beiden Diagramme besser funktioniert, da bisher, nach unserem Wissen, noch kein direkter Vergleich beider Diagramme vorliegt.

## 2. Material und Methoden

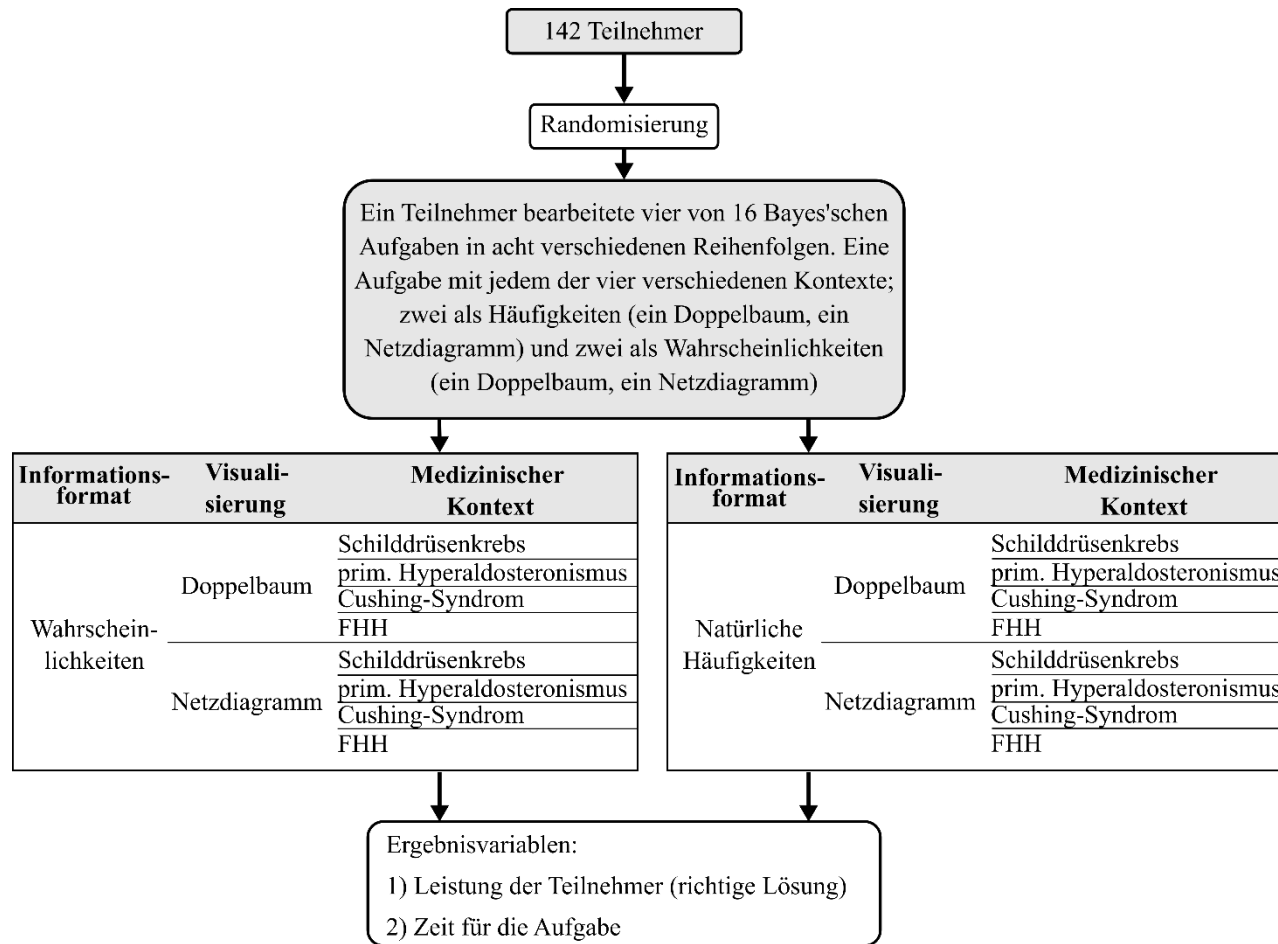
### 2.1 Teilnehmer

An dieser Studie nahmen 142 Medizinstudenten (90 Frauen, 51 Männer, 1 Unbekannter) teil, die zwischen November und Dezember 2020 alle Fälle bearbeitet haben. Die Teilnehmer wurden durch Öffentlichkeitsarbeit (d. h. Studienflyer) und Mundpropaganda an verschiedenen medizinischen Fakultäten in Deutschland rekrutiert. Die Teilnehmer waren im Durchschnitt 24,3 Jahre alt ( $SD = 2,9$ ), das durchschnittliche klinische Semester betrug 8,7 von maximal 12 ( $SD = 2,3$ ), die Schulprüfungsnote (1 = beste bis 6 = schlechteste) war  $M = 1,5$  ( $SD = 0,5$ ) und die Note der ersten ärztlichen Prüfung war  $M = 2,3$  ( $SD = 1,0$ ). Die Teilnehmer absolvierten durchschnittlich 2,7 Famulaturen ( $SD = 1,4$ ), die jeweils einen Monat dauerten.

Die Ethikkommission der LMU München genehmigte die Studie (Nummer: 20-0871). Alle Studenten wurden darüber informiert, dass die Teilnahme an der Studie freiwillig und anonym ist, und gaben ihr schriftliches Einverständnis zur Studienteilnahme. Jeder Student erhielt eine finanzielle Aufwandsentschädigung von zehn Euro.

### 2.2 Studiendesign

Abbildung 2 zeigt das  $2 \times 2 \times 4$  Studiendesign: 2 (*Informationsformat*: Wahrscheinlichkeiten vs. Häufigkeiten)  $\times$  2 (*Visualisierung*: Doppelbaum vs. Netzdiagramm)  $\times$  4 (*Kontexte*: Nebenziel der Studie). Zur Lösung der Aufgaben wurde die elektronische Fallsimulationsplattform CASUS<sup>62</sup> verwendet. Zunächst füllte jeder Studierende einen allgemeinen soziodemographischen Fragebogen aus (Alter, Geschlecht, klinisches Semester, Schulprüfungsnote, Note der ersten ärztlichen Prüfung, Anzahl der Famulaturen - als mögliche Kontrollvariablen, die die Leistung der Teilnehmer beeinflussen können) und las einen Einführungstext. Der Einführungstext informierte die Teilnehmer darüber, dass die Zeit zum Lösen der Fälle nicht begrenzt war und aufgezeichnet wurde. Die Teilnehmer konnten Taschenrechner benutzen. Jedem Studenten wurden randomisiert die Fälle zugeteilt, sodass jedem Studenten eine andere Reihenfolge der Aufgaben präsentiert wurde.



**Abbildung 2** - Studiendesign: 2 (Informationsformat: Wahrscheinlichkeiten vs. Häufigkeiten) x 2 (Visualisierung: Doppelbaum vs. Netzdiagramm) x 4 (medizinische Kontexte: Nebenziel der Studie). Zu den 4 medizinischen Kontexten gehörten Schilddrüsenkrebs, primärer Hyperaldosteronismus, das Cushing-Syndrom und die familiäre hypokalziurische Hyperkalzämie. Die Reihenfolge der Aufgaben wurde für jeden Teilnehmer randomisiert. Grafik adaptiert nach Kunzelmann et.al. (2022)

### 2.3 Medizinischer Kontext

Jeder Studierende löste vier Fälle unter Anwendung der Bayes Statistik, die aus verschiedenen Bereichen der klinischen Endokrinologie ausgewählt wurden: Cushing-Syndrom,<sup>63</sup> familiäre hypokalziurische Hyperkalzämie (FHH),<sup>64</sup> primärer Hyperaldosteronismus<sup>65</sup> und Schilddrüsenkrebs.<sup>14</sup>

Die Endokrinologie wurde ausgewählt, weil es in diesem Bereich der Inneren Medizin eine große Spanne unterschiedlich häufiger Erkrankungen in der Bevölkerung gibt. So sind beispielsweise Schilddrüsenknoten und primärer Hyperaldosteronismus recht häufig, während Krankheiten wie das Cushing-Syndrom und FHH sehr selten sind. Dies hat Einfluss auf die Interpretation der Befunde, da weniger Erfahrungswerte vorhanden sind. Dieser Aspekt kann in der täglichen klinischen Praxis eine Herausforderung darstellen. Diese genannten vier Kontexte wurden ausgewählt, um die klinische Realität darzustellen, indem zwei Szenarien, das Cushing-Syndrom und FHH, verwendet wurden, die den Studenten weniger vertraut sind, während die Szenarien Schilddrüsenkrebs und primärer Hyperaldosteronismus vertrauter sind. Alle Szenarien wurden im Curriculum behandelt. Die beiden genannten Beispiele, die in der Bevölkerung häufiger vorkommen werden im Curriculum auch intensiver behandelt, wodurch die Erfahrungswerte und Kenntnisse der Studierenden in diesem Gebiet ausgeprägter sind.<sup>66</sup>

Jeder Fall kann in vier verschiedenen Versionen dargestellt werden: Wahrscheinlichkeiten in Doppelbäumen, Wahrscheinlichkeiten in Netzdiagrammen, natürliche Häufigkeiten in Doppelbäumen und natürliche Häufigkeiten in Netzdiagrammen. Tabelle 1 zeigt beispielhaft den Schilddrüsenkrebsfall. Die entsprechenden Visualisierungen sind in Abbildung 1 dargestellt. Die anderen Fälle sind im Anhang zusammengestellt.

Jedem Teilnehmer wurden die oben genannten medizinischen Kontexte und Darstellungsformen gezeigt, die systematisch variiert wurden, so dass jeder Teilnehmer eine andere Abfolge von Aufgaben erhielt. Jeder Teilnehmer erhielt zwei medizinische Kontexte im Wahrscheinlichkeitsformat (einer visualisiert mit einem Doppelbaum und der andere mit einem Netzdiagramm) und die anderen beiden im Häufigkeitsformat (einer visualisiert mit einem Doppelbaum und der andere mit einem Netzdiagramm). Wie in Tabelle 1 dargestellt, wurden die Fragen in einem offenen Format gestellt. Am Ende jeder Aufgabe konnten die Teilnehmer ihre Antwort in ein dafür vorgesehenes Feld eintragen.



**Tabelle 1** - Fallbeschreibung des Schilddrüsenbeispiels. Grafik adaptiert nach Kunzelmann et.al. (2022).

<b>Fall</b>		<b>Schilddrüsenkarzinom</b>	
<b>Informationsformat</b>	<b>Wahrscheinlichkeiten</b>		<b>Häufigkeiten</b>
<b>Medizinischer Kontext</b>	<p>Stellen Sie sich vor, Sie arbeiten als Arzt in einer endokrinologischen Ambulanz. Hier werden u. a. Schilddrüsenultraschalluntersuchungen bei Verdacht auf Schilddrüsenknoten durchgeführt.</p> <p>Sie haben gerade einen Patienten konsultiert, bei dem sich in der Sonographie ein Schilddrüsenknoten zeigt, der aufgrund von Mikroverkalkungen Hinweise auf eine Malignität aufweist. Dieser Patient möchte wissen, wie wahrscheinlich es ist, dass es sich bei einem auffälligen sonographischen Befund tatsächlich um Schilddrüsenkrebs handelt.</p> <p>Für Ihre Antwort stehen Ihnen nur die folgenden Informationen zur Verfügung, die auf einer Stichprobe von Patienten mit Schilddrüsenknoten mit Verdacht auf Bösartigkeit aufgrund des Vorhandenseins von Mikroverkalkungen basieren.</p>		
<b>Visualisierung</b>	<ul style="list-style-type: none"> <li>• Wahrscheinlichkeits-Doppelbaum oder</li> <li>• Wahrscheinlichkeitsnetz-Diagramm (siehe Abbildung 1, links)</li> </ul>	<ul style="list-style-type: none"> <li>• Häufigkeits-Doppelbaum oder</li> <li>• Häufigkeitsnetz-Diagramm (siehe Abbildung 1, rechts)</li> </ul>	
<b>Fragestellung</b>	<p>Wie groß ist die Wahrscheinlichkeit, dass diese Person mit einem Schilddrüsenknoten mit Mikroverkalkungen tatsächlich Schilddrüsenkrebs hat?</p> <p>Antwort: _____</p>	<p>Wie viele Patienten, die einen Schilddrüsenknoten mit Mikroverkalkungen haben, haben tatsächlich Schilddrüsenkrebs?</p> <p>Antwort: ____ von ____</p>	

## 2.4 Kodierung

In Anlehnung an Gigerenzer und Hoffrage,<sup>24</sup> wurden bei den Aufgaben mit Wahrscheinlichkeiten die Antwort als richtig bewertet, wenn sie der exakten bayesianischen Lösung entsprach oder auf den nächsten ganzen Prozentpunkt darüber oder darunter gerundet wurde. Bei den Aufgaben mit natürlichen Häufigkeiten wurde die Antwort als richtig bewertet, wenn beide Zahlen der Häufigkeitsangaben korrekt angegeben wurden (z. B. bei der Lösung des Schilddrüsenkrebs-Problems "2000 von 13400", sowohl die 2000 als auch die 13400).

