

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



Bayesianische versus diagnostische ärztliche Informationsvermittlung an Patient*innen

Der Einfluss von Richtung der statistischen Information
und Visualisierung

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

vorgelegt von
Sarah Frederike Brose
aus
Filderstadt

2025

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Erstes Gutachten: Prof. Dr. med. Ralf Schmidmaier, MME

Zweites Gutachten: Prof. Dr. Martin Fischer

Drittes Gutachten: Prof. Dr. Julia Seiderer-Nack

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 17.02.2025

Inhaltsverzeichnis

Abbildungsverzeichnis.....	VI
Tabellenverzeichnis.....	VIII
1 Einleitung.....	1
1.1 Ärztliche Kommunikation mit Patient*innen.....	1
1.2 Diagnosen und statistische Informationen in der Medizin	2
1.2.1 Satz von Bayes.....	3
1.2.2 Diagnosefehler und deren Folgen.....	5
1.3 Verbesserungsmöglichkeiten des Verständnisses statistischer Informationen.....	7
1.3.1 Informationsformat.....	7
1.3.2 Informationsrichtung	8
1.3.3 Informationsvisualisierung	11
2 Zielsetzung	14
3 Material und Methoden.....	15
3.1 Modellentwicklung.....	15
3.1.1 Die vier Kontexte	15
3.1.2 Die variablen Parameter bei der Kommunikation statistischer Informationen	15
3.1.3 Pilotphase und die verschiedenen Versionen	16
3.2 Studienablauf	19
3.2.1 Untersuchungskollektiv	20
3.2.2 Bearbeitung der Casus-Fälle	21

3.3	Statistische Methoden.....	23
3.4	Ethikvotum	23
4	Ergebnisse.....	24
4.1	Soziodemografische Daten	24
4.2	Angaben zum Grundwissen	25
4.3	Ergebnisrichtigkeit.....	27
4.4	Lerneffekte	30
4.4.1	Verbesserung der Ergebnisrichtigkeit im Verlauf der Studie	30
4.4.2	Verbesserung der Ergebnisrichtigkeit durch mehrmaliges Anschauen	32
4.5	Ergebnisgeschwindigkeit.....	36
4.5.1	Bei falschen Antworten	38
4.5.2	Bei richtigen Antworten.....	38
4.5.3	Die diagnostische Effizienz.....	39
4.6	Subjektive Schwierigkeitsbeurteilung	40
5	Diskussion	42
5.1	Zusammenfassung.....	42
5.2	Diskussion der Methoden.....	42
5.3	Diskussion der Ergebnisse.....	45
5.3.1	Die Proband*innen als potenzielle Patient*innen.....	45
5.3.2	Der Einsatz diagnostischer Informationen ist von Vorteil.....	46
5.3.3	Das Häufigkeitsnetz verringert die Schwierigkeit von Bayesianischen Informationen	47

5.3.4	Mit Hilfe guter Kommunikationen können Patient*innen statistische Informationen angemessen verstehen	49
5.4	Limitationen.....	50
5.5	Ausblick.....	50
6	Literaturverzeichnis	52
7	Anhang.....	59
8	Zusammenfassung	72
9	Danksagung	74
10	Publikationsliste.....	75
11	Affidavit.....	76
12	Übereinstimmungserklärung.....	77

Abbildungsverzeichnis

Abbildung 1: Bayesianischer vs. diagnostischer Baum	10
Abbildung 2: Häufigkeitsnetz	13
Abbildung 3: Häufigkeitsnetz zum Fallkontext „Schilddrüsenkarzinom“	19
Abbildung 4: Der Studienablauf	21
Abbildung 5: Höchster Bildungsabschluss	24
Abbildung 6: Aktueller Beruf	25
Abbildung 7: Medizinisches Grundwissen	26
Abbildung 8: Mathematisches Grundwissen	26
Abbildung 9: Ergebnisrichtigkeit nach einmaligem Anschauen, wobei alle Aufgaben ausgewertet wurden	27
Abbildung 10: Ergebnisrichtigkeit nach einmaligem Anschauen, wenn nur die erste Aufgabe ausgewertet wurde	30
Abbildung 11: Ergebnisrichtigkeit nach mehrmaligem Anschauen, wobei alle Aufgaben ausgewertet wurden	32
Abbildung 12: Durchschnittliche Häufigkeit der Betrachtung des Videos	35
Abbildung 13: Subjektiv leichtester Fall	40
Abbildung 14: Subjektiv schwierigster Fall	40
Abbildung 15: Häufigkeitsnetz zum Fallkontext „Primärer Hyperaldosteronismus“	60
Abbildung 16: Häufigkeitsnetz zum Fallkontext „Familiäre hypokalziurische Hyperkalzämie“	62
Abbildung 17: Häufigkeitsnetz zum Fallkontext „Morbus Cushing“	64
Abbildung 18: Flyer der Studie	65

Abbildung 19: Informationsschrift und Einwilligungserklärung (Seite 1/6)	66
Abbildung 20: Informationsschrift und Einwilligungserklärung (Seite 2/6)	67
Abbildung 21: Informationsschrift und Einwilligungserklärung (Seite 3/6)	68
Abbildung 22: Informationsschrift und Einwilligungserklärung (Seite 4/6)	69
Abbildung 23: Informationsschrift und Einwilligungserklärung (Seite 5/6)	70
Abbildung 24: Informationsschrift und Einwilligungserklärung (Seite 6/6)	71

Tabellenverzeichnis

Tabelle 1: Bayesianische vs. diagnostische Informationen	9
Tabelle 2: Das 2×2×4 Design	17
Tabelle 3: Die Aufgabenversionen zum Fallkontext „Schilddrüsenkarzinom“ ...	18
Tabelle 4: Parameterschätzungen des verallgemeinerten linearen gemischten Modells zur Vorhersage der Leistung der Proband*innen, nachdem sie das Video einmal angeschaut haben	29
Tabelle 5: Parameterschätzungen des verallgemeinerten linearen gemischten Modells zur Vorhersage der Leistung der Proband*innen, nachdem sie das Video mehrmalig angeschaut haben	34
Tabelle 6: Deskriptive Ergebnisse über alle Kontexte und über alle gelösten Aufgaben, nachdem das Video einmal angesehen wurde	37
Tabelle 7: Die Aufgabenversionen zum Fallkontext "Primärer Hyperaldosteronismus"	59
Tabelle 8: Die Aufgabenversionen zum Fallkontext "Familiäre hypokalziurische Hyperkalzämie"	61
Tabelle 9: Die Aufgabenversionen zum Fallkontext "Morbus Cushing"	63

Teile dieser Arbeit wurden bereits als Paper in der internationalen Fachzeitschrift PLOS ONE veröffentlicht:

Brose SF, Binder K, Fischer MR, Reincke M, Braun LT, Schmidmaier R. Bayesian versus diagnostic information in physician-patient communication: Effects of direction of statistical information and presentation of visualization. PLOS ONE. 2023;18(6):e0283947.

- DOI: 10.1371/journal.pone.0283947
- Siehe (1)

Dabei haben die verschiedenen Autoren im Paper folgende Rollen übernommen:

- Sarah Frederike Brose: Datenkuration, Untersuchung, Visualisierung, Schreiben – Originalentwurf
- Karin Binder: Datenkuration, formale Analyse, Methodik, Überwachung, Visualisierung, Schreiben – Überprüfung und Bearbeitung
- Martin R. Fischer: Supervision, Schreiben – Überprüfen und Bearbeiten
- Martin Reincke: Supervision, Schreiben – Überprüfen und Bearbeiten
- Leah T. Braun: Konzeptualisierung, Datenkuration, Finanzierungsbeschaffung, Methodik, Supervision, Schreiben – Überprüfung und Bearbeitung
- Ralf Schmidmaier: Konzeptualisierung, Supervision, Schreiben – Überprüfung und Bearbeitung

Für ähnliche Teile zwischen dieser Arbeit und dem Paper ist auf dieselbe Rollenverteilung wie im Paper hinzuweisen. Neue Teile dieser Arbeit wurden selbstständig von Sarah Frederike Brose verfasst.

Aus dem Paper übernommene Abbildungen und Tabellen wurden in dieser Arbeit entsprechend gekennzeichnet.

1 Einleitung

1.1 Ärztliche Kommunikation mit Patient*innen

Kommunikation (von lateinisch *communicatio* = Mitteilung) ist der Austausch oder die Übertragung von Informationen. Diese stellt besonders in der Medizin eine zentrale Kompetenz dar, die aus dem klinischen Alltag nicht wegzudenken ist. Bereits am ersten Arbeitstag muss sie von jungen Ärzt*innen im Rahmen der Visite beherrscht werden (2). Ob in der Inneren Medizin, Chirurgie oder Psychiatrie – überall ist eine gute Kommunikation gefragt (3), um eine gemeinsame Entscheidungsfindung mit Patient*innen zu ermöglichen. Nur wer das beherrscht, kann klinische Tätigkeiten selbstständig ausführen (4) und so die Erwartungen als Ärzt*in erfüllen (5). Häufig ist die Realität aber eine andere: Immer wieder wird von der mangelnden Fähigkeit von Studierenden im praktischen Jahr berichtet, eine strukturierte und gezielte Visite durchzuführen (5). Doch wie kommt es zu diesen Missständen? Ein Problem lässt sich in der Lehre von Medizinstudierenden erkennen. So wurde lange Zeit der Lehre kommunikativer und praktischer Fähigkeiten im Vergleich zu Forschung und Krankenversorgung wenig Aufmerksamkeit geschenkt (6), wodurch es an großen Fortschritten auf dem Gebiet der Ausbildungsforschung mangelte (7). Um die Weiterentwicklung und Modernisierung der medizinischen Curricula zu erleichtern, wurde ab 2009 ein nationaler kompetenzbasierter Lernzielkatalog für Medizin erarbeitet (8) und konnte im Jahr 2015 verabschiedet werden (9). Darin sind die Kompetenzen aufgeführt, die ein*e Medizinstudent*in nach seinem*ihrem erfolgreichen Abschluss des Studiums erworben haben sollte (9). Die ärztliche Gesprächsführung wird darin als übergeordnete Kompetenz genannt. Die Umsetzung des kompetenzbasierten Curriculums an den einzelnen Universitäten und die dadurch erhofften Verbesserungen bleiben abzuwarten (10). An der Ludwig-Maximilians-Universität München wurde das Medizinische Curriculum München implementiert (6). Darin ist die Kompetenz der Kommunikation im Rahmen des Kursus „Patient*innen-orientierte Kommunikation“ bereits vorhanden, es gilt aber diese weiter systematisch zu fördern (4). Dass dies auch im Interesse der Studierenden ist, wurde in einer Umfrage deutlich, in der Studierende die ärztliche Gesprächsführung als eins der drei wichtigsten Lernziele in einer Famulatur angaben (11). Grundsätzlich

Einleitung

stehen dabei drei verschiedene Kommunikationsstile zur Auswahl: die ärztlichzentrierte Kommunikation, die patient*innenzentrierte Kommunikation und die patient*innenzentrierte Kommunikation mit Bedürfnisorientierung. Diese haben unterschiedliche Auswirkungen auf die Bewertung der Ärzt*innen durch die Patient*innen und auf die Informationsverarbeitung. Bei der patient*innenzentrierten Kommunikation werden die Ärzt*innen im Gegensatz zur ärztlichzentrierten Kommunikation als einfühlsamer und sozial sowie fachlich kompetenter wahrgenommen (12). Sie ist effektiv bei der Vermittlung von Sachinformationen (12) und wird häufig bevorzugt (13). Die patient*innenzentrierte Kommunikation mit Bedürfnisorientierung ist mit einem geringeren Wissenserwerb verbunden (12), eignet sich aber besonders für Patient*innen, die sich ihrer persönlichen Bedürfnisse bewusst sind (14) und wenn eine gute Beziehung aufgebaut werden soll (12). Patient*innen mit einer hohen Angstneigung hingegen bevorzugen häufig eine ärztlichzentrierte Kommunikation (15). Ebenso beeinflussen wissenschaftliche und emotionale Formulierungen die ärztliche Kommunikation mit Patient*innen in unterschiedlicher Weise (16). Der angemessene Einsatz bestimmter Kommunikationsstile hängt also von den spezifischen Patient*innengruppen und den Beratungszielen ab. Darüber hinaus hat sich gezeigt, dass Patient*innen dazu neigen, dem ärztlichen Rat zu folgen und sich gegen ihre eigenen, ursprünglichen Behandlungspräferenzen zu entscheiden (17-19). In Situationen, in denen die Patient*innenpräferenzen von großer Bedeutung sind, sollten Empfehlungen mit Vorsicht eingesetzt werden (20). Wenn die ursprüngliche Präferenz bereits mit der ärztlichen Empfehlung übereinstimmt, steigt die Entscheidungssicherheit und Zufriedenheit des Patient*innen (19). Eine gute Kommunikation gibt Ärzt*innen die Möglichkeit, ihr Wissen über eine bestimmte Krankheit sowie diagnostische und therapeutische Überlegungen an ihre Patient*innen weiterzugeben; es ist wichtig, dass diese Informationen von Patient*innen richtig verstanden werden.

1.2 Diagnosen und statistische Informationen in der Medizin

Ein Szenario, in dem gute Kommunikation von besonders hoher Bedeutung ist, ist das Stellen und Mitteilen von Diagnosen. Jedoch gibt es hier sowie bei der Mitteilung von Chancen und Risiken von Diagnoseverfahren oder medizinischen

Einleitung

Behandlungen oftmals große Defizite (21, 22). Im klinischen Alltag müssen Ärzt*innen häufig Patient*innen erklären, wie groß die Wahrscheinlichkeit für eine bestimmte Krankheit ist. Grund hierfür ist meist ein positives Testergebnis. Viele Menschen gehen davon aus, dass wenn ein Test bei fast jedem* jeder erkrankten Patienten* Patientin positiv ist, ein positiv getestete*r Patient*in auch mit einer hohen Wahrscheinlichkeit die Krankheit hat. Dass dies aber häufig nicht der Fall ist, kann mit dem Satz von Bayes erklärt werden.

1.2.1 Satz von Bayes

Der Satz von Bayes, auch Bayes-Theorem genannt, wurde im 18. Jahrhundert vom englischen Mathematiker Thomas Bayes formuliert. Die Wahrscheinlichkeit für das Eintreten eines Ereignisses, wenn eine bestimmte Bedingung vorliegt, nennt man bedingte Wahrscheinlichkeit. Der Satz von Bayes hilft, diese zu berechnen. Dabei macht man sich einer Rückwärtsinduktion zunutze, mit deren Hilfe man Schlussfolgerungen umkehren kann. Geht man von zwei Ereignissen A und B aus, kann man für jedes dieser Ereignisse ihre Wahrscheinlichkeit angeben. Diese formuliert man als:

$P(A)$ A-priori-Wahrscheinlichkeit des Ereignisses A

$P(B)$ A-priori-Wahrscheinlichkeit des Ereignisses B

A-priori-Wahrscheinlichkeiten sind Anfangswahrscheinlichkeiten, die angenommen werden müssen. Unter ihrer Annahme lassen sich bedingte Wahrscheinlichkeiten beschreiben. Im Fall der Ereignisse A und B lauten die bedingten Wahrscheinlichkeiten:

$P(A|B)$ Bedingte Wahrscheinlichkeit, dass das Ergebnis A eintritt, wenn das Ereignis B bereits eingetreten ist

$P(B|A)$ Bedingte Wahrscheinlichkeit, dass das Ergebnis B eintritt, wenn das Ereignis A bereits eingetreten ist

Einleitung

Im medizinischen Kontext ermöglicht dies Ärzt*innen zu erläutern, wie wahrscheinlich bestimmte Testergebnisse und Krankheiten sind. Im Folgenden werden die Ereignisse A und B für die Ereignisse des Vorhandenseins eines positiven Testergebnisses und des Vorliegens einer Krankheit definiert. Ebenso lassen sich entsprechende Gegenwahrscheinlichkeiten definieren, die für das Gegenteil der Ereignisse stehen und die durch einen Strich über dem Buchstaben veranschaulicht werden:

$P(A^*)$ A-priori-Wahrscheinlichkeit für das Vorliegen einer Krankheit

$P(\bar{A}^*)$ A-priori-Wahrscheinlichkeit für das Vorliegen keiner Krankheit

$P(B^*)$ A-priori-Wahrscheinlichkeit für ein positives Testergebnis

$P(\bar{B}^*)$ A-priori-Wahrscheinlichkeit für ein negatives Testergebnis

Daraus lassen sich die entsprechenden bedingten Wahrscheinlichkeiten definieren. Beispiele hierfür sind:

$P(A^* | B^*)$ Bedingte Wahrscheinlichkeit, dass eine Krankheit vorliegt, wenn das Testergebnis positiv ist

$P(\bar{A}^* | \bar{B}^*)$ Bedingte Wahrscheinlichkeit, dass keine Krankheit vorliegt, wenn das Testergebnis negativ ist

$P(B^* | A^*)$ Bedingte Wahrscheinlichkeit, dass das Testergebnis positiv ist, wenn eine Krankheit vorliegt

$P(\bar{B}^* | \bar{A}^*)$ Bedingte Wahrscheinlichkeit, dass das Testergebnis negativ ist, wenn keine Krankheit vorliegt

In der medizinischen Statistik haben manche dieser Wahrscheinlichkeiten besonders große Bedeutung. Deshalb wurden ihnen eigene Fachbegriffe

Einleitung

gewidmet. So wird $P(A^*)$ als die Prävalenz, $P(B^*|A^*)$ als die Sensitivität, $P(\bar{B}^*|\bar{A}^*)$ als die Spezifität, $P(A^*|B^*)$ als der positive prädiktive Wert und $P(\bar{A}^*|\bar{B}^*)$ als der negative prädiktive Wert bezeichnet. Die Sensitivität und die Spezifität sind Testparameter, die in vielen großen Studien untersucht werden. Auch über die Prävalenz verschiedener Krankheiten liegen zahlreiche Daten vor. Im klinischen Alltag müssen Testergebnisse und deren Bedeutung häufig an Patient*innen vermittelt werden, wobei für Patient*innen nach einem positiven Testergebnis der positive prädiktive Wert meist von größter Bedeutung ist. Dieser sagt aus, wie wahrscheinlich es ist, dass ein positives Testergebnis auch wirklich bedeutet, dass der*die Patient*in eine bestimmte Krankheit hat.

Mithilfe der Formel von Bayes und Informationen zu Sensitivität, Falsch-Positiv-Rate (1-Spezifität) und Prävalenz (23) kann der positive prädiktive Wert berechnet werden:

$$\text{Positiver prädiktiver Wert} = \frac{\text{Sensitivität} \cdot \text{Prävalenz}}{\text{Sensitivität} \cdot \text{Prävalenz} + (1 - \text{Spezifität}) \cdot (1 - \text{Prävalenz})}$$

Bei Krankheiten mit hoher Prävalenz können die Sensitivität und der positive prädiktive Wert ähnliche Werte annehmen. Sensitivität = positiver prädiktiver Wert gilt aber nur, wenn Tests die gleiche Sensitivität und Spezifität haben und die Prävalenz = 0,5 ist. Bei seltenen Krankheiten (niedrige Prävalenz) unterscheidet sich der positive prädiktive Wert meist stark von der Sensitivität. Je kleiner die Prävalenz, desto mehr sinkt der positive prädiktive Wert in Relation zur Sensitivität (24).

1.2.2 Diagnosefehler und deren Folgen

Die Formel von Bayes ist ein Werkzeug in der medizinischen Entscheidungsfindung (25), führt aber oft zu kognitiven Überforderungen oder wird komplett ignoriert (26). Dieser „statistische Analphabetismus“ macht sich in der gesamten Gesellschaft bemerkbar (27). Wahrscheinlichkeiten sind bekannt dafür, schwierig an Laien kommunizierbar zu sein (27). So überschätzt ein Großteil der Patient*innen ihr Erkrankungsrisiko, wie beispielsweise Frauen ihr Brustkrebsrisiko (28). Studierenden und medizinischem Personal gelingt es nicht, den positiven prädiktiven Wert eines HIV-Tests (29-31) zu folgern. Und selbst Ärzt*innen, zu deren alltäglichen Aufgabe es gehört, Patient*innen die Bedeutung

Einleitung

eines positiven Testergebnisses erklären zu können, verwechseln die Sensitivität mit dem positiven prädiktiven Wert (32) und können somit nicht den richtigen positiven prädiktiven Wert angeben (33-35). Dadurch wird die Illusion vermittelt, dass falsch positive Ergebnisse nicht auftreten und dass ein positives HIV-Testergebnis bedeutet, dass der*die Patient*in sicherlich infiziert ist (22). Frühere Studien konnten jedoch zeigen, dass Schulungen zu Bayesianischen Aufgaben (ohne direkte Schulung der Kommunikationsfähigkeiten; (36, 37)) sich lohnen und Medizinstudierende danach besser zwischen adäquater und inadäquater Kommunikation unterscheiden konnten.

Die Herkunft von Diagnosefehler ist multifaktoriell, wobei systemische als auch kognitive Faktoren eine Rolle spielen (38). Dabei überwiegt häufig das menschliche Versagen (39). In Untersuchungen zeigte sich, dass fehlerhafte Diagnoseprozesse von Studierenden vor allem auf fehlende Diagnosefähigkeiten und fehlendes Wissen zurückzuführen sind (40-43). Jedoch sind kognitive Aktionen für eine richtige Falllösung wichtiger als diagnostisches Wissen (44, 45). Besonders bei falsch implizierten Diagnosen treten kognitive Bias auf und führen so zu Fehlern (46).

Auch der im klinischen Alltag verspürte Zeitdruck beeinflusst die Diagnoserichtigkeit der Ärzt*innen negativ (47). Deutschland hat europaweit die kürzesten Zeiten für ärztliche Informationsgespräche (48). Das bewusste Kurzhalten der Gespräche scheint jedoch unnötig zu sein und kann schädliche Konsequenzen für den Austausch zwischen Ärzt*innen und Patient*innen haben (48).

Solche Diagnosefehler sind häufiger als gedacht. Das „Institute of Medicine“ in den USA warnt, dass fast alle Menschen mindestens einen Diagnosefehler in ihrem Leben erfahren werden (49). Häufig resultieren daraus Schäden für die Patient*innen, die bis zu ihrem Tod reichen können (50). In der Krebsvorsorge können durch falsches Verständnis statistischer Informationen Überdiagnosen auftreten, welche wiederum mit Überbehandlungen wie Operationen, Chemotherapien und Bestrahlungen einhergehen können (51). Neben den einzelnen Patient*innen (4) leidet darunter auch die Gesellschaft durch beispielsweise steigende Kosten im Gesundheitssystem.

Es bedarf Möglichkeiten der Kommunikation, die ein fehlerfreies Verständnis statistischer Kommunikationen gewährleisten, um zukünftig Diagnosefehler und deren Folgen vermeiden zu können. Einige dieser Möglichkeiten sollen in dieser Studie untersucht werden.

1.3 Verbesserungsmöglichkeiten des Verständnisses statistischer Informationen

1.3.1 Informationsformat

Statistische Informationen können mithilfe verschiedener Formate angegeben werden. Das bekannteste, die Wahrscheinlichkeiten, wird bereits Schüler*innen ab jungem Alter gelehrt. Weniger verbreitet ist das Konzept der natürlichen Häufigkeiten, welches aus der Kognitionspsychologie stammt. Die Umsetzung als natürliche Häufigkeiten erfolgt, indem eine Wahrscheinlichkeit (z.B. 10%) durch zwei natürliche Zahlen (z.B. 10 von 100) ersetzt wird. Besonders bei Bayesianischen Aufgaben scheinen natürliche Häufigkeiten von Vorteil zu sein. In Studien zeigte sich, dass die Anzahl richtig gelöster Aufgaben mit natürlichen Häufigkeiten höher war als jene mit Wahrscheinlichkeiten (52-57). Dieser positive Effekt auf die Lösungsrichtigkeit scheint dabei unabhängig von den gewählten Proband*innen zu sein. Sowohl bei Studierenden (58-61) als auch bei Ärzt*innen (34) zeigten sich mit natürlichen Häufigkeiten bessere Leistungen. Der Profit durch dieses Informationsformats geht über die Richtigkeit hinaus: Auch die Zeit, die Proband*innen zum Bearbeiten der Aufgaben und Fällen ihrer Urteile benötigen, hat abgenommen (34, 35, 60, 62). Daraus lässt sich schlussfolgern, dass auch die Effizienz bei Aufgaben mit natürlichen Häufigkeiten höher ist, was auf Seite der Ärzt*innen zu einer besseren Kommunikation und auf Seite der Patient*innen zu einem besseren Verständnis führt (34, 35). Das Informationsformat sollte dabei stets in Verbindung mit der jeweilig verwendeten Visualisierung, auf die an späterer Stelle noch genauer eingegangen wird, gesehen werden. Es hat sich beispielsweise gezeigt, dass Bayesianisches Denken nur bei Baumdiagrammen und Vierfeldertafeln mit natürlichen Häufigkeiten unterstützt wird und nicht bei Visualisierungen mit Wahrscheinlichkeiten (63). Doch die Unbekanntheit der natürlichen Häufigkeiten birgt auch Gefahren. So neigen viele Proband*innen dazu, natürliche

Häufigkeiten nicht für ihre Berechnungen zu verwenden, sondern sie stattdessen in komplizierte Wahrscheinlichkeiten zurückzuwandeln, wodurch Fehler entstehen können (64). Trotzdem ist mittlerweile unumstritten, dass natürliche Häufigkeiten Bayesianisches Denken mehr als ursprünglich angenommen erleichtern (58) und auf Grundlage des aktuellsten Forschungsstands ihre Verwendung anstelle von Wahrscheinlichkeiten empfohlen wird (21, 59, 65). Dementsprechend wurden in der vorliegenden Studie nur natürliche Häufigkeiten verwendet.