### Statistik

Zur Auswahl der Stichprobengröße wurde die Power-Analyse früherer, ähnlich angelegter Studien zugrunde gelegt. Diesen Studien zufolge liegt die Aussagekraft des Formateffekts bei einer Stichprobengröße von etwa N=120 Studierenden bei nahezu 100 % (95 % CI, 96,4 % bis 100 %).<sup>17,27</sup>

Für die statistische Analyse wurde die Programmiersprache R<sup>67</sup> verwendet. Verallgemeinerte lineare gemischte Modelle (mit einer Logit-Verknüpfungsfunktion) wurden verwendet, um 1) das korrekte Ergebnis der bayesianischen Aufgabe und 2) die Zeit zur Lösung der Aufgabe vorherzusagen. In diesem Modell wurde die Wahrscheinlichkeits-Doppelbaum-Version als Referenzkategorie festgelegt und die potenziellen Erklärungsfaktoren *Häufigkeiten* und *Netzdiagramm* über Dummy-Kodierung einbezogen.

Zur Bewertung der diagnostischen Effizienz wurde eine Punktzahl nach Binder et al. (2021) verwendet<sup>3</sup>. Dieser Wert für die diagnostische Effizienz ist der Median der Bearbeitungszeit geteilt durch den Anteil der richtigen Schlussfolgerungen. Niedrigere Werte dieses Scores bedeuten mehr richtige und schnellere Diagnosen. Der Median wurde anstelle des arithmetischen Mittels verwendet, da die Bearbeitungszeiten stark rechtsschief sind. So wie es bei den Bearbeitungszeiten der Teilnehmer normalerweise auch der Fall ist. Die im Voraus geplante Strategie für sehr lange Bearbeitungszeiten bestand darin, alle Teilnehmer mit einer Bearbeitungszeit von mehr als dem dreifachen Interquartilsabstand auszuschließen.

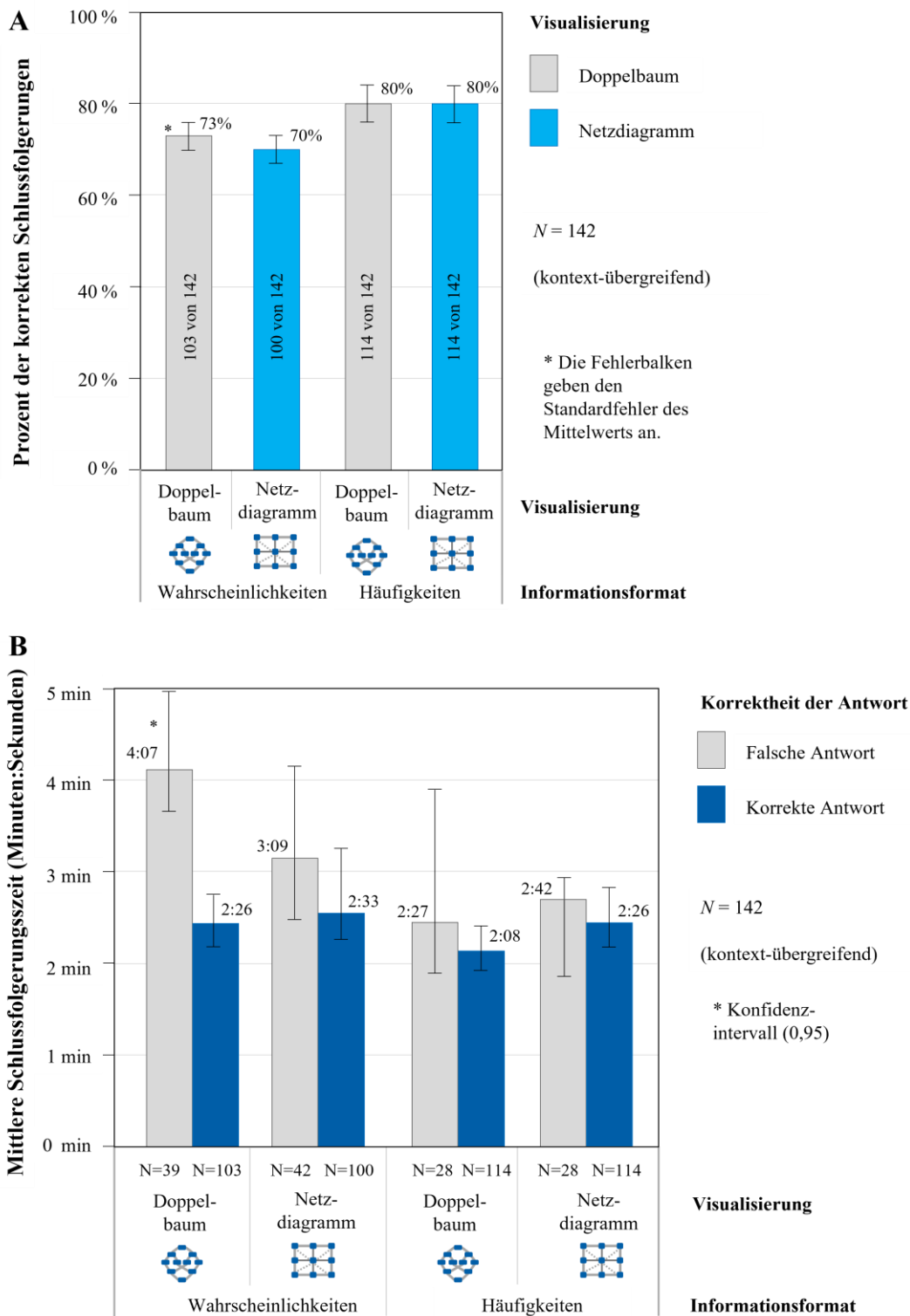
### 3. Ergebnisse

#### 3.1 Leistung der Teilnehmer (Genauigkeit)

Wie aus Abbildung 3 und Tabelle 2 hervorgeht, war die Zahl der richtigen Antworten der Teilnehmer in allen Versionen hoch. Die Teilnehmer, denen die Wahrscheinlichkeitsversion vorgelegt wurde, beantworteten 73 % der Fälle in der Doppelbaumversion und 70 % der Fälle in der Netzdiagrammversion richtig. Bei den Versionen mit natürlicher Häufigkeit war das Ergebnis mit 80 % richtig gelöster Fälle für beide Visualisierungsformate gleich.

**Tabelle 2** - Kontextübergreifende Gesamtergebnisse. Prozentsatz der richtigen Schlussfolgerungen, Zeit für die Diagnose mit Median, 1. (Q1) und 3. (Q3) Quartil, Punktzahl für die diagnostische Effizienz. Der Wert für die diagnostische Effizienz wurde berechnet, indem die mediane Zeit durch den Prozentsatz der richtigen Schlussfolgerungen geteilt wurde. N gibt die Anzahl der Teilnehmer an. Grafik adaptiert nach Kunzelmann et.al. (2022).

		N	Richtige Schlussfolgerungen	Zeit bis zu einer Diagnose			Zeit bis zu einer korrekten Diagnose			Diagnose-Effizienz-Score: Mittlere Zeit für alle Diagnosen/richtige Schlussfolgerungen [m:ss]	
				Q1	Median	Q3	N	Q1	Median		Q3
<b>Wahrscheinlichkeiten</b>	Doppelbaum	142	73 %	1:56	2:50	4:26	103	1:44	2:26	3:40	3:53
	Netzdiagramm	142	70 %	1:46	2:50	4:16	100	1:46	2:33	3:49	4:03
<b>Häufigkeiten</b>	Doppelbaum	142	80 %	1:32	2:14	3:25	114	1:33	2:08	3:19	2:48
	Netzdiagramm	142	80 %	1:41	2:28	3:37	114	1:41	2:26	3:40	3:05

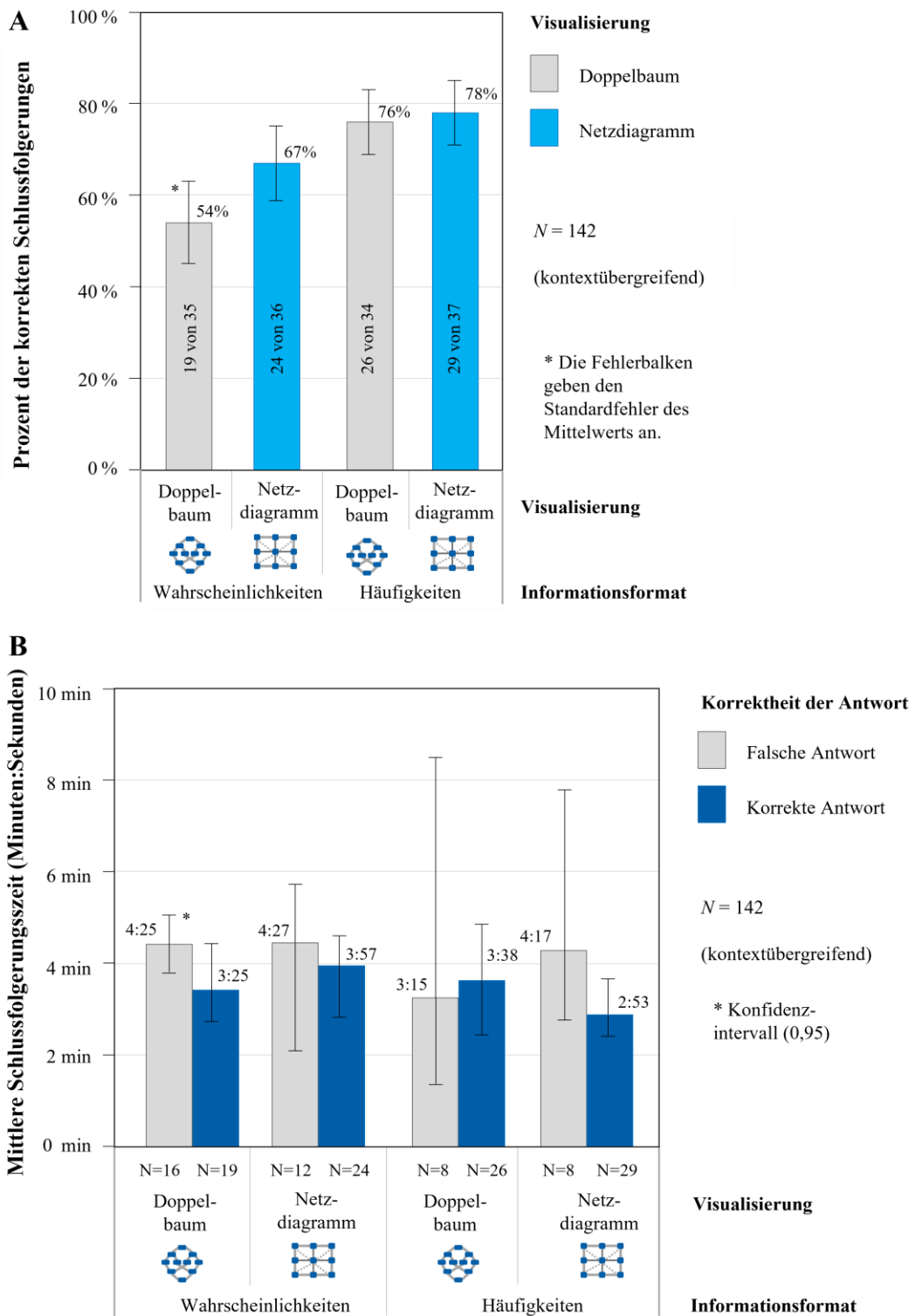


**Abbildung 3** - (A) Leistung der Teilnehmer bei den Aufgaben zum bayesianischen Schlussfolgern (kontextübergreifend). (B) Mittlere Zeit für das Lösen einer bayesianischen Aufgabe richtig oder falsch zu lösen (kontextübergreifend). Grafik adaptiert nach Kunzelmann et.al. (2022).