1.3.2 Informationsrichtung

Unabhängig vom Informationsformat können statistische Informationen in zwei verschiedene Richtungen übermittelt werden, im Folgenden als Informationsrichtungen bezeichnet. Diese Möglichkeiten der Informationsrichtung sind Bayesianische oder diagnostische Informationen. Sowohl im Medizinstudium als auch in der klinischen Praxis kommen meistens Bayesianische Informationen zum Einsatz (siehe Tabelle 1, links): der Anteil der Personen mit einer Krankheit (Prävalenz), der Anteil der Personen mit auffälligen/positiven Tests bei Vorliegen einer Krankheit (Sensitivität) und der Anteil der Personen mit auffälligen/positiven Tests bei Nichtvorliegen einer Krankheit (Falsch-Positiv-Rate). Andererseits ist es möglich, statistische Informationen in umgekehrter Richtung zu liefern, die nachfolgend als diagnostische Informationen bezeichnet werden (siehe Tabelle 1, rechte Seite). Die diagnostischen Informationen umfassen den Anteil der Personen mit einem auffälligen Test, den Anteil der Personen, die bei einem auffälligen/positiven Test tatsächlich erkrankt sind (positiver prädiktiver Wert), und den Anteil der Personen, die trotz eines unauffälligen Tests erkrankt sind (Falsch-Negativ-Rate).

Tabelle 1: Bayesianische vs. diagnostische Informationen

Bayesianische Informationen	Diagnostische Informationen
<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 50 Patient*innen ein Schilddrüsenkarzinom. • Von diesen 50 Patient*innen mit diagnostiziertem Schilddrüsenkarzinom haben 20 Patient*innen ein auffälliges Sonographie-Ergebnis. • Andererseits haben von 950 Patient*innen, die kein Schilddrüsenkarzinom haben, trotzdem 110 Patient*innen eine auffällige Sonographie. 	<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 130 Patient*innen ein auffälliges Sonographie-Ergebnis. • Von diesen 130 Patient*innen mit auffälligem Sonographie-Ergebnis haben 20 Patient*innen tatsächlich ein Schilddrüsenkarzinom. • Andererseits haben von 870 Patient*innen mit einer unauffälligen Sonographie trotzdem 30 Patient*innen ein Schilddrüsenkarzinom.

Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Möchte man Bayesianische Informationen grafisch in einem Baumdiagramm darstellen, sähe das wie Tabelle 1 aus. Für Patient*innen steht meist die Frage im Vordergrund, wie wahrscheinlich sie bei einem auffälligen sonographischen Befund tatsächlich Schilddrüsenkrebs haben. Um diese anhand von Bayesianischen Informationen zu beantworten, muss man zunächst die Anzahl der Patient*innen mit einem auffälligen sonographischen Befund berechnen: $20 + 110 = 130$ Patient*innen. Die Anzahl der Patient*innen, die sowohl einen auffälligen Befund als auch Krebs haben, steht direkt in der Aufgabenstellung: 20 Patient*innen. Die Frage kann also wie folgt richtig beantwortet werden: 20 von 130 Patient*innen mit einem auffälligen sonographischen Befund haben Schilddrüsenkrebs. Werden diagnostische Informationen visuell dargestellt, ergibt sich der sogenannten "diagnostischen Baum" (siehe Abbildung 1; (66, 67)). Hier kann die Frage nach dem positiven prädiktiven Wert direkt beantwortet werden, ohne dass gerechnet bzw. der Satz von Bayes angewendet werden muss. Die Antwort bleibt dieselbe wie im ersten Beispiel: 20 von 130 Patient*innen mit einem auffälligen sonographischen Befund haben Schilddrüsenkrebs. Jedoch sind diagnostische Angaben nur dann möglich, wenn alle hierfür benötigten Häufigkeiten im Vorfeld bekannt sind. Wie groß der tatsächliche Nutzen im medizinischen Alltag einer diagnostischen verglichen mit einer Bayesianischen Informationsrichtung wäre, ist bisher kaum bekannt.

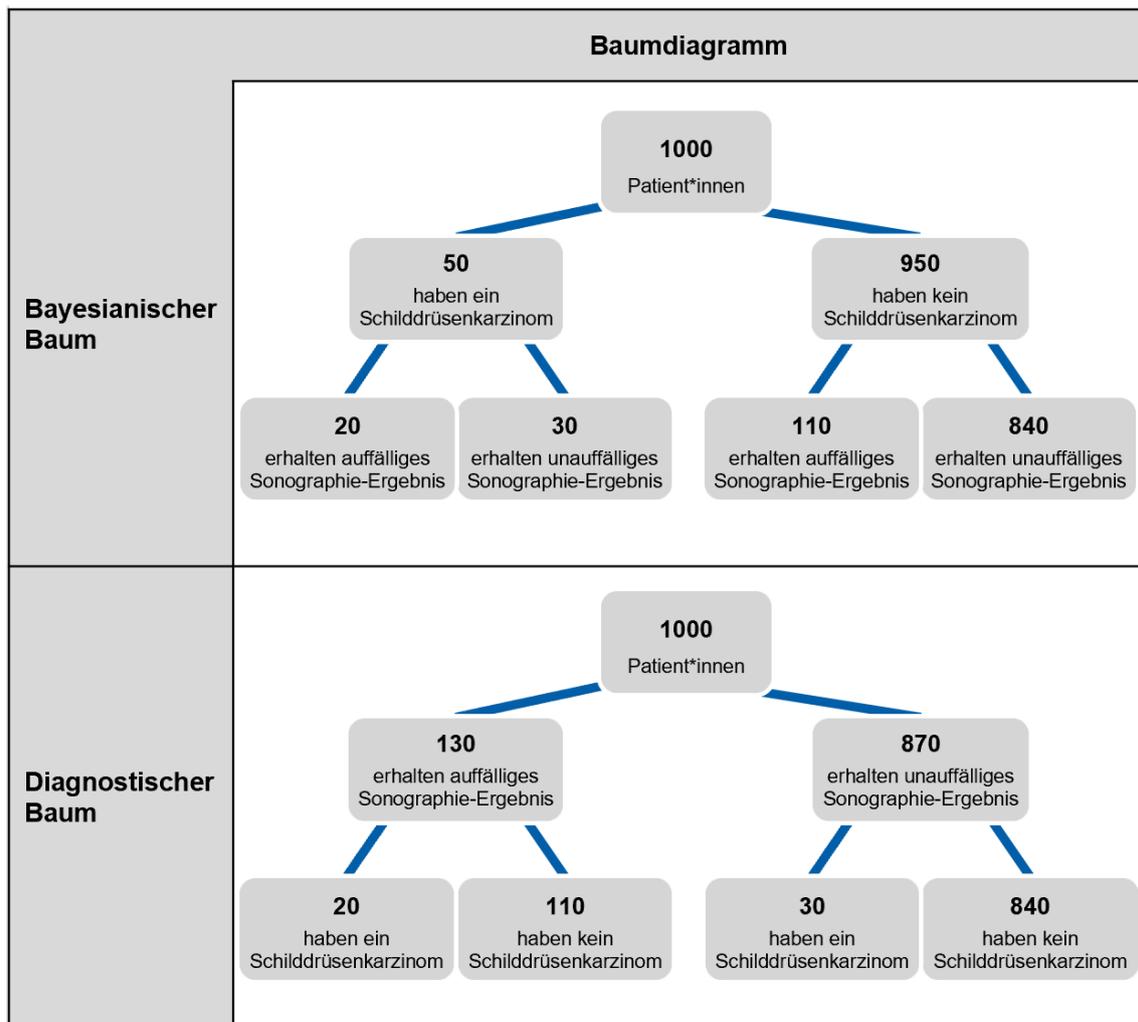


Abbildung 1: Bayesianischer vs. diagnostischer Baum

Die gezeigten statistischen Informationen entsprechen denen aus Tabelle 1. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

In einer beeindruckenden Studie zur Kommunikation in HIV-Beratungsstellen haben Prinz et al. (30) gezeigt, dass etwa die Hälfte der Ärzt*innen in der Lage ist, Sensitivität und Spezifität richtig zu benennen, aber kaum jemand kann den Patient*innen den positiven prädiktiven Wert nennen. In der vorliegenden Studie liegt der Schwerpunkt auf den Patient*innen, und es wird der folgenden Frage nachgegangen: Gibt es einen Unterschied im Verständnis der Patient*innen (in Bezug auf den positiven Vorhersagewert), wenn Ärzt*innen mit Bayesianische Informationen kommunizieren, im Gegensatz dazu, wenn diagnostische Informationen kommuniziert werden?

1.3.3 Informationsvisualisierung

Es gibt verschiedene Ansätze, wie man Diagnosefehler Bayesianischer Aufgaben versuchen kann zu reduzieren. Einer davon sind Visualisierungen. Ihr Einsatz führt zu einer Erhöhung der Lösungsraten (54, 68, 69) und auch zu einer Verringerung der für die Aufgabe benötigten Zeit (35, 62). Hierbei steht eine Reihe von Visualisierungen zur Verfügung, deren Eignung bereits in zahlreichen Studien überprüft wurde. Die bekanntesten Möglichkeiten stellen Vierfeldertafeln und Baumdiagramme (siehe Abbildung 1) dar. Diese Visualisierungen zeichnen sich besonders durch ihre Einfachheit aus, sie jederzeit schnell nur mit Papier und Stift darstellen zu können (61). Sowohl für Baumdiagramme (55) als auch für Vierfeldertafeln (61) konnte in Studien eine Steigerung der Lösungsrichtigkeit nachgewiesen werden. Teilweise zeigte sich dieser Vorteil aber nur bei der Präsentation mit natürlichen Häufigkeiten, bei Wahrscheinlichkeiten waren sie nur eingeschränkt hilfreich (59). Durch Baumdiagramme wurde zusätzlich die Bearbeitungszeit reduziert (60) und somit die Effizienz optimiert (35). Durch den zusätzlichen Schritt des Markierens der für die angeforderte Schlussfolgerung relevanten Zweige, wird ein zusätzlicher Beitrag zum Bayesianischen Verständnis geleistet (55). Die Struktur beider Visualisierungen führt zu einer Reduktion Bayesianischer Fehler, auch wenn es bei Aufgaben von Vierfeldertafeln andere sind als bei Aufgaben mit Baumdiagrammen (57). Aber es haben sich auch Schwierigkeiten bei der Verwendung von Informationsvisualisierungen gezeigt. So können Baumdiagramm manche Proband*innen daran hindern, den richtigen Nenner zu finden, wodurch wiederum ein falscher Zähler ausgewählt wird (70).

Eine Weiterentwicklung des Baumdiagramms, die sich in Studien positiv auf die Leistung auswirkte, ist der Doppelbaum (61). Mittels dieser grafischen Darstellung lassen sich sowohl Wahrscheinlichkeiten als auch Häufigkeiten abbilden. Jedoch muss dafür wiederum in Kauf genommen werden, dass Schnittwahrscheinlichkeiten fehlen und sich kreuzende Äste und doppelte Knoten auftreten (61).

Im Folgenden soll noch der Forschungsstand seltenerer angewandeter Visualisierungsmöglichkeiten aufgezeigt werden. Eine davon ist das Einheitsquadrat. Dieses zeichnet sich durch eine gute Effizienz aus, welche auf

Einleitung

einen Zugewinn prozeduralen und konzeptionellen Wissens in Bezug auf die Bayes-Regel zurückzuführen ist (71). Der Nutzen von Piktogramm ist bislang nicht eindeutig geklärt. Während in manchen Untersuchungen hiermit gute Ergebnisse erzielt werden konnten (72, 73), zeigte sich in anderen dadurch ein schlechtes Verständnis bedingter Wahrscheinlichkeiten, wie sie in Bayesianischen Aufgaben erfragt werden (74). Ebenso werden Venn-Diagramme subjektiv als schwierig zu interpretieren angesehen, obwohl objektiv ihre Lösungsrate der von Baumdiagrammen gleicht (74). Das Roulette-Wheel-Diagramm scheint wenig dabei zu helfen, Diagnosefehler Bayesianischer Aufgaben zu vermeiden und ist Piktogrammen unterlegen (72).

Das Häufigkeitsnetz (siehe Abbildung 2), welches in dieser Studie zum Einsatz kam, stellt eine neue Visualisierungsmöglichkeit für Wahrscheinlichkeiten dar. Es wurde aus dem (Doppel-)Baumdiagramm und der Vierfeldertafel entwickelt, indem es alle Vorteile dieser Visualisierungen in sich vereint: Sowohl alle 4 Randwahrscheinlichkeiten (siehe dicke Linien in Abbildung 2), alle 4 Schnittwahrscheinlichkeiten (siehe gestrichelte Linien in Abbildung 2) als auch alle 8 bedingten Wahrscheinlichkeiten (siehe dünne Linien in Abbildung 2) können hiermit visualisiert werden. Randwahrscheinlichkeiten sind Wahrscheinlichkeiten für ein Ereignis, während Schnittwahrscheinlichkeiten Wahrscheinlichkeiten dafür sind, dass zwei dieser Ereignisse gleichzeitig eintreten. Bedingte Wahrscheinlichkeiten geben die Wahrscheinlichkeit an, dass ein Ereignis auftritt, unter der Bedingung, dass ein anderes Ereignis bereits eingetreten ist. Ebenso ist es auch möglich, natürliche Häufigkeiten (siehe x in Abbildung 2) und Wahrscheinlichkeiten (siehe y in Abbildung 2) zeitgleich darzustellen. Diesen Zahlen können ohne Verrechnungen direkt abgelesen werden und vereinfachen somit Bayesianische Aufgaben. Ein weiterer Vorteil gegenüber Baumdiagrammen ist, dass beide „Leserichtungen“, also Bayesianische Informationen und diagnostische Informationen, gleichermaßen ersichtlich sind. Im Mittelpunkt des Häufigkeitsnetzes kann die Stichprobengröße angegeben werden. Diese kann in medizinischen Fällen die Gesamtzahl aller Patient*innen sein, auf die sich alle anderen Häufigkeiten beziehen. Erste Untersuchungen haben gezeigt, dass diese Darstellung die Leistungsfähigkeit gegenüber einer Erklärung mittels eines reinen Textes verbessert (61).