In einem verallgemeinerten linearen gemischten Modell (GLMM) zur Vorhersage der Genauigkeit waren die nicht standardisierten Regressionskoeffizienten für natürliche Häufigkeiten ( $b_1 = 0,75$ ,  $SE = 0,25$ ,  $z = 2,96$ ,  $p = 0,003$ ) und die Reihenfolge der Aufgabe (d. h. wurde die Aufgabe als erste, zweite, dritte oder vierte Aufgabe gestellt;  $b_2 = 0,36$ ,  $SE = 0,11$ ,  $z = 3,12$ ,  $p = 0,002$ ) signifikant von Null verschieden. Es lag also ein Lerneffekt vor. Die erste Aufgabe wurde im Durchschnitt schlechter bearbeitet. Deskriptive Ergebnisse, die sich nur auf die erste Aufgabe beziehen, finden sich in Abbildung 4. Wird nur die erste Aufgabe analysiert, ist der Effekt des Häufigkeitsformats stärker ausgeprägt. Andere Variablen, wie die Visualisierung, der Kontext, das Alter, das Geschlecht und die Schulabschlussnote, hatten dagegen keinen signifikanten Einfluss auf die Leistung der Teilnehmer. Das vollständige Modell ist in Tabelle 3 - Ergebnisse des GLMM zur Vorhersage der Genauigkeit der Antwort. zu finden.

**Tabelle 3** - Ergebnisse des GLMM zur Vorhersage der Genauigkeit der Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).

<b>Ergebnisse des GLMM zur Vorhersage der Genauigkeit der Antwort</b>				
*: signifikant auf einem 5 %-Niveau (zweiseitige Prüfung)				
	(nicht standardisiert) Regressionskoeffizient	Standardabweichung	z-Wert	p-Wert
Achsenabschnitt	3,40	1,18	2,88	0,004*
Informationsformat: Häufigkeiten	0,75	0,25	2,96	0,003*
Visualisierung: Netzdiagramm	-0,10	0,25	-0,41	0,68
Schulabschlussnote	-0,76	0,41	-1,86	0,06
Reihenfolge der Auf- gaben	0,36	0,11	3,12	0,002*
Klinisches Semester	-0,18	0,09	-1,88	0,06



**Abbildung 4** - (A) Leistung der Teilnehmer bei der ersten Aufgabe zum bayesianischen Schlussfolgern (kontextübergreifend). (B) Mittlere Zeit für das richtige oder falsche Lösen der ersten bayesianischen Schlussfolgerungsaufgabe (kontextübergreifend). Grafik adaptiert nach Kunzelmann et.al. (2022).

## 3.2 Zeit für die Lösung der Aufgabe und Effizienz

Tabelle 2 zeigt nicht nur den mittleren Zeitaufwand der Teilnehmer für jeden Fall, sondern auch den mittleren Zeitaufwand für die korrekte Lösung der Aufgabe (siehe auch Abbildung 3 - (A) Leistung der Teilnehmer bei den Aufgaben zum bayesianischen Schlussfolgern (kontextübergreifend). (B) Mittlere Zeit für das Lösen einer bayesianischen Aufgabe richtig oder falsch zu lösen (kontextübergreifend). Grafik adaptiert nach Kunzelmann et.al. (2022).). Wie in Tabelle 2 und Abbildung 3 zu sehen ist, wurden die Aufgaben mit natürlichen Häufigkeiten schneller gelöst. Die Version mit dem Häufigkeits-Doppelbaum war diejenige, die am schnellsten gelöst wurde, nämlich mit einer mittleren Zeit von 2:14 Minuten. Es ist auch auffällig, dass richtige Lösungen durchweg etwas schneller gegeben werden als falsche Antworten.

Ein lineares gemischtes Modell zur Schätzung der Zeit bis zur korrekten Diagnose wurde modelliert, wobei acht Ausreißer mit einer Fertigstellungszeit von mehr als  $3 \times$  Interquartilsbereich ausgeschlossen wurden.<sup>68</sup> Das Modell zeigte, dass die Visualisierung (*Doppelbaum* vs. *Netzdiagramm*) keinen signifikanten Einfluss auf die Geschwindigkeit der Antwort hatte ( $p = 0.89$ ). Die Darstellung der natürlichen Häufigkeiten ( $p = 0,04$ ) und die Reihenfolge der Aufgaben (d. h., wurde die Aufgabe als erste, zweite, dritte oder vierte Aufgabe präsentiert;  $p < 0,001$ ) hatten jedoch einen signifikanten Einfluss auf die Lösungsgeschwindigkeit (Tabelle 4). Darüber hinaus benötigte der Kontext des *Cushing-Syndroms* signifikant mehr Zeit, ohne dass sich die berichteten Auswirkungen des Informationsformats und der Reihenfolge der Aufgaben änderten. Das heißt, der inhaltliche Kontext hatte in diesem Fall wohl Einfluss auf den erhöhten Zeitbedarf.

**Tabelle 4** - Ergebnisse des LMM zur Vorhersage der Geschwindigkeit der richtigen Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).

<b>Ergebnisse des LMM zur Vorhersage der Geschwindigkeit der richtigen Antwort</b>				
<b>(ohne 8 Ausreißer mit Zeit &gt; 3× Interquartilsbereich)</b>				
*: signifikant auf einem 5%-Niveau (zweiseitige Prüfung)				
	(standardisierter)	SE (Standardfehler)	t-Wert	p-Wert
	Regressionskoeffizient			
Achsenabschnitt	0,04	0,07	0,53	0,60
Informationsformat: Natürliche Häufigkeiten	-0,08	0,04	-2,07	0,04*
Visualisierung: Netzdiagramm	0,006	0,04	0,14	0,89
Schulabschlussnote	0,06	0,07	0,89	0,37
Reihenfolge der Aufgaben	-0,31	0,04	-8,11	<0,001*
Kontext: Cushing	0,12	0,05	2,46	0,01*
Kontext: FHH	0,08	0,05	1,60	0,11
Kontext: Schilddrüsenkrebs	0,01	0,05	0,25	0,81

Ein zusätzliches linear gemischtes Modell zur Schätzung der Zeit bis zur Antwort - unabhängig davon, ob die Antwort richtig oder falsch war - zeigt ein sehr ähnliches Bild (signifikanter Effekt des Formats  $p = 0,42$  und der Reihenfolge der Aufgabe  $p = 0,003$ ; siehe Tabelle 5). Darüber hinaus benötigten die Kontexte *Cushing-Syndrom* und *FHH* im letztgenannten linearen gemischten Modell signifikant mehr Zeit, ohne dass sich die berichteten Effekte des Informationsformats und der Reihenfolge der Aufgaben veränderten. Andere Variablen hatten keinen signifikanten Einfluss auf die Bearbeitungszeit.



**Tabelle 5** - Ergebnisse des LMM zur Vorhersage der Geschwindigkeit Antwort. Grafik adaptiert nach Kunzelmann et.al. (2022).

<b>Ergebnisse des LMM zur Vorhersage der Geschwindigkeit der Antwort</b>				
<b>(ohne 8 Ausreißer mit Zeit &gt; 3×Interquartilsbereich)</b>				
*: signifikant auf einem 5%-Niveau (zweiseitige Prüfung)				
	(standardisierter) Regressionskoeffizient	SE (Standardfehler)	t-Wert	p-Wert
Achsenabschnitt	0,01	0,06	0,17	0,86
Informationsformat: Natürliche Häufigkeiten	-0,09	0,03	-3,04	0,003*
Visualisierung: Netzdiagramm	-0,03	0,03	-0,82	0,42
Schulabschlussnote	0,06	0,06	1,08	0,28
Reihenfolge der Aufgaben	-0,32	0,03	-10,20	<0,001*
Kontext: Cushing	0,11	0,04	2,91	0,004*
Kontext: FHH	0,09	0,04	2,35	0,02*
Kontext: Schilddrüsenkrebs	0,05	0,04	1,25	0,21

Darüber hinaus zeigt Tabelle 2 den Wert zur Analyse der *diagnostischen Effizienz*. Der Score für die diagnostische Effizienz dividiert die mittlere Zeit für die Aufgabe durch den Anteil der richtigen Schlussfolgerungen. Der beste Wert für die diagnostische Effizienz wurde bei der Häufigkeits-Doppelbaum-Version erzielt (2:48 min), während der schlechteste Wert bei der Wahrscheinlichkeitsnetz-Diagramm-Version erzielt wurde (4:03 min).

## 4. Diskussion

### 4.1 Häufigkeits-Doppelbäume und Häufigkeitsnetze erhöhen die Effizienz im Vergleich zu den entsprechenden Wahrscheinlichkeitsvisualisierungen

Diese Studie untersuchte den Einfluss des Informationsformats (Häufigkeiten vs. Wahrscheinlichkeiten) und verschiedener Visualisierungen (Doppelbaum vs. Netzdiagramm) auf die Effizienz, das bedeutet auf die *Genauigkeit* (bestimmt über die Zahl der richtigen Antworten) und *Zeit für das Lösen der Aufgabe*, von Medizinstudenten bei bayesianischen Schlussfolgerungsaufgaben. Es konnte gezeigt werden, dass Häufigkeitsvisualisierungen häufiger zu richtigen Lösungen und sogar zu schnelleren Lösungen führen als die entsprechenden Wahrscheinlichkeitsvisualisierungen. Die Visualisierungstypen Doppelbaum und Netzdiagramm führten nicht zu einem signifikanten Unterschied in der Leistung der Medizinstudenten beim Lösen von Bayes-Aufgaben.

### 4.2 Auch Wahrscheinlichkeits-Doppelbäume und Wahrscheinlichkeitsnetz-Diagramme führen bei Medizinstudenten zu hohen Erfolgsquoten

Im Gegensatz zu den Ergebnissen von Binder et al. (2020)<sup>25</sup> führten das Wahrscheinlichkeits-Doppelbaum- und das Wahrscheinlichkeitsnetz-Diagramm in unserer Studie zu bemerkenswert guten Ergebnissen in Form von richtigen Antworten (73 % und 70 %). Bei Binder et al. waren die richtigen Antworten der Teilnehmer bei der Wahrscheinlichkeits-Doppelbaum- und der Wahrscheinlichkeitsnetz-Version (jeweils 34 %) deutlich weniger häufig als bei der Häufigkeits-Doppelbaum- (60 %) und der Häufigkeitsnetz-Version (61 %), in jedem Fall aber deutlich weniger häufig richtig als in unserer Studie. Der Grund dafür könnte unter anderem darin liegen, dass unterschiedliche Stichproben betrachtet wurden: Die vorliegende Studie wurde mit Medizinstudenten durchgeführt. Bei Binder et al. nahmen Studenten verschiedener Fachrichtungen teil, wobei 117 der Studenten aus dem Bereich der Pharmazie stammten. Das gute Abschneiden der Medizinstudenten bei den Wahrscheinlichkeitsversionen in unserer Studie könnte darauf hinweisen, dass Medizinstudenten möglicherweise auf Grund der Vorlesungen zu medizinischer Statistik in der Lage sind, zwischen den Wahrscheinlichkeitsbegriffen zu unterscheiden. Sie können somit  $P(C|T+)$  von den typischen Störfaktoren  $P(T+|C)$  und  $P(C \cap T+)$  differenzieren.<sup>25</sup> Das konzeptionelle Verständnis, das notwendig ist, um  $P(C|T+)$  von  $P(T+|C)$  und  $P(C \cap T+)$  zu unterscheiden, scheint also bereits entwickelt worden zu sein. Da beim Doppelbaum und beim Netz die richtigen

Werte nur noch erkannt werden müssen, wäre ein passives Wissen bereits ausreichend, die richtige Antwort zu identifizieren. Das notwendige prozedurale Wissen, um  $P(C|T+)$  aus der Prävalenz, der Sensitivität und der falsch-positiven Rate zu berechnen, scheint noch zu fehlen.<sup>3</sup> Bei anderen Formaten, bayesianische Fragestellungen zu präsentieren, wie beispielsweise im einfachen Baum oder als Text, müssen Berechnungen durchgeführt werden, das heißt man muss über das nötige prozedurale Wissen zur Lösung verfügen. Der Lernaufwand, Wissen aktiv verfügbar zu haben, ist größer als der Lernaufwand, die richtige Lösung passiv zu erkennen.

### **4.3 Vergleich von Doppelbäumen und Netzdiagrammen mit einfachen Baumdiagrammen**

Wassner et. al. 2002 beschreiben bereits die Erweiterung eines klassischen Baumdiagramms zu einem Doppelbaum. Sie sahen den wesentlichen Vorteil des Doppelbaums in der sequentiellen Darstellung der Informationen und der Ablesbarkeit in beide Richtung, was das logische Verständnis der Zusammenhänge der gegebenen Zahlen erhöht.<sup>69</sup>

Doppelbäume und Netzdiagramme enthalten bereits die korrekte Lösung der bayesianischen Aufgabe. In unserer Studie mussten die Teilnehmer daher nur die richtige Lösung unter vielen anderen Wahrscheinlichkeits- oder Häufigkeitsinformationen auswählen. Dies erklärt, warum die Leistung in unserer Studie höher und die Medianzeit niedriger ist als bei Binder et al. (2021)<sup>3</sup>, die den Einfluss *einfacher* Baumdiagramme, mit Wahrscheinlichkeiten oder Häufigkeiten, auf die Leistung und den Zeitaufwand der Teilnehmer bei den Aufgaben untersuchten. Das Lösen von bayesianischen Aufgaben mit einem *einfachen* Baumdiagramm erfordert mehr Denkschritte und ist daher komplexer und demnach auch zeitaufwändiger.<sup>23</sup> Beispielsweise müssen in *einfachen* Baumdiagrammen die Anzahl an Personen mit positivem Testergebnis aufsummiert werden. Diese Information ist im Doppelbaum in der unteren Ebene bereits enthalten. Außerdem müssen beim einfachen Baumdiagramm die relevanten Äste und Knoten des Baumes erst identifiziert und richtig miteinander verrechnet werden.

### **4.4 Daten sollten bereits mit Visualisierungen veröffentlicht werden - Algorithmus zur Erstellung der Darstellungen**

Wenn neue Daten zu Krankheiten und ihren Tests bzw. Screeningverfahren veröffentlicht werden, wäre es sinnvoll, alle Daten bereits mit einer Visualisierung beispielsweise durch einen Häufigkeitsdoppelbaum oder ein Häufigkeitsnetz zu präsentieren. Hierzu benötigt man lediglich die Prävalenz, die Größe der entsprechenden Stichprobe, die Testsensitivitäten und die Testspezifität

bzw. die falsch-positiv Rate. Diese sollte, sofern sie in Prozentzahlen vorliegen, zunächst in Häufigkeiten umformuliert werden. Anhand dieser Daten lassen die Diagramme in wenigen Schritten erzeugen, vergleiche Abbildung 1.

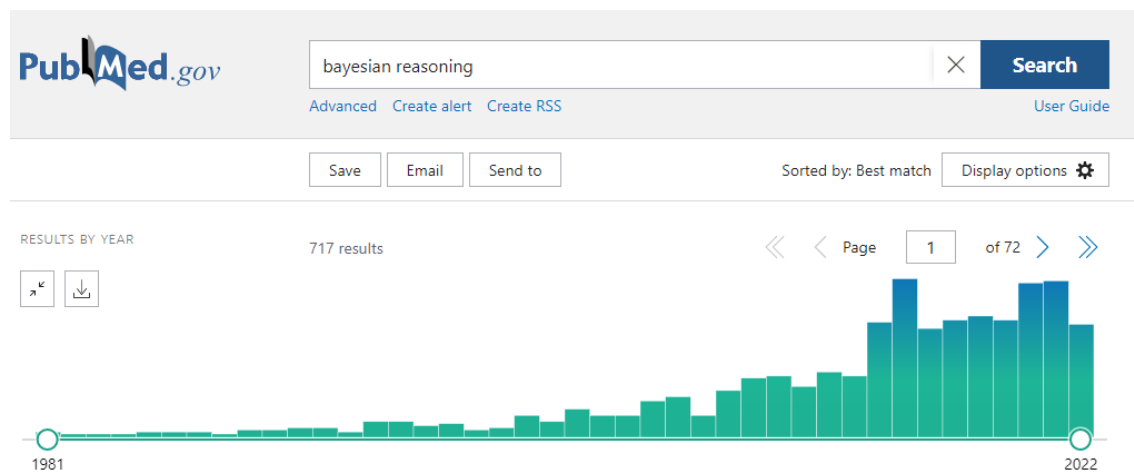
Die Stichprobengröße selbst, beziehungsweise eine didaktisch optimierte Repräsentation der Stichprobengröße, wird beim Doppelbaum beispielsweise ganz oben und auch ganz unten in den Kasten eingefügt. Eine Ebene tiefer von oben aus gelesen folgen dann die Daten zu Prävalenz, es werden die Anzahl der erkrankten und gesunden Personen, bzw. eine didaktisch gewählte Repräsentation davon, in die beiden Kästen eingetragen.

Vom unteren Knoten ausgehend folgen dann in der nächsten Ebene die Häufigkeiten der Personen mit positivem bzw. negativem Testergebnis. Die mittlere Ebene stellt dann die gemeinsamen Häufigkeiten dar, also die Kombinationen von erkrankten Personen mit positivem Testergebnis, erkrankte Personen mit negativem Testergebnis, sowie auch entsprechend der Gesunden mit positivem Testergebnis und die Gesunden mit negativem Testergebnis.

Die relevanten Zahlen, beispielsweise zum positiven Vorhersagewert, sollten immer markiert werden, da auf diese Weise fehlerhaftes Ablesen der Werte vermieden werden kann. Die Publikation von Daten könnten auf diese Weise bereits didaktisch sinnvoll publiziert werden, sodass Mediziner die Informationen schneller und richtiger interpretieren könnten. Um die Doppelbaumdiagramme und das Häufigkeitsnetz auch ohne vorheriges Training leicht interpretieren zu können, wäre zudem denkbar, optisch farblich oder durch unterschiedliche Schriftarten und/oder Schriftgrößen abgesetzt zusätzliche, erläuternde statistischen Informationen an den Rand der jeweiligen Zellen zu schreiben, wie z. B. Testsensitivität und die Testspezifität bzw. die falsch-positiv Rate.

### **4.5 Einflüsse auf Lösung bayesianischer Problemstellungen**

Das bessere Abschneiden der Studenten in dieser Studie könnte auch damit zusammenhängen, dass die Zahl der Veröffentlichungen über bayesianisches Denken und Entscheidungsfindung im letzten Jahrzehnt stark zugenommen hat, vergleiche Abbildung 5.<sup>12</sup>



**Abbildung 5** - Bildschirmfoto zur Recherche mit den Begriffen *bayesian reasoning medical* in der Datenbank PubMed. Das Histogramm zeigt, wie sich die Häufigkeit der zu dieser Begriffskombination gefundenen Artikel im Zeitraum von 1981 bis 2022 entwickelt hat. Allein in den letzten *fünf* Jahren wurden 44 % (367 von 833) der seit 1987 erschienen Artikel publiziert. Stand der Recherche: 5.9.2022.

Infolgedessen erhält das Thema inzwischen viel Aufmerksamkeit, weshalb sich mehr Menschen damit beschäftigen. Dadurch ist bereits anzunehmen, dass bayesianisches Schlussfolgern mittlerweile in der medizinischen Lehre stärker vertreten ist. Wie in ähnlichen Studien von vor zehn Jahren zu sehen ist, waren nur etwa 30 %<sup>70</sup> – 40 %<sup>71</sup> Teilnehmer in der Lage, die bayesianische Aufgabe richtig zu lösen, während in dieser Studie über 70 % der Teilnehmer auf die richtige Lösung wählten. Im Alltag eines Studierenden könnte das Wissen zu Bayes generell dadurch gefördert werden, dass Dozenten ihre Daten in Häufigkeitsdoppelbäumen oder Netzen darstellen und somit die Studierenden auf die richtige Interpretation bayesianischer Aufgabestellungen trainieren.

Die guten Ergebnisse könnten zusätzlich darauf zurückzuführen sein, dass *evidenzbasierte Medizin* in der medizinischen Lehre enorm an Bedeutung gewonnen hat, im Zuge dessen ist auch die Biomathematik und somit die Epidemiologie, Statistik und Entscheidungsfindung präsenter geworden ist.<sup>72</sup> Daher werden in der medizinischen Ausbildung mehr Diagramme dieser Art verwendet, weshalb ein gewisser Trainingseffekt für Medizinstudenten nachgewiesen werden kann. Es gibt im Eignungstest für das Medizinstudium in Deutschland einen gesonderten Testabschnitt, der ausschließlich die Kompetenz prüft, naturwissenschaftliche Zusammenhänge in Diagrammen und Tabellen zu interpretieren.<sup>73</sup> Da der Eignungstest für das Medizinstudium im Auswahlverfahren der Hochschulen einen hohen Stellenwert hat und über das Auswahlverfahren der Hochschulen 60 % aller Medizinstudienplätze vergeben werden, kann man davon ausgehen, dass die Medizinstudenten eine positive Selektion hinsichtlich der Menschen darstellen, die Diagramme und Tabellen gut interpretieren können. Im Rahmen der Zusatzinformationen könnte man in zu-

künftigen Studien abfragen, ob die Teilnehmer am Eignungstest für das Medizinstudium teilgenommen haben und ob diese Teilnahme einen Einfluss auf die richtigen Antworten und die diagnostische Effizienz hat.

Auch das Training der Medizinstudenten, Multiple-Choice-Fragen gut beantworten zu können, trägt zu dem Bias bei, Aufgaben schnell verstehen,<sup>74</sup> und Muster schnell erkennen zu können.<sup>75</sup> In der vorliegenden Studie kann das gute Abschneiden der Studenten in der korrekten der Lösung der bayesianischen Fragestellungen durch dieses Multiple-Choice-Training gefördert sein. Jede Aufgabe war nach dem gleichen Schema aufgebaut und die richtige Lösung stand immer an der gleichen Stelle, sodass das konzeptionelle Wissen zur bayesianischen Statistik ausgereicht hat, die Aufgaben schnell und richtig zu lösen. Die Tendenz, die sich durch das Multiple-Choice-Training ergibt, kann man nur schwer ausgleichen. Am einfachsten wäre, Studenten in frühen Studiensemestern für diese Art der Untersuchungen zu rekrutieren, die noch nicht so viel Routine mit Multiple-Choice-Fragen haben. Das hätte aber den Nachteil, dass man viele Themen nicht untersuchen kann, da das erforderliche klinische Verständnis noch nicht vorliegen würde.

Jedoch ist der Lerneffekt durch Mustererkennung auch fehleranfällig. Wenn Studienteilnehmer glaubten, die richtige Zahl im Diagramm erkannt zu haben, haben sie potenziell in den Folgeaufgaben ähnliche Zahlen an der gleichen Stelle im Diagramm ausgewählt. Sie haben dadurch möglicherweise Fehler gemacht, weil sie die falsche Wahrscheinlichkeit ausgewählt haben.<sup>25</sup> Wie genau die Herangehensweise der Studienteilnehmer hinsichtlich der Lösung und Erkennen der Lösung im Diagramm ablief, könnte mittels Eye-Trackingstudien genauer untersucht werden.<sup>76</sup> Die Analyse der bei der Lösung der Aufgabe entstandenen Fehlermuster kann zur Verbesserung der Visualisierungen für die Zukunft dienen.

In zukünftigen Studien sollten noch mehr unterschiedliche Darstellungsformen getestet werden als in dieser Studie. Um Lerneffekte zu minimieren, könnten außerdem die Beschriftungen der Diagramme verändert werden, damit die Studierenden nicht immer nur denselben Ast oder Kasten lesen müssen und somit im Voraus wissen, wo sie hinschauen müssen. Alternativ könnten auch die Formulierungen der Problembeschreibungen und der Fragen variiert werden, um dem Lerneffekt durch Ablenkung entgegenzuwirken.

Die Informationen zur Semesterzahl deuteten darauf hin, dass die Studenten die Kurse bereits hatten. Das ist allerdings nur eine indirekte Beurteilung über die Zahl des Fachsemesters. Die Aussagekraft wäre besser, wenn man bei zukünftigen Studien ergänzend zu den demographischen Grundlagen zusätzlich abfragt, welche themenrelevanten Kurse die Studenten bereits belegt hatten.