Einleitung

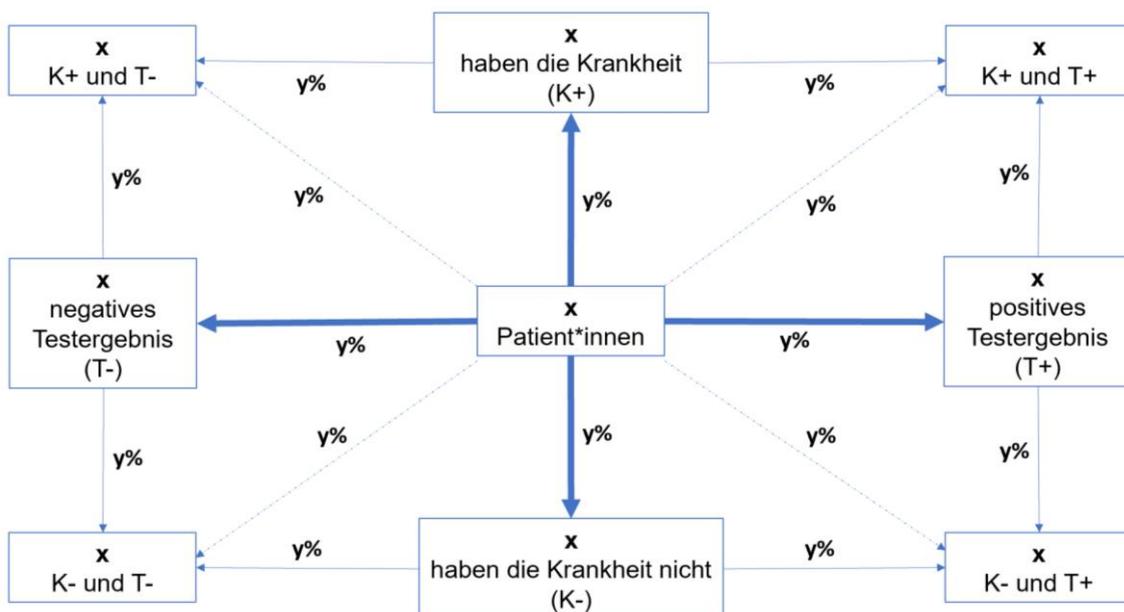


Abbildung 2: Häufigkeitsnetz

Das Verständnis rein auditiver Kommunikation Bayesianischer Informationen ist bislang kaum untersucht. Meist wurden den Visualisierungen Aufgaben mit Fallerkklärungen in Textform gegenübergestellt.

2 Zielsetzung

In ärztlichen Gesprächen mit Patient*innen sollte sowohl auf das, was gesagt wird, als auch darauf, wie es gesagt wird, Wert gelegt werden. Ziel dieser Studie war es, die Richtigkeit und Schnelligkeit des Verständnisses statistischer Informationen von Patient*innen zu verbessern. Hierzu wurde eine Studie entwickelt, um die Auswirkung verschiedener Parameter auf die Kommunikation zu überprüfen. Bislang wurden verschiedene visuelle Erklärungen (Texte oder Grafiken) miteinander verglichen, aber auditive Erklärungen von Testergebnissen, wie sie im ärztlichen Gespräch häufig erfolgen, noch kaum beleuchtet. Ebenso wurde noch nicht untersucht, wie groß der Vorteil einer diagnostischen Informationsrichtung anstelle der üblichen Bayesianischen sein könnte. Im Fokus dieser Untersuchung standen zwei Fragestellungen:

- Wie stark unterscheiden sich die Verständnisrichtigkeit und -schnelligkeit zwischen einer Bayesianischen und einer diagnostischen Informationsrichtung?
- Wie sehr beeinflusst eine Informationsvisualisierung mittels eines Häufigkeitnetzes zusätzlich zu einer auditiven Erklärung das Verständnis verschiedener statistischer Informationen im Gegensatz zu einer rein auditiven Erklärung?

Zu vermuten war, dass eine diagnostische Informationsrichtung gegenüber der Bayesianischen die Effizienz der Kommunikation verbessert. Die visuell-auditive Erklärung ist wahrscheinlich vor allem bei den Bayesianischen Aufgaben von Vorteil. Bei den diagnostischen Aufgaben war ein Nutzen unklar und somit interessant zu testen, da es sowohl Berichte über eine Reduzierung (75) als auch Erhöhung der kognitiven Belastung (76) gibt. Mittels der von den Proband*innen bearbeiteten Aufgaben konnte dies systematisch getestet werden. Dies geschah vor dem Hintergrund, dass ein richtiges Verständnis statistischer Informationen auf Seite der Patient*innen zu weniger Unklarheiten und auf Seite der Krankenhäuser zu beispielsweise geringeren Kosten durch wiederholte Testungen führen kann. Langfristig sollte hiermit auch ein Beitrag zu einem besseren medizinischen Curriculum und so einer besseren Ausbildung der Studierenden und Ärzt*innen geleistet werden.

3 Material und Methoden

3.1 Modellentwicklung

3.1.1 Die vier Kontexte

Es wurden vier verschiedene Kontexte aus dem endokrinologischen Fachbereich gewählt und hieraus vier verschiedene Aufgaben entwickelt. Die Kontexte stellten realitätsnahe, klinische Fälle dar, die sich mit verschiedenen Krankheiten und den dazugehörigen Tests zur Diagnose befassten. Die entsprechenden Häufigkeiten wurden der aktuellen Literatur entnommen. Der erste Kontext befasste sich mit einem auffälligen bzw. unauffälligen Sonographie-Ergebnis in Zusammenhang mit der Diagnose eines bzw. keines Schilddrüsenkarzinoms (77). Beim zweiten Kontext wurden Häufigkeiten eines auffälligen bzw. unauffälligen Aldosteron-Renin-Quotienten und einem bzw. keinem primären Hyperaldosteronismus gegeben (78). Im dritten Fall wurde gezeigt, inwiefern ein positives bzw. negatives Ergebnis in der Calcium-Kreatinin-Ratio mit der Diagnose einer familiären hypokalziurischen Hyperkalzämie bzw. eines primären Hyperparathyreoidismus zusammenhängt (79). Im vierten Kontext wurde ein positives bzw. negatives Ergebnis im 8-mg-Dexamethason-Hemmtest mit einem Morbus Cushing bzw. ektope Cushing-Syndrom in Verbindung gebracht (80).

3.1.2 Die variablen Parameter bei der Kommunikation statistischer Informationen

Im Vorfeld wurden Parameter ausgearbeitet, welche sich auf das Verständnis statistischer Informationen auswirken könnten. Dabei knüpfte die Studie an vorherigen Studien zur Kommunikation statistischer Informationen an und baute darauf auf (35, 55, 59, 60, 63). So ließen sich zwei interessante Parameter identifizieren: Der erste Parameter war die Informationsvisualisierung mit Häufigkeitsnetz vs. ohne Visualisierung. Im ärztlichen Gespräch mit Patient*innen kommen im Alltag häufig keine Visualisierungen zum Einsatz. Ob diese zu einer besseren Kommunikation beitragen können, konnte so systematisch untersucht werden. Als Visualisierung wurde in dieser Studie das Häufigkeitsnetz gewählt, eine neue Visualisierungsmöglichkeit, die dabei helfen kann, Wahrscheinlichkeiten besser zu verstehen (61). Der zweite Parameter, die

Informationsrichtung, konnte Bayesianisch oder diagnostisch sein. Patient*innen werden häufig mit Bayesianischen Angaben konfrontiert, also mit der Prävalenz, Sensitivität und Falsch-Positiv-Rate. Diagnostische Informationen wären der Anteil der Personen mit positivem/auffälligem Befund, der positive prädiktive Wert und die Falsch-Negativ-Rate. Als Informationsformat wurden stets die natürlichen Häufigkeiten gewählt und Wahrscheinlichkeiten nicht in das Modell aufgenommen. Vorangegangene Studien haben gezeigt, dass natürliche Häufigkeiten fast immer besser verstanden werden als Wahrscheinlichkeiten (55, 59, 60, 65). Somit war es wahrscheinlich, dass auch dieses Mal Häufigkeiten von Vorteil sind.

3.1.3 Pilotphase und die verschiedenen Versionen

Anhand dieses Modells erfolgte eine explorative Pilotphase mit 5 Probeproband*innen. Diese nahmen freiwillig daran teil und erhielten keine Aufwandsentschädigung. Es wurde darauf geachtet, dass verschiedene Altersklassen, Geschlechter und Berufsgruppen vorhanden waren, um ein möglichst objektives, realitätsnahes Testergebnis zu erhalten. Die Probeproband*innen bearbeiteten 4 Aufgaben. Dabei entsprachen der Ablauf und die Materialien denen der späteren Studie. Zum Lösen der Fälle wurde eine angemessene Zeit (30-45min) benötigt. Es zeigten sich Unterschiede in der Lösungsrichtigkeit zwischen den Versionen. Aufgrund dieser voraussichtlich guten und differenzierten Auswertbarkeit der dabei zu prüfenden Fragestellungen wurde dieses Modell für die Studie festgelegt. Zusätzlich wurde das Layout verbessert.

So ergab sich als finales Modell folgendes $2 \times 2 \times 4$ Design (siehe Tabelle 2) mit 16 Versionen:

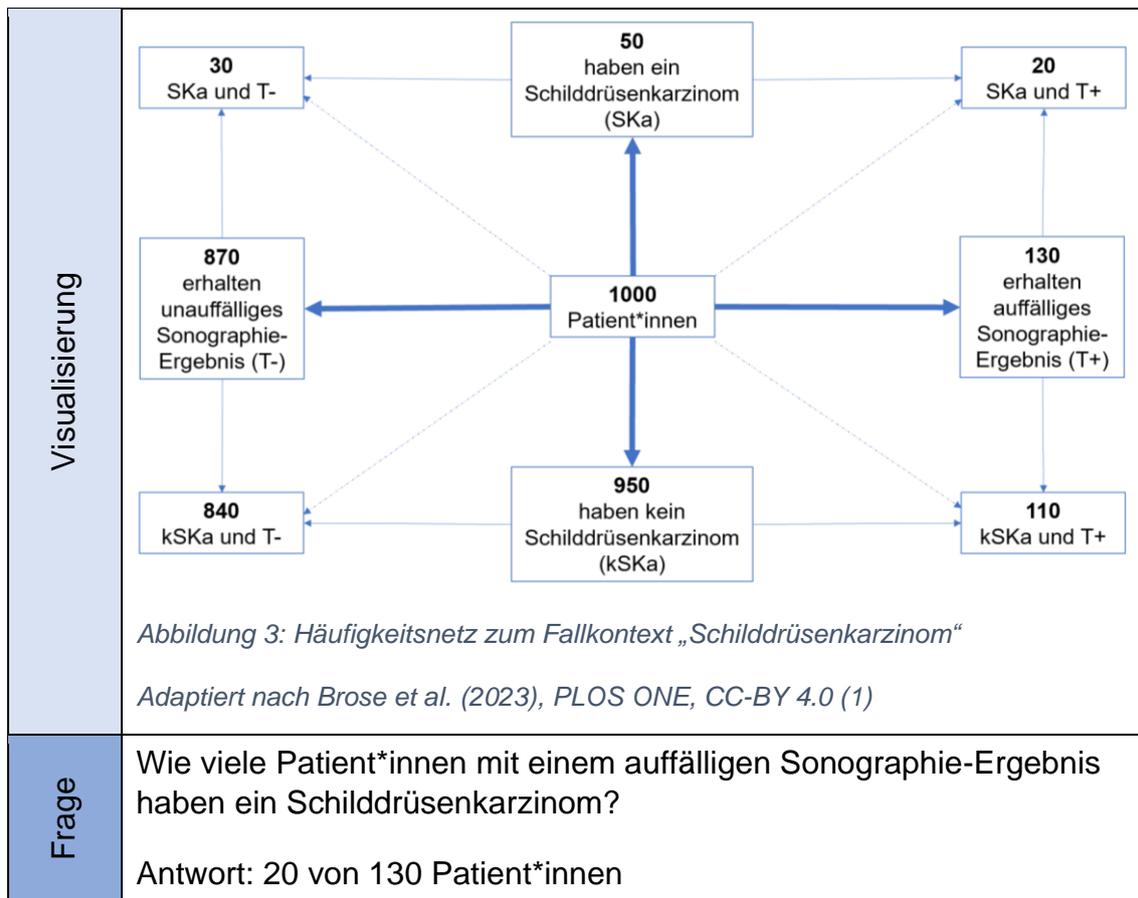
Tabelle 2: Das 2 x 2 x 4 Design

Information- richtung	Informations- visualisierung	Kontext
Bayesianisch	Ohne Visualisierung	Schilddrüsenkarzinom
		Primärer Hyperaldosteronismus
		Morbus Cushing
		Familiäre hypokalziurische Hyperkalzämie
	Häufigkeitsnetz	Schilddrüsenkarzinom
		Primärer Hyperaldosteronismus
		Morbus Cushing
		Familiäre hypokalziurische Hyperkalzämie
Diagnostisch	Ohne Visualisierung	Schilddrüsenkarzinom
		Primärer Hyperaldosteronismus
		Morbus Cushing
		Familiäre hypokalziurische Hyperkalzämie
	Häufigkeitsnetz	Schilddrüsenkarzinom
		Primärer Hyperaldosteronismus
		Morbus Cushing
		Familiäre hypokalziurische Hyperkalzämie

Die Versionen unterschieden sich in zwei verschiedenen Informationsrichtungen (Bayesianisch vs. diagnostisch), zwei verschiedenen Informationsvisualisierungen (ohne Visualisierung vs. mit Häufigkeitsnetz) und vier verschiedenen Kontexten (Schilddrüsenkarzinom, primärer Hyperaldosteronismus, Morbus Cushing und familiäre hypokalziurische Hyperkalzämie, wobei die Kontexte jedoch in der Studie nicht vorrangig von Interesse waren). Hierdurch gab es zu jedem Kontext vier verschiedene Aufgabenversionen. Deren Zuordnung zu den klinischen Kontexten variierte zwischen den Proband*innen und wurde zufällig zugeteilt. Nachfolgende Tabelle 3 zeigt beispielhaft die Aufgabenversionen zum Kontext des Schilddrüsenkarzinoms mit dem dazugehörigen Häufigkeitsnetz (siehe Abbildung 3). Die Aufgabenversionen und Häufigkeitsnetze der anderen Kontexte sind im Anhang zu finden (Tabelle 7-9 sowie Abbildung 15-17).

Tabelle 3: Die Aufgabenversionen zum Fallkontext „Schilddrüsenkarzinom“

	Version 1	Version 2	Version 3	Version 4
	Diagnostisch		Bayesianisch	
Einleitung	<p>Eben wurde bei Ihnen eine Sonographie der Schilddrüse durchgeführt, um den Verdacht eines Schilddrüsenkarzinoms erhärten oder verwerfen zu können. Jetzt möchten wir Ihre Untersuchungsergebnisse gemeinsam besprechen.</p> <p>Leider muss ich Ihnen mitteilen, dass sich in der Untersuchung bei Ihnen ein auffälliges Sonographie-Ergebnis gezeigt hat. Ich möchte Ihnen nun erklären, was ein auffälliger Untersuchungsbefund genau bedeutet.</p> <p>(In diesem Häufigkeitsnetz wurden Patient*innen auf das Vorliegen eines auffälligen Sonographie-Ergebnisses sowie eines Schilddrüsenkarzinoms untersucht. Die beiden Kriterien einer auffälligen bzw. unauffälligen Sonographie und von einem bzw. keinem Schilddrüsenkarzinoms sind hierbei sowohl einzeln als auch in Kombination visualisiert.)</p>			
Informationsrichtung	<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 130 Patient*innen ein auffälliges Sonographie-Ergebnis. • Von diesen 130 Patient*innen mit auffälligem Sonographie-Ergebnis haben 20 Patient*innen tatsächlich ein Schilddrüsenkarzinom. • Andererseits haben von 870 Patient*innen mit einer unauffälligen Sonographie trotzdem 30 Patient*innen ein Schilddrüsenkarzinom. 		<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 50 Patient*innen ein Schilddrüsenkarzinom. • Von diesen 50 Patient*innen mit diagnostiziertem Schilddrüsenkarzinom haben 20 Patient*innen ein auffälliges Sonographie-Ergebnis. • Andererseits haben von 950 Patient*innen, die kein Schilddrüsenkarzinom haben, trotzdem 110 Patient*innen eine auffällige Sonographie. 	
Informations-visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung



Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

3.2 Studienablauf

Zur Durchführung der Studie wurde die Online-Lernplattform Casus (<https://lmu.casus.net>) genutzt. Diese ermöglicht das multimediale, fallorientierte Lernen medizinischer Fallbeispiele (81). Es wurden acht verschiedene Varianten der Studie mit je vier zu bearbeitenden Fällen erstellt, wobei in jeder Variante jeder Kontext und jede Aufgabenversion (also zwei Bayesianische Aufgaben und zwei diagnostische Aufgaben, jeweils einmal mit und einmal ohne Visualisierung) einmal vertreten waren. Sie unterschieden sich jedoch hinsichtlich der Reihenfolge der Kontexte und Versionen, die dabei systematisch variiert wurden. So waren jede Aufgabenversion und jeder Kontext zweimal an erster Stelle, zweimal an zweiter Stelle usw. Dadurch sollte vermieden werden, dass die Ergebnisse von möglichen Reihenfolgeeffekten (d.h. Lerneffekten aber auch Ermüdungseffekten) oder Vorkenntnissen der Krankheiten beeinflusst wurden. Durch eine vorgegebene Bearbeitungsreihenfolge konnte dies zudem kontrolliert werden. Zudem wurde auf eine gleichmäßige Verteilung der Aufgabenvarianten geachtet. Für alle Fälle wurden Videos gedreht, in denen ein*e potenzielle*r

Ärzt*in zu den Proband*innen sprach. Die Proband*innen sollten sich in die Lage eines*einer potenziellen Patienten*Patientin hineinversetzen. Der*die Ärzt*in leitete hierbei den klinischen Kontext zunächst ein und teilte dem*der Patienten*Patientin sein*ihr Untersuchungsergebnis mit. Das Untersuchungsergebnis enthielt statistische Informationen. Bei der Hälfte der Videos wurde zusätzlich die Visualisierung eingeblendet und so die Zahlen anhand eines Häufigkeitsnetzes visualisiert erläutert. Entsprechend dem 2×2×4 Design gab es 16 verschiedene Videos. Diese sind unter folgendem Link zu finden: <https://www.doi.org/10.17605/OSF.IO/FZNWG>

3.2.1 Untersuchungskollektiv

Als Untersuchungskollektiv wurden 110 Proband*innen herangezogen, welche potenzielle Patient*innen darstellen sollten. Die Teilnahme der Proband*innen an der Studie war freiwillig sowie anonym und wurde mit einem Proband*innenhonorar in Höhe von zehn Euro entlohnt. Die Finanzierung der Studie erfolgte Drittmittel-gefördert. Als Proband*innen konnten alle Personen ab einem Alter von mindestens 18 Jahren eingeschlossen werden. Auch wenn Patient*innen in Kliniken oft älter sind, gibt's es auch jüngere Personen, die selbst krank sind oder ältere erkrankte Angehörige zu Untersuchungen begleiten. Darüber hinaus gab es keine weiteren Ein- oder Ausschlusskriterien. Im klinischen Alltag reicht das Patient*innenspektrum von Laien bis hin zu Experten und jeder kann potenziell Patient*in sein, sodass auch ein beruflicher Hintergrund im mathematischen oder medizinischen Bereich keine Einschränkung war. Von der gezielten Auswahl von Proband*innen, die tatsächlich an den jeweiligen Krankheiten leiden, wurde abgesehen, da sie dadurch stark abgelenkt und in ihren Antworten emotional beeinflusst werden könnten. Eine Inhomogenität der Proband*innen bezüglich Alter, Geschlecht, höchstem Bildungsabschluss und aktuellem Beruf stellte kein Problem dar und war sogar erwünscht, um die Vielfalt der Patient*innen aus dem realen klinischen Alltag zu repräsentieren. Vorkenntnisse waren nicht erforderlich und lediglich der Titel der Studie „Kommunikation statistischer Informationen an Patient*innen“ wurde im Vorfeld mitgeteilt. Zur Teilnahme, welche innerhalb des Studienzeitraums jederzeit zeitlich flexibel von Zuhause durchführbar war, benötigten die Proband*innen lediglich einen Computer mit Internetzugang. Die Proband*innen wurden mithilfe

einer Studienausschreibung rekrutiert. Diese wurde über Social Media verbreitet und in München vor Kliniken, Universitäten und Bibliotheken aufgehängt. Da dadurch besonders medizinisches Personal, Studierende und jüngere Menschen angesprochen wurden, ist von einem gewissen Bias auszugehen. Die Vergabe der Teilnahmeplätze erfolgte nach dem Prinzip „Wer zuerst kommt, mahlt zuerst“. Der Studienzeitraum erstreckte sich von August 2021 bis Juni 2022.

3.2.2 Bearbeitung der Casus-Fälle

Die interessierten Proband*innen wurden gebeten sich per E-Mail zu melden. Daraufhin bekamen sie alle wichtigen Informationen zum Ablauf der Studie sowie ihre PIN zum Einloggen für Casus mitgeteilt. Ebenso erhielten sie die Einwilligungsschrift und das Auszahlungsformular mit der Bitte, beides ausgefüllt zurückzuschicken. Die Casus-Fälle waren in drei Teile untergliedert, wobei die Proband*innen linear durch die Folien geleitet wurden. Eine Möglichkeit, zu vorherigen Folien zurückzukehren, bestand dabei nicht. Die Benutzung eines Taschenrechners war erlaubt, wenn auch nicht zwingend notwendig. Jeder Casus-Fall war nach dem in Abbildung 4 dargestellten Schema aufgebaut:

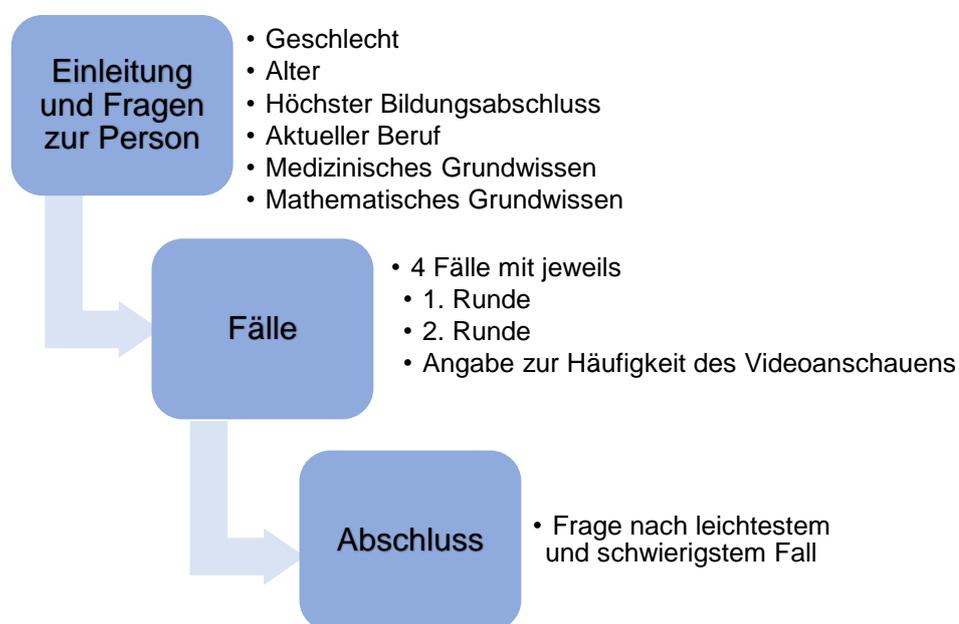


Abbildung 4: Der Studienablauf

Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Im ersten Teil der Studie wurden nach einem einleitenden Text die demografischen Daten (Geschlecht, Alter, höchster Bildungsabschluss und

aktueller Beruf) und Angaben zum medizinischen und mathematischen Grundwissen erhoben. Das Grundwissen sollte hierbei mit einer Note zwischen 1 (sehr gut) und 6 (ungenügend) bewertet werden. Im Hauptteil bearbeiteten die Proband*innen die eigentlichen Fälle. Die Aufgabe der Proband*innen bestand darin, ein Video mit einem fiktiven ärztlichen Gespräch mit Patient*innen anzuschauen. Nach dem Betrachten des jeweiligen Videos wurde die Zeitmessung gestartet und die Proband*innen wurden dazu aufgefordert, eine Frage zu beantworten, welche bereits vor dem Anschauen des Videos mitgeteilt wurde. Dies ist realitätsnah, da Patient*innen häufig eine Frage stellen, bevor die Ärzt*innen die entsprechenden Informationen liefert. Die Frage war sowohl bei den Videos, in denen die Informationen Bayesianisch vorgestellt wurden, als auch in den Videos, in denen die Informationen diagnostisch erklärt wurden, dieselbe: „Wie viele Patient*innen mit einem positiven/auffälligen Testergebnis haben die Krankheit x?“. Somit wurde stets nach dem positiven prädiktiven Wert gefragt. In einen Lückentext (___ Patient*innen von ___ Patient*innen) mussten dabei zwei Zahlen angegeben werden. Die Bearbeitung jedes Falls umfasste zwei Runden. In der ersten Runde wurden die Proband*innen ausdrücklich darauf hingewiesen, das Video nur einmal anzusehen. Da im klinischen Alltag die Ergebnisse fast immer nur einmal mitgeteilt werden, wurden die Proband*innen bewusst bereits nach einmaligem Anschauen dazu aufgefordert, eine Antwort abzugeben. In der zweiten Runde durften die Proband*innen sich dasselbe Video nochmals beliebig oft ansehen und erneut dieselbe Frage beantworten, wobei die in der zweiten Runde gemessene Zeit für die Auswertung nicht relevant war. Stattdessen sollten sie zusätzlich angeben, wie oft sie sich das Video angeschaut hatten, bis sie sich des Ergebnisses sicher waren. Dieses Bearbeiten der mehreren Runden ermöglichte es auszuwerten, inwiefern die Ergebnisrichtigkeit mit der Häufigkeit des Anschauens des Videos korrelierte. Die Zeit, welche die Proband*innen zum Bearbeiten der Studie hatten, war unbegrenzt, wurde aber aufgezeichnet. Nachdem die Proband*innen alle Aufgaben des Hauptteils bearbeitet hatten, folgte zum Abschluss eine letzte Frage zur Schwierigkeit der Aufgaben. Hierbei sollten die eben bearbeiteten Aufgaben in einer Sortierantwort so angeordnet werden, dass der Fall, der am leichtesten empfunden wurde, ganz oben stehen sollte, während der am schwierigsten empfundene Fall ganz unten sein sollte.