## 4.6 Psychologische Aspekte von Diagrammen und Persönlichkeitseigenschaften / Optimierung der Präsentationsform

In zukünftigen Studien sollte auch der Einfluss der Darstellungsform des Netzdiagramms genauer untersucht werden. In einem Netzdiagramm könnte man sowohl die Häufigkeiten in den Knoten als auch die Wahrscheinlichkeiten an den Pfeilen darstellen. Auf diese Weise können alle Informationen auf einen Blick dargestellt werden, was zu weniger Fehlern und einer höheren diagnostischen Effizienz führen kann. Außerdem könnte man zusätzliche Diagramme darstellen, bei denen die richtige Lösung bereits farbig hervorgehoben ist, denn durch visuelle Hilfen, wie Farben und Formen kann die Aufmerksamkeit des Lesers gelenkt werden und die Daten der Grafiken somit leichter interpretiert werden.<sup>77</sup>

Zudem kann man davon ausgehen, dass verschiedene Persönlichkeitsmerkmale zu einer besseren Problemlösungsleistung führen. Dazu gehören numerische Fähigkeiten, visuell-räumliche Fähigkeiten,<sup>78,79</sup> Arbeitsgedächtnis,<sup>80</sup> kognitive Reflexion,<sup>70</sup> kognitive Verarbeitung,<sup>39</sup> kognitive Ressourcen,<sup>81</sup> und kognitive Fähigkeiten.<sup>82</sup> Es hat sich gezeigt, dass numerische Fähigkeiten und visuell-räumliche Fähigkeiten zu unterschiedlichen Lösungseffizienzen bei Bayes-Aufgaben führen.<sup>42</sup> Daher sollte mehr Forschung betrieben werden, um zu vergleichen, ob Medizinstudenten andere kognitive Fähigkeiten oder Persönlichkeitsmerkmale haben, wie beispielsweise die Fähigkeit, bayesianische Aufgaben und Visualisierungen zu verstehen, als Studenten anderer Fachrichtungen. Durch eine solche Evaluierung könnte die medizinische Ausbildung langfristig dahingehend modifiziert werden, dass nicht nur unterstützende Formen der Visualisierung, sondern auch ein gezieltes Training numerischer und visuell-räumlicher Fähigkeiten zur Förderung des erfolgreichen Lösens bayesianischer Aufgaben angeboten werden.<sup>42</sup>

Die Ergebnisse dieser Studie tragen zur Verbesserung der Ausbildung im medizinischen Bereich bei. Die Ausbildung der Medizinstudenten sollte an die Ergebnisse dieser Studie angepasst werden, da dies auch ihre Kompetenz in der klinischen Praxis erhöhen würde. Bayesianische Probleme sollten unter Verwendung von natürlichen Frequenzen gelehrt werden, die in Doppelbäumen oder Netzdiagrammen modelliert werden. Diese Probleme lassen sich schneller lösen, weil dieses Schema für die Studierenden leichter zu verstehen ist. Außerdem ist diese Strategie einprägsamer als die bayesianische Formel.<sup>32</sup>

Man könnte auch die Präsentationsform bayesianischer Fragestellungen optimieren, indem man sie so visualisiert, dass Farbfehlsichtigkeit kein Problem darstellen kann. Somit sollte immer eine Visualisierungsoption bestehen, die ausschließlich eine schwarz-weiße Grafik enthält. Außerdem sollte man Diagramme ergänzen, die Fehlsichtigkeit unterstützen, somit entweder in einer Onlineversion auditiv unterstützt oder in einer Druckversion durch Brailleschrift ergänzt.

## 4.7 Vorteile für die medizinische Ausbildung durch Häufigkeits-Doppelbäume und Häufigkeitsnetze

Bayesianische Probleme so darzustellen, dass sie schnell gelöst werden können, ist in der täglichen medizinischen Praxis von großer Bedeutung. Das zeigt sich zum Beispiel daran, dass ein Arzt in einer Hausarztpraxis durchschnittlich fünf Minuten Zeit hat, um einen Patienten zu behandeln.<sup>83</sup> Es ist unwahrscheinlich, dass Ärzte bei jeder neuen bayesianischen Entscheidung die Daten neu berechnen, dafür ist leider nicht genug Zeit. Außerdem haben die Ärzte möglicherweise keinen Zugang zu den neuesten Erkenntnissen, wie Sensitivitäten oder falsch-positive Raten. Im klinischen Alltag wird man neue Erkenntnisse einmal durchdenken und dann einfache Wenn-dann-Regeln formulieren, die helfen, schnell Entscheidungen zu treffen (z. B. wenn der neue Test eine höhere Sensitivität hat, wird der positive prädiktive Wert höher als bei dem Test zuvor). Gerade in der Notfallmedizin, wo der Zeitdruck noch höher ist, wären Heuristiken (statt konkreter Berechnungen) und Visualisierungen mit Doppelbäumen oder Netzdiagrammen noch hilfreicher.<sup>84,85</sup>

Im klinischen Alltag könnte es hilfreich sein, den Ärzten die Visualisierungen immer in der gleichen Form zu präsentieren, damit sie die richtige Lösung schnell erkennen können, ohne zusätzliche Berechnungen durchführen zu müssen. Besonders gut hierfür wären Doppelbäume- und Häufigkeitsnetze, die bereits alle relevanten Daten enthalten, wie das Beispiel aus Abbildung 1. Diese Abbildungen sollten bereits von den Gesellschaften in den Leitlinien implementiert werden. Auf diese Weise muss nicht jeder Mediziner ein eigenes Diagramm erstellen.

Es gibt noch immer ein großes Defizit bei klinisch tätigen Medizinerinnen in der Kenntnis von Statistik. Beispielweise kannten in einer Studie fast 20 % der Medizinprofessoren und -dozenten die korrekte Definition der Sensitivität nicht und 40 % kannten die Definition von Spezifität nicht.<sup>16</sup> Fehler in der Interpretation statistischer Daten können durch besseres didaktisches Training im Rahmen des medizinischen Curriculums reduziert werden.<sup>86</sup> Jenny et al. 2018 konnten dies bereits zeigen, da in ihrer Studie bereits ein 90-minütiges Training zu einer durchschnittlichen Erhöhung der Anzahl richtiger Antworten von zuvor 50 % auf 90 % führen konnte.<sup>16</sup> Zudem wäre es ideal, bereits in der Ausbildung damit zu beginnen und Medizinstudenten zu erklären, wie sie selbständig Häufigkeits-Doppelbäume und Häufigkeitsnetze erstellen können. Somit hätten sie eine Herangehensweise bayesianische Probleme zu lösen, die sie für jedes neue Problem der gleichen Art anwenden könnten. Es ist unrealistisch, dass Ärzte im Alltag bedingte Wahrscheinlichkeiten jedes Mal erneut berechnen. Im Nationalen Kompetenzbasierten Lernzielkatalog Medizin (NKLM), der für die deutschen Universitäten gilt, ist für die Studierenden die Ausbildung zur Erstellung von Wahrscheinlichkeits- oder Häufigkeitsdiagrammen bisher nicht vorgesehen. Insbesondere das Bayes Theorem findet keine gesonderte Erwähnung.<sup>66</sup> Im künftigen NKLM sollte die bayesianische Statistik, sowie deren didaktische Aufbereitung mit Hilfe von Häufigkeitsdoppelbäumen und



-netzen inkludiert werden. In Lehrbüchern zur Mathematik in Biowissenschaften und Statistik in Medizin wird der Satz von Bayes zwar eingeführt, jedoch häufig nur unter Verwendung der Formel von Bayes und der Anwendung von Wahrscheinlichkeiten.<sup>2,4</sup> Didaktisch ansprechende und verständlich erklärte Darstellungen der bayesianischen Argumentation haben immer noch Seltenheitswert. Auch bei Internet-Recherchen zu diesem Thema wird rasch offensichtlich, dass gute Statistikkenntnisse und die Fähigkeit, dieses Thema ansprechend zu erklären, nur selten koinzident sind. Solange sich das nicht grundlegend ändert, werden die Widerstände der angehenden und bereits fertig ausgebildeten Mediziner statistischen Inhalten gegenüber nur mit Mühe überwunden werden können.

Der Einfluss des Netzdiagramms sollte in künftigen Forschungsarbeiten genauer untersucht werden, unter anderem da es möglich ist, in einem Netzdiagramm, sowohl die Häufigkeiten in den Knoten als auch die Wahrscheinlichkeiten an den Verzweigungen des Netzdiagramms darzustellen.<sup>25</sup> Auf diese Weise können alle Informationen auf einen Blick dargestellt werden, was zu weniger Fehlern und einer höheren diagnostischen Effizienz führen kann.

Unsere Studie kann dazu beitragen, die Ausbildung im medizinischen Bereich zu verbessern. Bayesianische Probleme sollten basierend auf aktuellen Erkenntnissen der Literatur und auch den Ergebnissen der vorliegenden Studie mit natürlichen Häufigkeiten unterrichtet werden, die in Doppelbäumen oder Netzdiagrammen modelliert werden. Diese Probleme sind schneller zu lösen, weil dieses Schema für die Lernenden leicht zu verstehen ist. Außerdem scheint diese Strategie einprägsamer zu sein als die bayesianische Formel selbst.<sup>32</sup>

## 5. Limitationen und Ausblick

In dieser Studie wurde nicht gezeigt, ob man Medizinstudenten beibringen kann, selbständig Doppelbäume und Häufigkeitsnetze zu zeichnen, um bayesianische Aufgaben zu lösen. Dies sollte in zukünftigen Studien für Netzdiagramme untersucht werden. Wenn Medizinstudenten selbst in der Lage wären, bayesianische Kontexte händisch darzustellen, könnten viele Fehler bei der medizinischen Entscheidungsfindung im Klinikalltag vermieden werden.

Fachgremien sollten diskutieren, ob sie in Leitlinien, Metaanalysen und Reviews bereits Häufigkeitsdoppelbäume und -netze implementieren wollen, um die Präsentation für die Ärzte zu standardisieren. So könnte ein Trainingseffekt entstehen und Mediziner wären mehr darauf vorbereitet bayesianische Kontexte schnell und richtig zu interpretieren. Aufgrund der Überlastung von Ärzten im klinischen Alltag durch enormen bürokratischen Aufwand ist es unrealistisch, dass Ärzte auch noch die Zeit finden, solche Diagramme händisch für jeden neuen klinischen Kontext selbst erstellen. Um diese Diagramme richtig verstehen zu können, macht es durchaus Sinn im Medizinstudium den Studierenden beizubringen solche Diagramme händisch zu skizzieren. Im weiteren Verlauf dann reicht es die bereits dargestellten Diagramme korrekt interpretieren zu können.

In dieser Studie war der medizinische Kontext kein entscheidender Faktor. Zukünftige Studien könnten untersuchen, ob sich die Entscheidungsfähigkeit ändert, wenn die Teilnehmer bereits mit dem Kontext vertraut sind. Die Fragen könnten an das jeweilige Thema angepasst werden, beispielsweise endokrinologische klinische Fälle für Endokrinologen. Außerdem könnten in zukünftigen Studien gemischte, also aus verschiedenen Themengebieten kombinierte Kontexte, verwendet werden, um inhaltliche Verzerrungen zu minimieren. In diesem Zusammenhang könnte auch der Effekt des *cognitive overloading* untersucht werden. Das heißt die Überforderung eine Aufgabe zu lösen, wenn sie zu viele parallele Denkprozesse erfordert.<sup>87</sup> In komplexeren Fällen, wie beispielsweise in der vorliegenden Studie im Fall des Cushing-Syndroms, könnte dies eine mögliche Erklärung von verlängerten Zeiten bis zur Lösung eines bayesianischen Problems sein.

Im Hinblick auf die ökologische Validität, also die Generalisierbarkeit der Ergebnisse auf das Verhalten in der realen Welt, konzentriert sich die vorliegende Studie nur auf die Fähigkeit, medizinische bayesianische Aufgaben zu lösen. Unsere Studie bezieht sich nicht auf die Korrektheit der Entscheidungsfindung in konkreten Fällen von realen Patienten, da dies Thema klinischer Diagnostik ist. Hierfür könnten in Zukunft realistische Übungen anhand von simulierten Patientenfällen durchgeführt. Wichtig ist zu beachten, dass die Interpretation von Bayes Informationen und klinische Diagnostik separat zu trainierende Aspekte medizinischer Entscheidungsfindung sind. Die korrekte und schnelle Interpretation von bayesianischen Diagrammen bedeutet nicht automatisch, dass die klinische Diagnosefähigkeit von Ärzten und Medizinstudierenden besser wird.

Nicht nur im Hinblick auf die Verbesserung der Lehre in der Medizin und der Kompetenz im klinischen Alltag, sondern auch in der Arzt-Patienten-Kommunikation könnte die visuelle Darstellung von Daten in Form von Doppelbäumen oder Netzdiagrammen zu positiven Ergebnissen führen, da bereits gezeigt wurde, dass die visuelle Darstellung von statistischen Informationen eine bessere Arzt-Patienten-Kommunikation fördert.<sup>35</sup> Auch die Patienten könnten davon profitieren, indem sie anhand der Diagramme in der Lage wären die Interpretation von Testergebnissen besser nachvollziehen zu können.

Außerdem gibt es Erkenntnisse über den Einfluss der kognitiven Fähigkeiten von Individuen beim bayesianischen Schlussfolgern. Zu diesen kognitiven Fähigkeiten zählen beispielsweise Denkprozesse und das Arbeitsgedächtnis, welches Teil des Kurzzeitgedächtnisses ist, der für die unmittelbare bewusste Verarbeitung von Wahrnehmungen und Sprache zuständig ist.<sup>88</sup>

Brase und Hill (2015) stellten fest, dass es keinen Einfluss von kognitiven Fähigkeiten auf das Interpretieren von bayesianischen Fragestellungen geben sollte. Yin et al. (2020) fanden jedoch, dass die Interaktion zwischen dem Arbeitsgedächtnis von Menschen und dem Verständnis des Wahrscheinlichkeitsformats bayesianischer Aufgaben signifikant war. Daher wäre es interessant, in zukünftigen Arbeiten zusätzlich Aufgaben zu kognitiven Fähigkeiten und deren Einfluss auf die Lösung von bayesianischen Problemen zu implementieren.

Es scheint auch einen signifikanten Leistungsunterschied beim Lösen von bayesianischen Problemen zwischen Personen mit hohen und niedrigen räumlichen Vorstellungsvermögen zu geben.<sup>89</sup> Für Menschen mit einer hohen räumlichen Vorstellungsfähigkeit ist die Interaktion zwischen den Textangaben der Fallbeschreibungen und den visuellen Darstellungen teilweise ablenkend und überfordernd und führt zu verminderter Leistungsfähigkeit im Vergleich zu den Versionen, in denen ausschließlich Texte gegeben waren.<sup>90</sup> Zukünftige Studien sollten genauer untersuchen, ob Personen mit gutem Vorstellungsvermögen interne Repräsentationen der Daten entwickeln, während sie den Text lesen und sie daher länger brauchen neue präsentierte Grafiken zu interpretieren, da sie Umdenken müssen. Im Vergleich hierzu sollte untersucht werden, ob Menschen ohne räumliches Vorstellungsvermögen dieses Problem nicht haben, da sie keine interne Repräsentation entwickeln und sie daher keinen Konflikt mit vorhandenen Grafiken haben. Zur Prüfung diese Hypothese soll zum einen untersucht werden, ob sich die Versuchspersonen hinsichtlich des räumlichen Vorstellungsvermögens unterscheiden, beispielsweise mittels der in den medizinischen Eignungstests angewandten Tests zur Überprüfung der räumlichen Vorstellungsfähigkeit. Zusätzlich sollten die Studienteilnehmer dann gefragt werden, ob die präsentierten, eventuell ungewohnten Grafiken zu einer *gefühlten* Verzögerung bei der Bearbeitung geführt haben. Diese Ergebnisse könnte man dann mit den tatsächlich gemessenen Verzögerungen und der Richtigkeit in Abhängigkeit vom räumlichen Vorstellungsvermögen der Studienteilnehmer statistisch untersuchen.

Darüber hinaus wurde nicht untersucht, welches Informationsformat (Natürliche Häufigkeiten vs. Wahrscheinlichkeiten) von den Medizinstudenten bevorzugt wurde. In zukünftigen Untersuchungen könnte dies durch eine strukturierte Befragung nach der Studie ermittelt werden, um herauszufinden, was die Studierenden dachten und welche Informationen für sie bei der Lösung der Aufgaben hilfreich waren und welche nicht. Daher kann nicht nachvollzogen werden, ob die Teilnehmer Wahrscheinlichkeiten in natürliche Frequenzen umwandelten.<sup>44</sup> Man könnte auch als Zusatz nach einer eingetippten Antwort direkt in der Aufgabe fragen, wie der Lösungsweg war, um direkt in der Auswertung die Lösungsstrategie beurteilen zu können. Zusätzlich könnte man eine Freitextoption für Kommentare ergänzen, um zu erfragen, was gut war und was abgelenkt hat. So könnte man neue Informationen herausfinden, die den Studienleitern sonst nicht bewusst wären.

### **Fazit für die Praxis**

Wenn Medizinstudenten ein vollständiges Bild der möglichen Wahrscheinlichkeiten oder Häufigkeiten innerhalb einer Visualisierung präsentiert wird, sind sie bemerkenswert gut darin, die richtige Wahrscheinlichkeit oder Häufigkeit aus den in der Visualisierung angegebenen Zahlen auszuwählen. Die Visualisierung natürlicher Häufigkeiten sowohl in Doppelbäumen als auch in Netzdiagrammen hilft Medizinstudenten, bayesianische Aufgaben nicht nur genauer, sondern auch effizienter zu beantworten.

Diese Erkenntnis sollte zum Wohle der Patienten in der medizinischen Ausbildung berücksichtigt werden. Die Vorlesungen in Statistik und Biomathematik sollten mehr Wert auf die didaktische erfolgreiche Präsentation legen als auf den rein mathematischen Formalismus. Dadurch hätten mehr Studenten ein Erfolgserlebnis beim Lernen, die Ergebnisse würden sich ebenso wie die Entscheidungsgeschwindigkeit verbessern und so hätte der Patient auch einen Vorteil, sowohl bei der Diagnose, der Aufklärung als auch bei der Entscheidung für die richtige Therapiealternative.

Zusammenfassend können diese Erkenntnisse den klinischen Alltag von Medizinern und Patienten erleichtern und richtige Diagnosestellungen und Entscheidungsfindungsprozesse verbessern.

## Literaturverzeichnis

1. Tschirk W. *Statistik: Klassisch oder Bayes*. Springer; 2014:113-116.
2. Timischl W. *Mathematische Methoden in Den Biowissenschaften*. Springer; 2016:40-44.
3. Binder K, Krauss S, Schmidmaier R, Braun LT. Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. *Advances in Health Sciences Education*. 2021:1-17.
4. Hilgers R-D, Bauer P, Scheiber V. *Einführung in die medizinische Statistik*. Springer-Verlag; 2007:132-146.
5. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*. 1763(53):370-418.
6. Grünwald R. Bayes Statistik einfach und verständlich erklärt. <https://novostat.com/statistik-blog/bayes-statistik-einfach-und-verstaendlich-erklart.html>. Published 2021. Accessed 06.09.2022.
7. Gallistel CR. Bayes for beginners: probability and likelihood. *APS Observer*. 2015;28.
8. Holt CA, Smith AM. An update on Bayesian updating. *Journal of Economic Behavior & Organization*. 2009;69(2):125-134.
9. Sox HC, Higgins MC, Owens DK. *Medical Decision Making*. Wiley; 2013:61-88.
10. Eddy D. Probabilistic reasoning in clinical medicine: Problems and opportunities. Judgment under uncertainty: Heuristics and biases. Edited by: Kahneman D, Slovic P and Tversky A. 1982. In: Cambridge, England, Cambridge University Press; 1982.
11. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Academic Medicine*. 1998;73(5):538-540.
12. Gigerenzer G. What are natural frequencies? *Bmj*. 2011;343.
13. Wegwarth O, Gigerenzer G. Overdiagnosis and overtreatment: evaluation of what physicians tell their patients about screening harms. *JAMA internal medicine*. 2013;173(22):2086-2088.
14. Diederich S, Feldkamp J, Grußendorf M, Reincke M. *Referenz Endokrinologie und Diabetologie*. Georg Thieme Verlag; 2020:181-182.
15. Bayer S. Satz von Bayes. <https://www.sofatutor.com/mathematik/videos/satz-von-bayes>. Published 2022. Accessed 06.09.2022.
16. Jenny MA, Keller N, Gigerenzer G. Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany. *BMJ open*. 2018;8(8):e020847.
17. Binder K, Krauss S, Bruckmaier G, Marienhagen J. Visualizing the Bayesian 2-test case: The effect of tree diagrams on medical decision making. *PLoS One*. 2018;13(3):e0195029.
18. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *Bmj*. 2009;339.
19. Salerno S, Laghi A, Cantone M-C, Sartori P, Pinto A, Frija G. Overdiagnosis and overimaging: an ethical issue for radiological protection. *La radiologia medica*. 2019;124(8):714-720.
20. Braun LT, Zottmann JM, Adolf C, et al. Representation scaffolds improve diagnostic efficiency in medical students. *Medical education*. 2017;51(11):1118-1126.
21. ALQahtani DA, Rotgans JI, Mamede S, Mahzari MM, Al-Ghamdi GA, Schmidt HG. Factors underlying suboptimal diagnostic performance in physicians under time pressure. *Medical education*. 2018;52(12):1288-1298.
22. Reani M, Davies A, Peek N, Jay C. How do people use information presentation to make decisions in Bayesian reasoning tasks? *International Journal of Human-Computer Studies*. 2018;111:62-77.
23. Ayal S, Beyth-Marom R. The effects of mental steps and compatibility on Bayesian reasoning. *Judgment and Decision Making*. 2014;9(3):226.
24. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. *Psychological review*. 1995;102(4):684.

25. Binder K, Krauss S, Wiesner P. A new visualization for probabilistic situations containing two binary events: the frequency net. *Frontiers in Psychology*. 2020;11:750.
26. Eichler A, Böcherer-Linder K, Vogel M. Different Visualizations Cause Different Strategies When Dealing With Bayesian Situations. *Frontiers in Psychology*. 2020;11.
27. McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological bulletin*. 2017;143(12):1273.
28. Gigerenzer G, Hoffrage U. Overcoming difficulties in Bayesian reasoning: a reply to Lewis and Keren (1999) and Mellers and McGraw (1999). 1999.
29. Kunzelmann AK, Binder K, Fischer MR, Reincke M, Braun LT, Schmidmaier R. Improving Diagnostic Efficiency with Frequency Double-Trees and Frequency Nets in Bayesian Reasoning. *MDM policy & practice*. 2022;7(1):23814683221086623.
30. Fenton N, Neil M. Comparing risks of alternative medical diagnosis using Bayesian arguments. *Journal of Biomedical Informatics*. 2010;43(4):485-495.
31. Krauss S, Weber P, Binder K, Bruckmaier G. Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit–Forschungsdesiderate und einige Antworten. *Journal für Mathematik-Didaktik*. 2020;41(2):485-521.
32. Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *Journal of experimental psychology: general*. 2001;130(3):380.
33. Binder K, Krauss S, Bruckmaier G. Effects of visualizing statistical information – an empirical study on tree diagrams and  $2 \times 2$  tables. *Frontiers in Psychology*. 2015;6(1186).
34. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. In: American Association for the Advancement of Science; 2000.
35. Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science & Medicine*. 2013;83:27-33.
36. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychological science in the public interest*. 2007;8(2):53-96.
37. Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical DNA evidence. *Jurimetrics*. 2002;43:147.
38. Hoffrage U, Hafenbrädl S, Bouquet C. Natural frequencies facilitate diagnostic inferences of managers. *Frontiers in Psychology*. 2015;6:642.
39. Sirota M, Juanchich M, Haggmayer Y. Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic bulletin & review*. 2014;21(1):198-204.
40. Barbey AK, Sloman SA. Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*. 2007;30(3):241-254.
41. Brase GL, Hill WT. Good fences make for good neighbors but bad science: a review of what improves Bayesian reasoning and why. *Frontiers in Psychology*. 2015;6:340.
42. Brase G. Which cognitive individual differences predict good Bayesian reasoning? Concurrent comparisons of underlying abilities. *Memory & Cognition*. 2020:1-14.
43. Hasher L, Zacks RT. Automatic and effortful processes in memory. *Journal of experimental psychology: General*. 1979;108(3):356.
44. Weber P, Binder K, Krauss S. Why can only 24% solve bayesian reasoning problems in natural frequencies: frequency phobia in spite of probability blindness. *Frontiers in psychology*. 2018;9:1833.
45. Kirkwood BR, Sterne JA. *Essential medical statistics*. John Wiley & Sons; 2010.
46. Khan A, Breslav S, Glueck M, Hornbæk K. Benefits of visualization in the mammography problem. *International Journal of Human-Computer Studies*. 2015;83:94-113.
47. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *science*. 2011;333(6048):1393-1400.