3.3 Statistische Methoden

Die Daten und Ergebnisse wurden anonymisiert erhoben und ausgewertet. Die Ergebnisauswertung wurden mit der Statistiksoftware R (82) durchgeführt, wobei für die statistischen Analysen (verallgemeinerte) lineare gemischte Regressionsmodellen und das Paket "lme4" (83) verwendet wurden. Einerseits wurde die Wahrscheinlichkeit einer richtigen Lösung bei der Beantwortung der Aufgaben durch die Proband*innen mit Hilfe von verallgemeinerten linearen gemischten Modellen (mit einer logistischen Verknüpfungsfunktion) vorhergesagt. Hierfür wurden die Ergebnisse der Proband*innen binär kodiert, wobei 0 für eine falsche Antwort und 1 für eine richtige Antwort stand. Damit ein Ergebnis als richtig kodiert wurde, mussten beide erfragten Zahlen richtig sein. Zum Beispiel beim Kontext des Schilddrüsenkarzinoms war die richtige Antwort „20 Patient*innen von 130 Patient*innen“. Diese wurde dementsprechend auch nur als solche gewertet, wenn sowohl „20“ als auch „130“ richtig angegeben wurde. Andersseits wurde auch die zur Lösung der Aufgabe benötigte Zeit mit Hilfe von linearen gemischten Modellen vorhergesagt. In beiden Modellen wurde die Richtung als Bayesianische Informationen und die Version ohne Visualisierung wurde als Referenzkategorie spezifiziert und die potenziellen Einflussfaktoren diagnostische Informationen und Häufigkeitsnetz über Dummy-Kodierung einbezogen. Zudem wurden auch verschiedene unabhängige Variablen (u.a. Alter, Beruf) als potenzielle, weitere Einflussfaktoren auf die Richtigkeit der Ergebnisse untersucht.

3.4 Ethikvotum

Die Studie wurde in Übereinstimmung mit der Deklaration von Helsinki durchgeführt. Dafür wurde ein Ethikvotum bei der „Ethikkommission bei der LMU München“ eingeholt. Die Projektnummer lautete 21-0024. Die erforderlichen Richtlinien wurden eingehalten. Alle Proband*innen wurden über den Inhalt und den Ablauf der Studie aufgeklärt. Sie hatten das Recht, die Studie zu jedem Zeitpunkt ohne negative Konsequenzen abubrechen. Die Belehrung der Proband*innen erfolgte schriftlich, wobei die Einwilligungserklärung von den Proband*innen und dem*der Studienverantwortlichen unterschrieben wurde.

4 Ergebnisse

4.1 Soziodemografische Daten

An der Studie nahmen 110 Proband*innen teil. Jedoch wurden nur die Ergebnisse von 109 Proband*innen ausgewertet, da ein*e Proband*in nur den Einleitungsteil aber nicht die Fälle bearbeitete. Von den 109 Proband*innen waren 71% weiblichen Geschlechts und 29% männlichen Geschlechts. Die Möglichkeit divers oder keine Angabe zu wünschen, wurde von niemandem ausgewählt. Das durchschnittliche Alter betrug 29,2 Jahre (SD 11,2 Jahre). Der*die jüngste Proband*in war 18 Jahre alt, wohingegen der*die Ältteste 64 Jahre alt war.

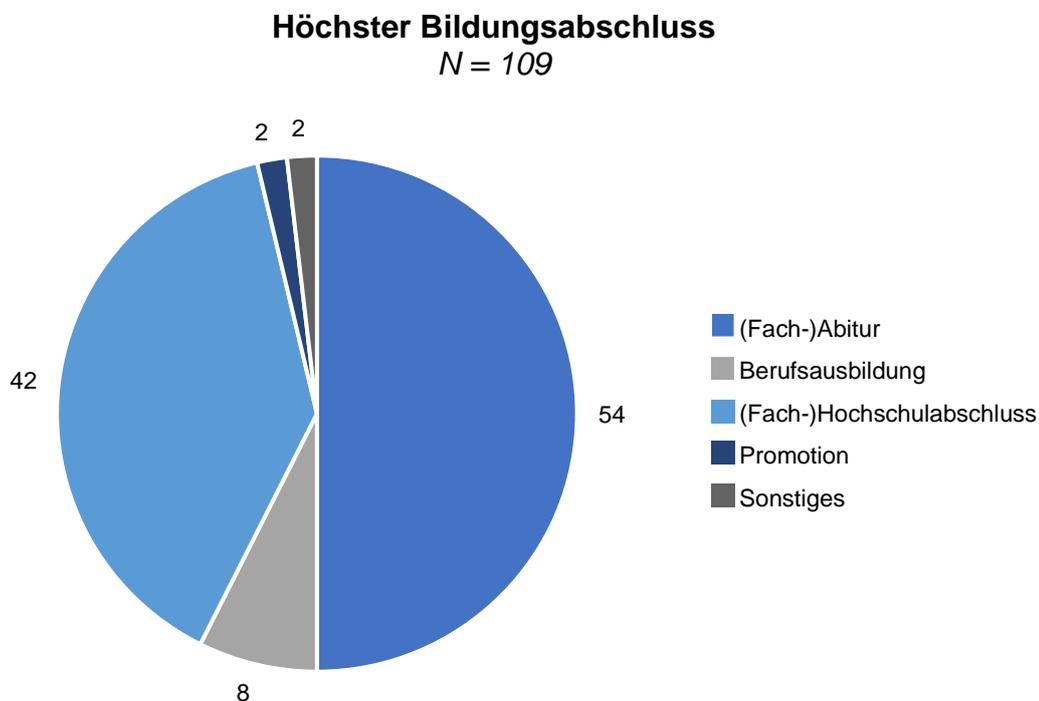


Abbildung 5: Höchster Bildungsabschluss

Die Hälfte der Proband*innen gab das (Fach-)Abitur als höchsten Bildungsabschluss an. An zweiter Stelle stand mit 39% der (Fach-) Hochschulabschluss. Die übrigen Nennungen verteilten sich in absteigender Häufigkeit auf die Möglichkeiten „Berufsausbildung“, „Sonstiges“ und „Promotion“. Die Möglichkeiten „Ohne Abschluss“, „Hauptschulabschluss“ und „Realschulabschluss“ wurden nie gewählt und sind deshalb in dem Kreisdiagramm (siehe Abbildung 5) nicht aufgeführt.

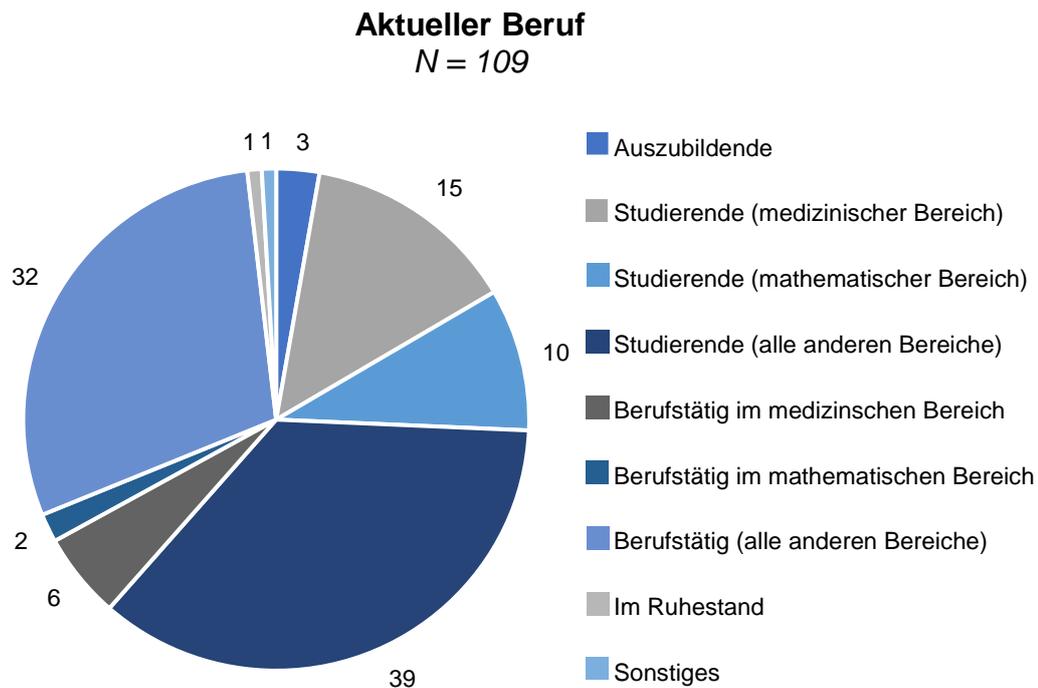


Abbildung 6: Aktueller Beruf

Ebenso wurde der aktuelle Beruf erfragt. 59% der Proband*innen gaben an, Studierende zu sein, 23% davon im medizinischen Bereich. 37% der Proband*innen waren berufstätig, davon 15% in medizinischen Berufen. Somit hatten 19% der Proband*innen einen medizinischen Hintergrund. 11% der Proband*innen studierten oder arbeiteten im mathematischen Bereich. Kein*e Proband*in war Schüler*in oder arbeitssuchend/arbeitslos, weshalb auch hier diese Möglichkeiten grafisch (siehe Abbildung 6) nicht abgebildet wurden.

4.2 Angaben zum Grundwissen

Des Weiteren wurde in der Studie das Grundwissen der Proband*innen erfragt. Dadurch lässt sich ein Bild über die Grundvoraussetzungen der Proband*innen schaffen und mögliche Korrelationen mit den Ergebnissen der Fälle aufdecken. Dazu wurden die Proband*innen gebeten, ihr medizinisches und mathematisches Grundwissen anhand von Schulnoten (1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = ausreichend, 5 = mangelhaft, 6 = ungenügend) anzugeben. Es waren nur ganze Zahlen wählbar. Dies erfolgte in Selbsteinschätzung und ist somit als sehr subjektiv anzusehen. Eine objektivere Methode zum Beispiel anhand der tatsächlichen damaligen Schulnoten wäre nicht realisierbar gewesen, da in der Schule in der Regel nur Mathematik, aber nicht Medizin unterrichtet

Ergebnisse

wurde. Außerdem wäre eine Besserung des Grundwissens nach Schulabschluss dadurch außer Acht gelassen worden. Da die meisten Proband*innen jedoch mit Schulnoten vertraut sein sollten, wurde die Spannweite „1“ bis „6“ gewählt.