48. Starns JJ, Cohen AL, Bosco C, Hirst J. A visualization technique for Bayesian reasoning. *Applied Cognitive Psychology*. 2019;33(2):234-251.
49. Zikmund-Fisher BJ, Witteman HO, Dickson M, et al. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical decision making*. 2014;34(4):443-453.
50. Sirota M, Kostovičová L, Juanchich M. The effect of iconicity of visual displays on statistical reasoning: evidence in favor of the null hypothesis. *Psychonomic Bulletin & Review*. 2014;21(4):961-968.
51. Yamagishi K. Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental psychology*. 2003;50(2):97.
52. Steckelberg A, Balgenorth A, Berger J, Mühlhauser I. Explaining computation of predictive values: 2×2 table versus frequency tree. A randomized controlled trial [ISRCTN74278823]. *BMC medical education*. 2004;4(1):1-8.
53. Friederichs H, Ligges S, Weissenstein A. Using tree diagrams without numerical values in addition to relative numbers improves students' numeracy skills: a randomized study in medical education. *Medical Decision Making*. 2014;34(2):253-257.
54. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC medical informatics and decision making*. 2013;13(2):1-15.
55. Micallef L, Dragicevic P, Fekete J-D. Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE transactions on visualization and computer graphics*. 2012;18(12):2536-2545.
56. Brase GL. The power of representation and interpretation: doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology*. 2014;26(1):81-97.
57. Brase GL. Pictorial representations in statistical reasoning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*. 2008;23(3):369-381.
58. Armstrong B, Spaniol J. Experienced probabilities increase understanding of diagnostic test results in younger and older adults. *Medical Decision Making*. 2017;37(6):670-679.
59. Kurzenhäuser S, Hoffrage U. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Medical Teacher*. 2002;24(5):516-521.
60. Obrecht NA, Anderson B, Schulkin J, Chapman GB. Retrospective Frequency Formats Promote Consistent Experience-Based Bayesian Judgments. *Applied cognitive psychology*. 2012;26(3):436-440.
61. Vonthein R, Burkholder I, Muche R, Rauch G. *Zeig mir mehr Biostatistik!: Mehr Ideen und neues Material für einen guten Biometrie-Unterricht*. Springer; 2017:87-98.
62. Fischer MR, Aulinger B, Baehring T. Computer-based-training (CBT): fallorientiertes lernen am PC mit dem CASUS/ProMediWeb-system. *DMW-Deutsche Medizinische Wochenschrift*. 1999;124(46):1401-1401.
63. Braun LT, Riester A, Oßwald-Kopp A, et al. Toward a diagnostic score in cushing's syndrome. *Frontiers in endocrinology*. 2019;10:766.
64. Bhangu JS, Selberherr A, Brammen L, Scheuba C, Riss P. Efficacy of calcium excretion and calcium/creatinine clearance ratio in the differential diagnosis of familial hypocalciuric hypercalcemia and primary hyperparathyroidism. *Head & neck*. 2019;41(5):1372-1378.
65. Montori VM, Young Jr WF. Use of plasma aldosterone concentration-to-plasma renin activity ratio as a screening test for primary aldosteronism. A systematic review of the literature. *Endocrinology and metabolism clinics of North America*. 2002;31(3):619-632, xi.
66. e.V. MFdBD. National Competence Based Learning Objectives Catalogue Medicine [Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM)]. 2015.
67. Team RC. R: A language and environment for statistical computing. 2013.

68. Rousseeuw PJ, Hubert M. Robust statistics for outlier detection. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*. 2011;1(1):73-79.
69. Wassner C, Krauss S, Martignon L. Muss der Satz von Bayes schwer verständlich sein? *Praxis der Mathematik*. 2002;44(1):12-16.
70. Sirota M, Juanchich M. Role of numeracy and cognitive reflection in Bayesian reasoning with natural frequencies. *Studia Psychologica*. 2011;53(2).
71. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *Jama*. 2007;298(9):1010-1022.
72. Wie lernt der zukünftige Arzt, klug zu entscheiden? [press release]. Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP), 29.11.2019 2019.
73. Universität-Heidelberg. Über den Test für Medizinische Studiengänge (TMS). [https://cip.dmed.uni-heidelberg.de/tms-info/tms-info/index.php?id=ueber\\_den\\_tms](https://cip.dmed.uni-heidelberg.de/tms-info/tms-info/index.php?id=ueber_den_tms). Published 2021. Accessed 19.04.2021, 2021.
74. Daig U, Hermann I. Das Kreuz mit den Kreuzen. Kritik an den Multiple-Choice-Prüfungen- und kein Ende. [The cross with crosses. Criticism of multiple-choice exams - and no end]. In: *Deutsches Ärzteblatt*. Vol 992002:1643-1644.
75. Freiwald T, Salimi M, Khaljani E, Harendza S. Pattern recognition as a concept for multiple-choice questions in a national licensing exam. *BMC medical education*. 2014;14(1):1-6.
76. Bruckmaier G, Binder K, Krauss S, Kufner H-M. An eye-tracking study of statistical reasoning with tree diagrams and  $2 \times 2$  tables. *Frontiers in psychology*. 2019;10:632.
77. Bechtel W. Diagrammatic reasoning. In: *Springer handbook of model-based science*. Springer; 2017:605-618.
78. Brase GL, Hill WT. Adding up to good Bayesian reasoning: Problem format manipulations and individual skill differences. *Journal of Experimental Psychology: General*. 2017;146(4):577.
79. Kellen V, Chan S, Fang X. Improving user performance in conditional probability problems with computer-generated diagrams. Paper presented at: International Conference on Human-Computer Interaction2013.
80. Yin L, Shi Z, Liao Z, Tang T, Xie Y, Peng S. The Effects of Working Memory and Probability Format on Bayesian Reasoning. *Frontiers in Psychology*. 2020;11:863.
81. Lesage E, Navarrete G, De Neys W. Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking & Reasoning*. 2013;19(1):27-53.
82. Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences*. 2000;23(5):645-665.
83. Osterloh F. Ärzten macht ihre Arbeit Spaß [Physicians enjoy their work]. *Deutsches Ärzteblatt*. 2012;109(24):1212-1213.
84. Marewski JN, Gigerenzer G. Heuristic decision making in medicine. *Dialogues in clinical neuroscience*. 2012;14(1):77.
85. Woike JK, Hoffrage U, Martignon L. Integrating and testing natural frequencies, naïve Bayes, and fast-and-frugal trees. *Decision*. 2017;4(4):234.
86. Gigerenzer G. Risk savvy: how to make good decisions. Allen Lane. In: Penguin Group, London; 2014.
87. Kirsh D. A few thoughts on cognitive overload. *Intellectica*. 2000;1(30).
88. Baddeley A. Working memory. *Science*. 1992;255(5044):556-559.
89. Ottley A, Peck EM, Harrison LT, et al. Improving Bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE transactions on visualization and computer graphics*. 2015;22(1):529-538.
90. Mosca A, Ottley A, Chang R. Does interaction improve bayesian reasoning with visualization? *arXiv preprint arXiv:210301701*. 2021.

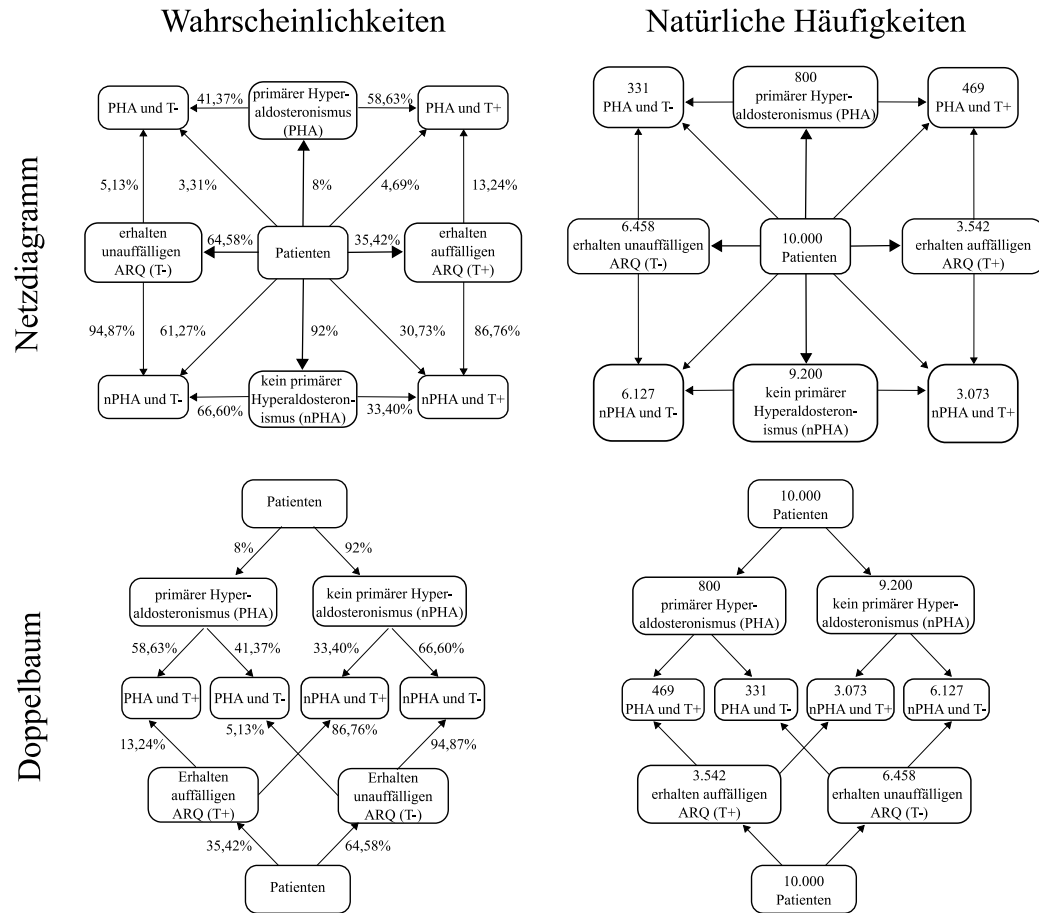


## **Anhang:**

<b>Anhang 1</b> - Fallbeschreibung <i>Primärer Hyperaldosteronismus (PHA)</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). .....	50
<b>Anhang 2</b> - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel <i>Primärer Hyperaldosteronismus</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). .....	51
<b>Anhang 3</b> - Fallbeschreibung <i>Cushing-Syndrom</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). ....	52
<b>Anhang 4</b> - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel <i>Cushing-Syndrom</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). .....	53
<b>Anhang 5</b> - Fallbeschreibung <i>Familiäre hypokalziurische Hyperkalzämie</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). .....	54
<b>Anhang 6</b> - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel <i>Familiäre hypokalziurische Hyperkalzämie</i> . Grafik adaptiert nach Kunzelmann et.al. (2022). .....	55

<b>Fall</b>		<b>Primärer Hyperaldosteronismus</b>	
<b>Informationsformat</b>	<b>Wahrscheinlichkeiten</b>	<b>Häufigkeiten</b>	
<b>Medizinischer Kontext</b>	<p>Stellen Sie sich vor, Sie arbeiten als Arzt in einer endokrinologischen Ambulanz. Hier wird unter anderem der Aldosteron-Renin-Quotient (ARQ) bei Verdacht auf primären Hyperaldosteronismus bestimmt.</p> <p>Sie haben gerade einen Patienten mit einem auffälligen Aldosteron-Renin-Quotienten konsultiert, der zur Abklärung des Verdachts auf primären Hyperaldosteronismus in die Ambulanz gekommen ist. Dieser Patient möchte wissen, wie wahrscheinlich es ist, dass er einen primären Hyperaldosteronismus hat, wenn er einen auffälligen Aldosteron-Renin-Quotienten hat.</p> <p>Für Ihre Antwort stehen Ihnen nur die folgenden Informationen zur Verfügung, die auf einer Stichprobe von Patienten mit einem auffälligen Aldosteron-Renin-Quotienten basieren.</p>		
<b>Visualisierung</b>	<ul style="list-style-type: none"> <li>• Wahrscheinlichkeits-Doppelbaum oder</li> <li>• Wahrscheinlichkeitsnetz-Diagramm (siehe Anhang 2, links)</li> </ul>	<ul style="list-style-type: none"> <li>• Häufigkeits-Doppelbaum oder</li> <li>• Häufigkeitsnetz-Diagramm (siehe Anhang 2, rechts)</li> </ul>	
<b>Fragestellung</b>	Wie hoch ist die Wahrscheinlichkeit, dass ein Patient mit einem auffälligen Aldosteron-Renin-Quotienten tatsächlich einen primären Hyperaldosteronismus hat?	Wie viele Patienten, die einen auffälligen Aldosteron-Renin-Quotienten haben, haben tatsächlich einen primären Hyperaldosteronismus?	
	Antwort: _____	Antwort: ____ von ____	

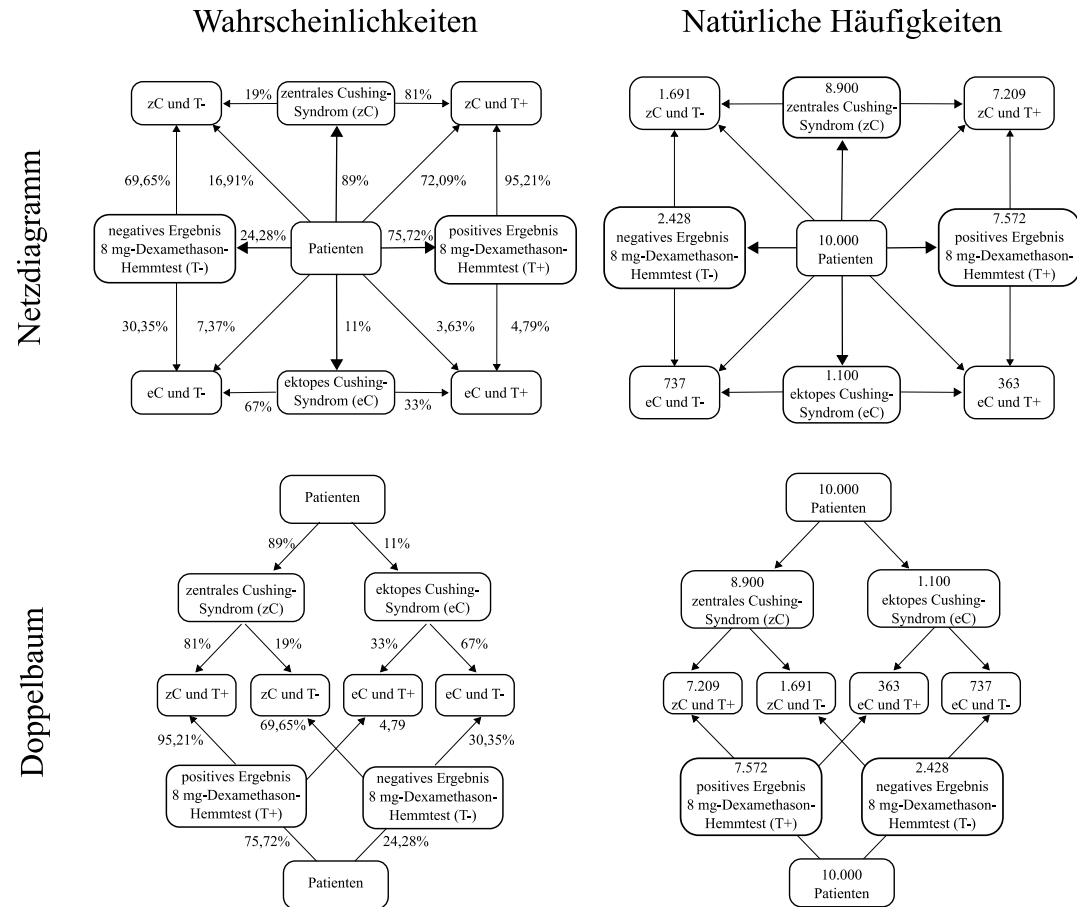
**Anhang 1** - Fallbeschreibung *Primärer Hyperaldosteronismus (PHA)*. Grafik adaptiert nach Kunzelmann et.al. (2022).



**Anhang 2** - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel *Primärer Hyperaldosteronismus*. Grafik adaptiert nach Kunzelmann et.al. (2022).

Fall		Cushing-Syndrom	
Informationsformat	Wahrscheinlichkeiten	Häufigkeiten	
<b>Medizinischer Kontext</b>	<p>Stellen Sie sich vor, Sie arbeiten als Arzt in einer endokrinologischen Ambulanz. Hier werden unter anderem 8 mg Dexamethason-Suppressionstests (DST) zur Differenzierung zwischen ektopem und zentralem Cushing-Syndrom durchgeführt.</p> <p>Sie haben gerade einen Patienten mit einem pathologischen 1 mg Dexamethason-Suppressionstest und erhöhten Serum-ACTH-Spiegeln beraten, der zur ätiologischen Abklärung bei der Neudiagnose eines ACTH-abhängigen Cushing-Syndroms einen 8 mg Dexamethason-Suppressionstest erhalten hat. Dieser Patient möchte wissen, wie wahrscheinlich es ist, dass er ein zentrales Cushing-Syndrom hat, wenn der 8-mg-Dexamethason-Suppressionstest positiv ist.</p> <p>Für Ihre Antwort stehen Ihnen nur die folgenden Informationen zur Verfügung, die auf einer Stichprobe von Patienten mit einem pathologischen 8-mg-Dexamethason-Suppressionstest und erhöhten Serum-ACTH-Werten basieren.</p>		
<b>Visualisierung</b>	<ul style="list-style-type: none"> <li>• Wahrscheinlichkeits-Doppelbaum oder</li> <li>• Wahrscheinlichkeitsnetz-Diagramm (siehe Anhang 4, links)</li> </ul>	<ul style="list-style-type: none"> <li>• Häufigkeits-Doppelbaum oder</li> <li>• Häufigkeitsnetz-Diagramm (siehe Anhang 4, rechts)</li> </ul>	
<b>Fragestellung</b>	Wie hoch ist die Wahrscheinlichkeit, dass diese Person mit einem positiven Ergebnis im 8-mg-Dexamethason-Suppressionstest tatsächlich ein zentrales Cushing-Syndrom hat?	Wie viele Patienten, die ein positives Ergebnis beim 8-Mg-Dexamethason-Suppressionstest haben, haben tatsächlich ein zentrales Cushing-Syndrom?	
	Antwort: _____	Antwort: ____ von ____	

**Anhang 3** - Fallbeschreibung *Cushing-Syndrom*. Grafik adaptiert nach Kunzelmann et.al. (2022).



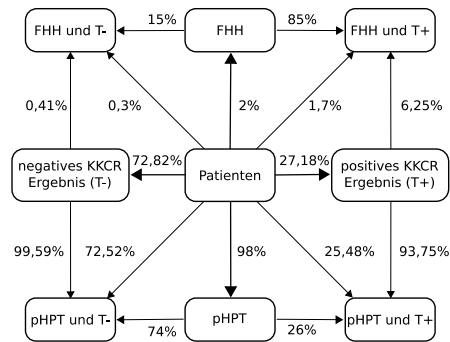
**Anhang 4** - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel *Cushing-Syndrom*. Grafik adaptiert nach Kunzelmann et.al. (2022).

Fall	Familiäre hypokalziurische Hyperkalzämie	
Informationsformat	Wahrscheinlichkeiten	Häufigkeiten
<b>Medizinischer Kontext</b>	<p>Stellen Sie sich vor, Sie arbeiten als Arzt in einer endokrinologischen Ambulanz. Hier werden unter anderem Kalziumbestimmungen im Urin bei Verdacht auf primären Hyperparathyreoidismus (pHPT) oder familiäre hypokalziurische Hyperkalzämie (FHH) durchgeführt.</p> <p>Sie haben gerade einen Patienten mit erhöhten Kalzium- und Parathormonwerten beraten, der sich zur Abklärung des Verdachts auf FHH in der Ambulanz vorgestellt und sein Kalzium-Kreatinin-Clearance-Ratio (KKCR) bestimmen lassen hat. Dieser Patient möchte wissen, wie wahrscheinlich es ist, dass er an einer FHH leidet, wenn das Ergebnis der KKCR positiv ist.</p> <p>Für Ihre Antwort stehen Ihnen nur die folgenden Informationen zur Verfügung, die auf einer Stichprobe von Patienten mit erhöhtem Kalzium und Nebenschilddrüsenhormon basieren.</p>	
<b>Visualisierung</b>	<ul style="list-style-type: none"> <li>• Wahrscheinlichkeits-Doppelbaum oder</li> <li>• Wahrscheinlichkeitsnetz-Diagramm (siehe Anhang 6, links)</li> </ul>	<ul style="list-style-type: none"> <li>• Häufigkeits-Doppelbaum oder</li> <li>• Häufigkeitsnetz-Diagramm (siehe Anhang 6, rechts)</li> </ul>
<b>Fragestellung</b>	Wie hoch ist die Wahrscheinlichkeit, dass diese Person mit einem positiven Ergebnis in der KKCR tatsächlich FHH hat?	Wie viele Patienten, die ein positives Ergebnis in der KKCR haben, haben tatsächlich FHH?
	Antwort: _____	Antwort: ____ von ____

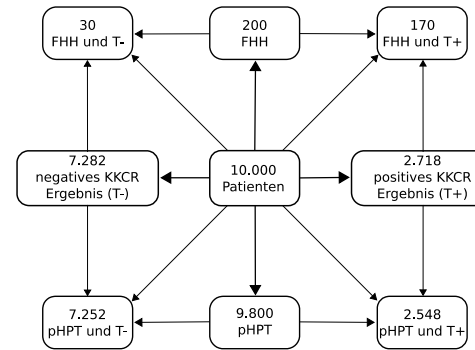
**Anhang 5** - Fallbeschreibung *Familiäre hypokalziurische Hyperkalzämie*. Grafik adaptiert nach Kunzelmann et.al. (2022).

Netzdiagramm

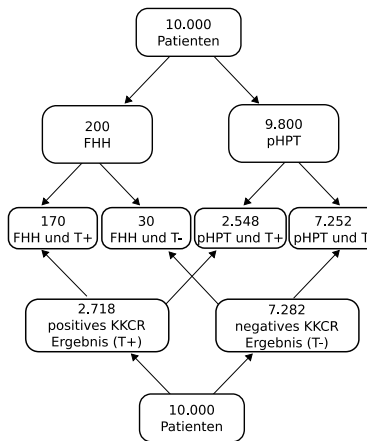
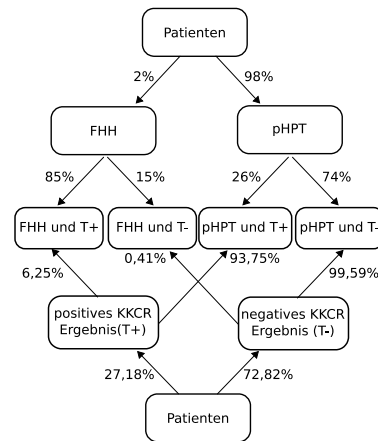
### Wahrscheinlichkeiten



### Natürliche Häufigkeiten



Doppelbaum



**Anhang 6** - Netzdiagramme und Doppelbäume in beiden Informationsformaten (Häufigkeiten vs. Wahrscheinlichkeiten) für das Beispiel *Familiäre hypokalziurische Hyperkalzämie*. Grafik adaptiert nach Kunzelmann et.al. (2022).

## **Danksagung**

Zuerst möchte ich meinem Doktorvater Herrn Prof. Dr. Ralf Schmidmaier für die Überlassung dieses spannenden Themas danken und die Möglichkeit unter seiner Leitung daran zu arbeiten.

Ganz besonderen Dank möchte ich meiner Betreuerin Frau Dr. med. Leah Braun aussprechen, die mich mit ihrer freundlichen und liebenswerten Art zu jeder Zeit hilfreich unterstützt hat. Sie hat mich in jeder Phase dieser Promotion professionell und persönlich begleitet und mich durch das große Vertrauen und die viele Freiheit bei der Durchführung dieser Arbeit stets motiviert. Sie stand mir zu jeder Zeit mit viel Geduld zur Seite und ist ein Vorbild für mich.

Besonderer Dank gilt Frau Prof. Dr. Karin Binder, sie gewährte mir mit großem Engagement während der Durchführung der gesamten Arbeit außerordentlich kompetente und bedeutende Unterstützung. Insbesondere inspirierte sie mich mit ihrer sehr persönlichen und positiven Art, mich mit mathematischen Problemen auseinanderzusetzen. Die Gespräche mit ihr hatten einen sehr positiven Einfluss auf mein wissenschaftliches Interesse. Ihre Teilhabe an meiner Promotion werde ich nie vergessen.

Zudem möchte ich Herrn Prof. Dr. Martin Fischer für seine Anregungen und Diskussionsbeiträge danken.

Mein Dank gilt außerdem allen Probanden, die durch ihre Teilnahme diese Studie ermöglicht haben.

Ganz besonderer Dank gilt meinen Eltern, die mir nicht nur meine Studienzeit ermöglicht haben und mir immer helfen meine Träume zu verwirklichen, sondern mir zudem stets das Gefühl geben, mich bei Allem zu unterstützen. Sie geben jederzeit das Beste für mich, stehen mir stets bedingungslos zur Seite und teilen ihre Lebenserfahrung mit mir.

Ich möchte meiner Schwester danken, meiner wertvollen Begleiterin in allen Lebensphasen, die für mich immer eine große Inspiration ist.

Außerdem danke ich meinen Großeltern und meinem Großonkel für die Unterstützung, Geduld und Ratschläge.

Mein persönlicher Dank gilt besonders auch meinen Freunden, die mich immer motiviert und aufgemuntert haben. Weiterer Dank gilt außerdem meinen Kollegen des Spitalzentrums Oberwallis in Visp, die mich in der letzten Phase der Fertigstellung dieser Arbeit stets unterstützt und für den nötigen Ausgleich gesorgt haben.



## Affidavit



### Eidesstattliche Versicherung

Kunzelmann, Alexandra Katharina

---

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

### ***Einfluss von Bayesian Reasoning auf die Diagnoserichtigkeit und - schnelligkeit bei Medizinstudierenden***

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 19.12.2022

---

Ort, Datum

Kunzelmann, Alexandra Katharina

---

Unterschrift Doktorandin bzw. Doktorand