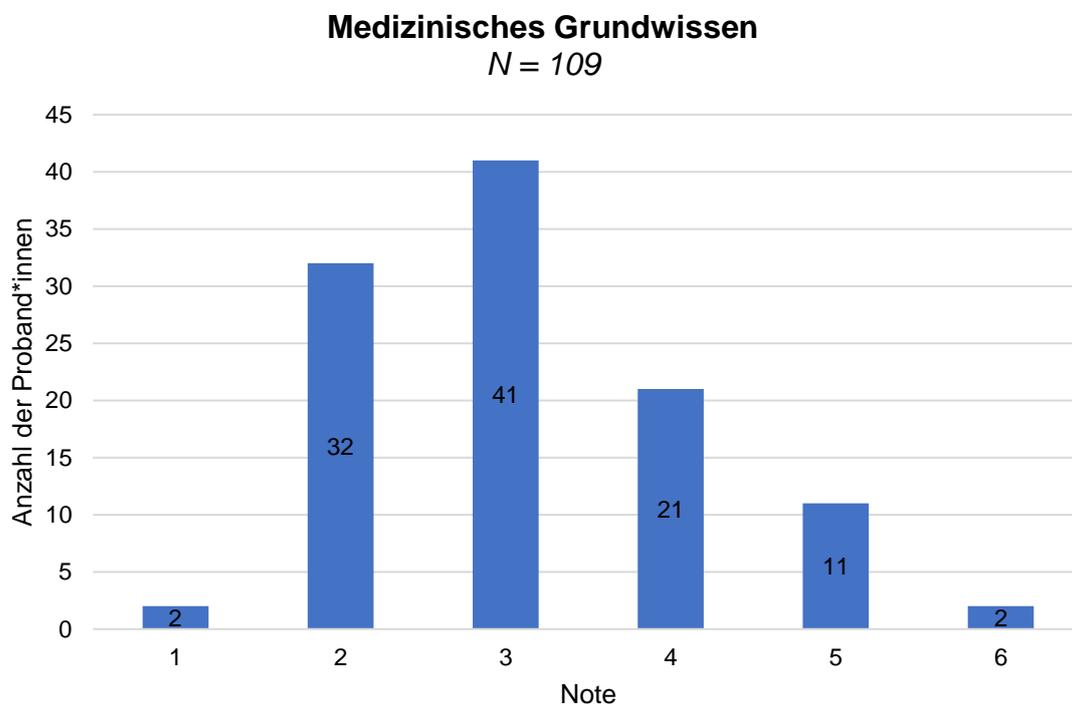


Abbildung 7: Medizinisches Grundwissen

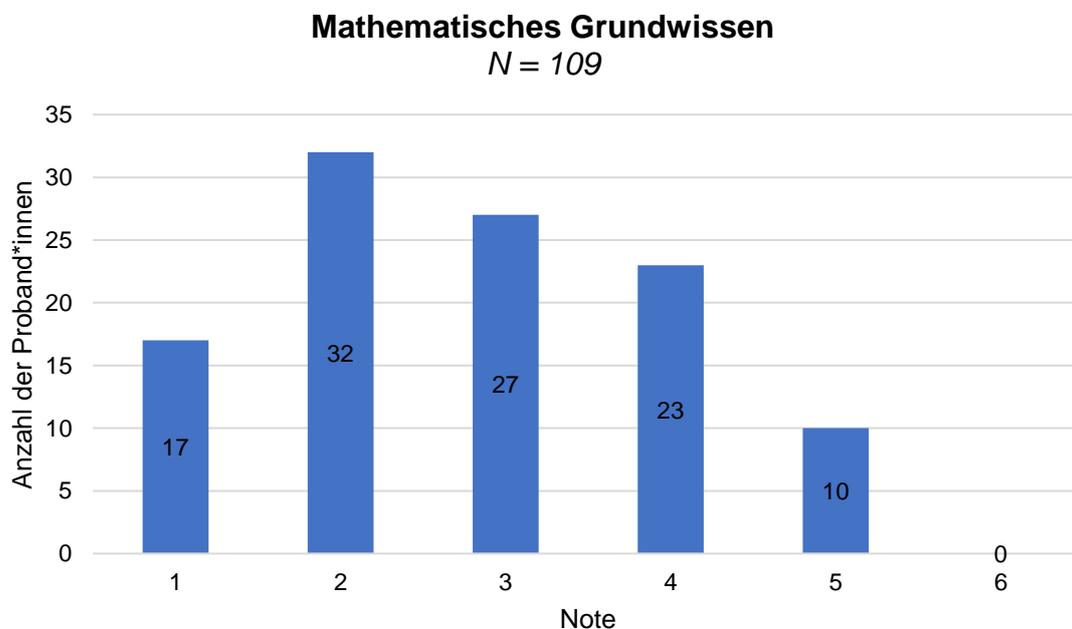


Abbildung 8: Mathematisches Grundwissen

Ergebnisse

Bei beiden Fragen ist eine Tendenz zur Mitte erkennbar: Beim medizinischen Grundwissen (siehe Abbildung 7) wurde mit 38% am häufigsten die Note 3 gewählt. Aber auch der Note 2 (29%) und der Note 4 (19%) ordneten sich einige Proband*innen zu. Die Noten 1, 5 und 6 wurden lediglich vereinzelt gewählt. Beim mathematischen Grundwissen (siehe Abbildung 8) sah die Verteilung ähnlich aus, war jedoch zu den besseren Noten hingehend verschoben und gleichmäßiger verteilt. So wurde hier mit 29% die Note 2 am häufigsten gewählt. Mit wenig Abstand folgten die Note 3 (25%) und die Note 4 (21%). Auch hier bildeten Note 1 und Note 5 die Schlusslichter. Kein*e Proband*in schätze sein* ihr mathematisches Grundwissen mit der Note 6 ein. Somit hatten die Proband*innen ein leicht besseres mathematisches (Durchschnittsnote 2,8) als ein medizinisches Grundwissen (Durchschnittsnote 3,1), was vermutlich auf die grundlegende mathematische Schulausbildung zurückzuführen ist.

4.3 Ergebnisrichtigkeit

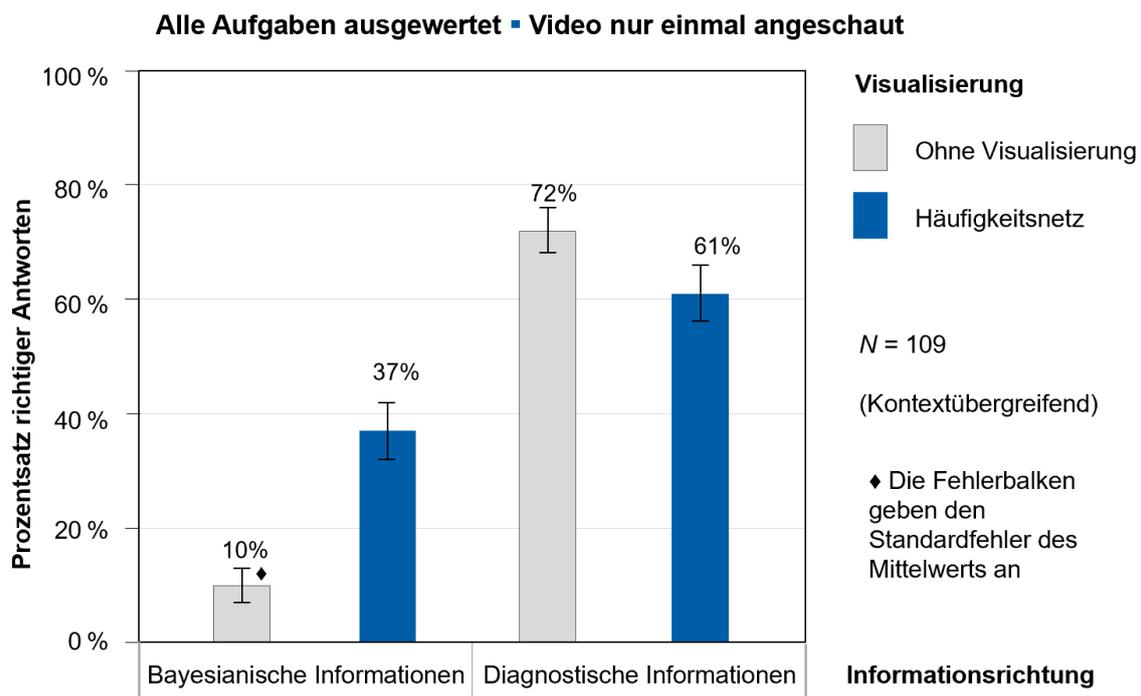


Abbildung 9: Ergebnisrichtigkeit nach einmaligem Anschauen, wobei alle Aufgaben ausgewertet wurden

$N = 109$. Die Fehlerbalken geben den Standardfehler des Mittelwerts an. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Vorweg ist zu sagen, dass sich die nachfolgend beschriebenen und in Abbildung 9 ersichtlichen Ergebnisse auf den Fall beziehen, wenn alle Aufgaben

Ergebnisse

ausgewertet wurden sowie das Video einmal angesehen wurde. Zwischen den verschiedenen Versionen wurden große Unterschiede die Richtigkeit der Ergebnisse betreffend ersichtlich. So führte das einmalige Ansehen des Videos bei der diagnostischen Version ohne Visualisierung in circa 7 von 10 Fällen zu einem richtigen Ergebnis. Somit übertraf diese Version sogar die diagnostische Version mit Häufigkeitsnetz, die nur in circa 6 von 10 Fällen von den Proband*innen gelöst werden konnte. Die beiden Bayesianischen Versionen wiesen die niedrigsten Lösungsraten auf. Hierbei schien aber die Visualisierung zu einer besseren Verständlichkeit geführt zu haben (37%). Ohne Visualisierung war nur 10% der Proband*innen ein richtiges Lösen der Aufgabe möglich.

Um die Auswirkungen der Informationsrichtung und des Häufigkeitsnetzes statistisch zu vergleichen, wurde ein verallgemeinertes lineares gemischtes Modell mit einer logistischen Verknüpfungsfunktion geschätzt. Damit konnte die Wahrscheinlichkeit vorhergesagt werden, dass Proband*innen die Aufgabe richtig lösen, nachdem das Video nur einmal, aber über alle Aufgaben hinweg, angesehen wurde. In diesem Modell wurde die Bayesianische Informationsrichtung ohne Visualisierung als Referenzkategorie spezifiziert und die möglichen Erklärungsfaktoren "diagnostische Information" und "Häufigkeitsnetz" über Dummy-Kodierung eingeschlossen. Da außerdem zu erwarten war, dass sich die Wirksamkeit des Häufigkeitsnetzes unterscheidet, je nachdem, ob Bayesianische oder diagnostische Informationen gezeigt wurden, wurde ein Interaktionsterm Informationsrichtung \times Häufigkeitsnetz modelliert. Darüber hinaus wurden von allen Proband*innen Alter, Geschlecht und der höchste Bildungsgrad erhoben. Diese Variablen sowie der Kontext der Aufgabe (welche Krankheit oder welches Syndrom) und die Reihenfolge der Aufgabe (d.h. wurde die Aufgabe als erste, zweite, dritte oder vierte Aufgabe gestellt) konnten ebenfalls als potenzielle Prädiktoren in das verallgemeinerte lineare Mischmodell implementiert werden. Im Folgenden werden die Ergebnisse des folgenden Modells vorgestellt:

$$Y = b_0 + b_1 \cdot \text{Informationsrichtung} + b_2 \cdot \text{Häufigkeitsnetz} + b_3 \cdot \text{Informationsrichtung} \times \text{Häufigkeitsnetz} + b_4 \cdot \text{Reihenfolge der Aufgabe} + b_5 \cdot \text{Alter} + b_6 \cdot \text{Geschlecht}$$

mit einer logistischen Verknüpfungsfunktion $e^Y/(1+e^Y)$.

Ergebnisse

Wie aus Tabelle 4 ersichtlich ist, war der Regressionskoeffizient für die Umkehrung der Richtung der statistischen Information von Bayesianischer zu diagnostischer Information signifikant, und die Darstellung eines Häufigkeitsnetzes führte ebenfalls zu einem signifikanten Regressionskoeffizienten. Wie bereits erwartet und deskriptiv belegt, hing die Wirksamkeit des Häufigkeitsnetzes davon ab, ob Bayesianische Informationen oder diagnostische Informationen präsentiert wurden: Der Interaktionseffekt zwischen Informationsrichtung und Visualisierung war signifikant. So nahm die positive Wirkung des Häufigkeitsnetzes signifikant ab, wenn anstelle von Bayesianischen Informationen diagnostische Informationen gezeigt werden. Wie die deskriptiven Ergebnisse bereits angedeutet hatten, gab es einen signifikanten Effekt im Zusammenhang mit der Reihenfolge der Aufgabe (d.h. wurde die Aufgabe als erste, zweite, dritte oder vierte Aufgabe gestellt). Ein weiterer statistisch signifikanter Prädiktor war das Alter der Proband*innen (je jünger der*die Proband*in in unserer Studie war, desto besser war die Gesamtleistung). Andere Variablen (z. B. der Kontext, das Geschlecht) waren dagegen keine signifikanten Einflussfaktoren.

*Tabelle 4: Parameterschätzungen des verallgemeinerten linearen gemischten Modells zur Vorhersage der Leistung der Proband*innen, nachdem sie das Video einmal angeschaut haben*

Unabhängige Variablen	Regressionsgewicht	Standardfehler	z	p
y-Achsenabschnitt	-0,53	0,25	-2,13	0,03
Informationsrichtung	1,77	0,22	7,97	<0,001
Häufigkeitsnetz	0,46	0,16	2,94	0,003
Informationsrichtung × Häufigkeitsnetz	-0,87	0,17	-5,08	<0,001
Reihenfolge der Aufgabe	0,69	0,16	4,24	<0,001
Alter	-0,62	0,25	-2,42	0,02
Geschlecht	0,32	0,25	1,29	0,20

Bemerkung: Andere mögliche unabhängige Variablen (wie der Kontext der Aufgabe (über Dummy-Kodierung) oder das höchste Bildungsniveau) waren keine signifikanten Prädiktoren (wie auch das Geschlecht kein signifikanter Prädiktor im Modell war), wenn sie in das Modell implementiert wurden. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Zusammenfassend lässt sich Folgendes feststellen: Die beste Leistung wurde erzielt, wenn die statistischen Informationen als diagnostische Informationen und nicht als Bayesianische Informationen präsentiert wurden. Das Häufigkeitsnetz (das die Bayesianische Informationsversion um diagnostische Informationen ergänzt) war hilfreich, wenn es sich um Bayesianische Informationen handelte. Das Häufigkeitsnetz stellte jedoch eher ein Hindernis als eine Hilfe dar, wenn es sich um diagnostische Informationen handelte (da das Häufigkeitsnetz dann nur Bayesianische, weniger relevante Informationen ergänzte).

4.4 Lerneffekte

4.4.1 Verbesserung der Ergebnisrichtigkeit im Verlauf der Studie

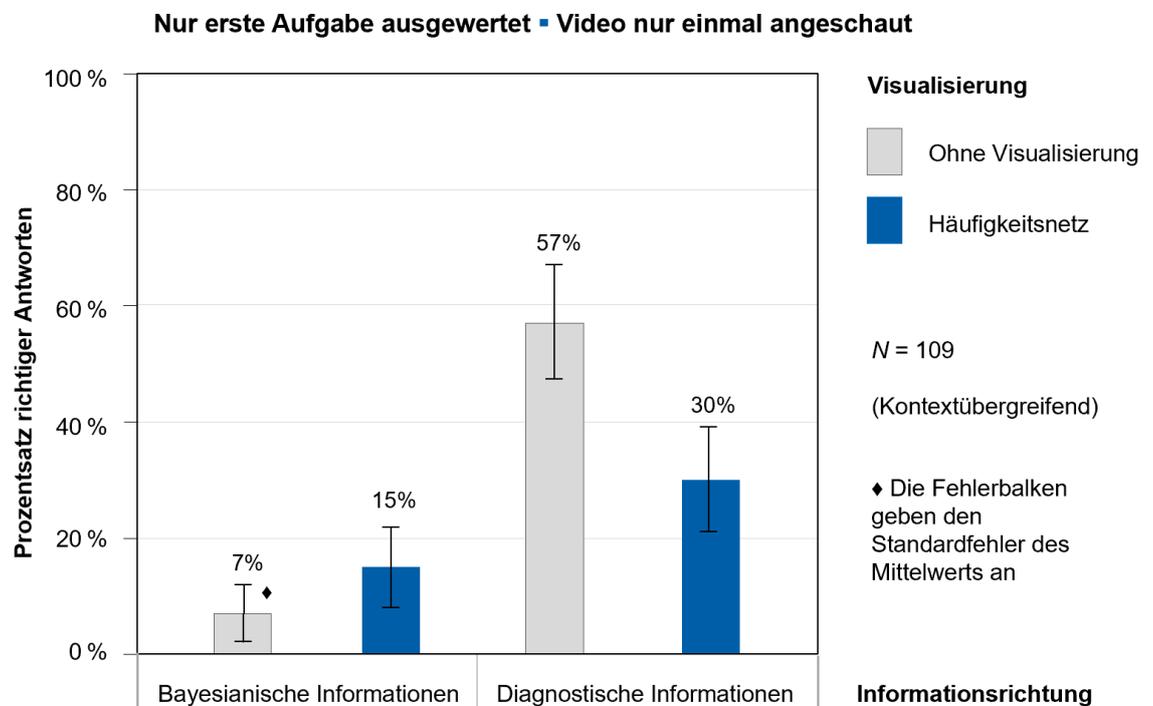


Abbildung 10: Ergebnisrichtigkeit nach einmaligem Anschauen, wenn nur die erste Aufgabe ausgewertet wurde

N = 109. Die Fehlerbalken geben den Standardfehler des Mittelwerts an. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Da die Proband*innen in der Studie vier, wenn auch unterschiedliche Fälle hintereinander bearbeiten mussten, war es von Interesse herauszufinden, ob das vorherige Lösen eines ähnlichen Falls die Wahrscheinlichkeit für das richtige Lösen eines darauffolgenden Falls erhöhte. Dazu wurde die Anzahl richtiger Lösungen berechnet, wenn jede*r Proband*in nur die erste Aufgabe nach

Ergebnisse

einmaligem Anschauen gelöst hatte, da hier noch kein Lerneffekt im Rahmen der Studie eingetreten sein konnte. Bayesianische Aufgaben ohne Visualisierung konnten nur 7% der Proband*innen lösen, mit Häufigkeitsnetz circa doppelt so viele (Zur Erinnerung: Die Darstellung eines Häufigkeitsnetzes in der Version mit Bayesianischen Informationen bedeutete, dass zusätzlich diagnostische Informationen dargestellt wurden.) Bei Aufgaben mit diagnostischen Informationen konnten deutlich höhere Lösungsraten erzielt werden (30% mit Häufigkeitsnetz, also mit zusätzlichen Bayesianischen Informationen). Im Unterschied zu den Bayesianischen Aufgaben schnitten die diagnostischen Aufgaben ohne Visualisierung mit doppelt so vielen richtig gelösten Fällen deutlich besser ab als jene mit Häufigkeitsnetz.

Vergleicht man diese Ergebnisse (siehe Abbildung 10) mit den Ergebnissen, wenn alle Aufgaben gelöst wurden (siehe Abbildung 9), fällt auf, dass sich nach Lösen aller Aufgaben deskriptiv höhere Lösungsraten in der Analyse zeigten. Die zuerst bearbeitete Aufgabe war dabei im Durchschnitt diejenige mit dem schlechtesten Anteil richtiger Lösungen. Die durchschnittliche Leistung der Proband*innen über alle Aufgaben hinweg stieg in der Version mit Bayesianischen Informationen ohne Visualisierung von 7% auf 10%. Mit Hilfe eines Häufigkeitsnetzes (das diagnostische Informationen hinzufügt) konnten 37% der Proband*innen (im Vergleich zu 15%) eine Bayesianische Aufgabe lösen; 72% der Proband*innen (im Vergleich zu 57%) konnten eine diagnostische Aufgabe ohne Visualisierung lösen; und 61% (im Vergleich zu 30%) der Proband*innen konnten eine diagnostische Informationsaufgabe mit Hilfe des Häufigkeitsnetzes lösen. Dies lässt auf einen Lerneffekt, genauer gesagt einen Sequenzeffekt, der Proband*innen im Verlauf der Studie rückschließen. Aufgaben, die in der Abfolge der vier Aufgaben später gestellt wurden, wurden mit größerer Wahrscheinlichkeit richtig gelöst. Dabei ist wiederum zu beachten, dass Aufgaben mit Bayesianischen Informationen und diagnostischen Informationen abwechselnd bearbeitet wurden (d.h. bei der Hälfte der Aufgaben wurden die Bayesianischen Informationen zuerst gezeigt, bei der anderen Hälfte der Aufgaben die diagnostischen Informationen). Besonders zu betonen ist die Tatsache, dass dieser Lerneffekt eintrat, obwohl die verschiedenen Fälle zwar strukturell ähnlich waren, sich aber doch alle auf unterschiedliche medizinische Sachverhalte bezogen. Im fiktiven Falle, dass Proband*innen mehrere Aufgaben

Ergebnisse

der gleichen Version lösen müssen, ist mit einer noch stärkeren Verbesserung zu rechnen.

4.4.2 Verbesserung der Ergebnisrichtigkeit durch mehrmaliges Anschauen

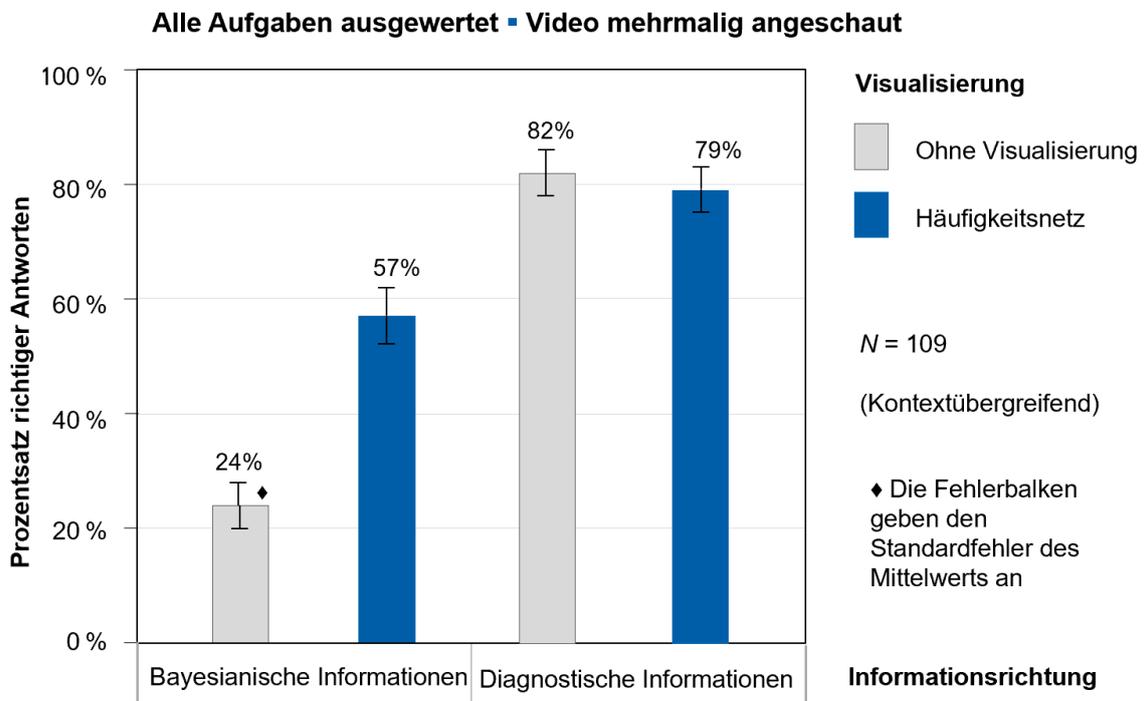


Abbildung 11: Ergebnisrichtigkeit nach mehrmaligem Anschauen, wobei alle Aufgaben ausgewertet wurden

$N = 109$. Die Fehlerbalken geben den Standardfehler des Mittelwerts an. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

In einer zweiten Runde bestand für die Proband*innen die Möglichkeit, dasselbe Video nochmals so oft anzusehen, bis sie sich des Ergebnisses sicher waren. Danach sollten sie ein erneutes Ergebnis sowie die Häufigkeit des Anschauens angeben. Eine hohe Steigerung der Ergebnisrichtigkeit konnte beim Bayesianischen Fall mit Häufigkeitsnetz gezeigt werden. Auch die diagnostischen Versionen, welche bereits nach einmaligem Anschauen häufig zu richtigen Ergebnissen geführt hatten, konnten nun von noch mehr Proband*innen gelöst werden. Besonders gewinnbringend war das mehrmalige Anschauen für die Version mit Bayesianischen Informationen und ohne Visualisierung, welche zuvor am schlechtesten abgeschnitten hatte. Hierbei ließ sich eine über doppelt so hohe Anzahl richtiger Ergebnisse feststellen. Im Gesamten betrachtet stieg die Leistung aller Aufgabenversionen nach mehrmaligem Betrachten, die

Ergebnisse

Reihenfolge der Lösungsrichtigkeiten der verschiedenen Versionen blieb aber gleich zu der nach einmaligem Anschauen. Nur wenige Proband*innen korrigierten ihre Antwort nach mehrmaliger Betrachtung der Videos zu einem falschen Ergebnis hin. Auffällig war außerdem, dass der Lerneffekt durch mehrmaliges Betrachten des Videos bei den Versionen mit Häufigkeitsnetz höher war als bei den rein auditiven Versionen.

Ebenso wurde ein weiteres verallgemeinertes lineares gemischtes Modell spezifiziert, in dem die abhängige Variable die Leistung der Proband*innen über alle Aufgaben hinweg war, nachdem das Video mehrmalig angesehen werden durfte (siehe Tabelle 5). Auch hier wurden die Richtung der Information und das Häufigkeitsnetz als unabhängige Variablen über Dummy-Kodierung implementiert. Zusätzlich konnten die Variablen Alter, Geschlecht, höchster Bildungsgrad, Kontext, Reihenfolge der Aufgabe und wie oft das Video angeschaut wurde als potenzielle Prädiktoren eingesetzt werden. Im Folgenden werden die Ergebnisse des folgenden Modells vorgestellt:

$$Y = b_0 + b_1 \cdot \text{Informationsrichtung} + b_2 \cdot \text{Häufigkeitsnetz} + b_3 \cdot \text{Informationsrichtung} \times \text{Häufigkeitsnetz} + b_4 \cdot \text{Reihenfolge der Aufgabe} + b_5 \cdot \text{Alter} + b_6 \cdot \text{Geschlecht} + b_7 \cdot \text{Häufigkeit des Ansehens des Videos}$$

wiederum mit einer logistischen Verknüpfungsfunktion. Wie im Modell zuvor war der Regressionskoeffizient für die Umkehrung der Richtung der statistischen Information von Bayesianischer zu diagnostischer Information erneut signifikant und die Darstellung eines Häufigkeitsnetzes führte ebenfalls zu einem signifikanten Regressionskoeffizienten. Wie zuvor hing die Wirksamkeit des Häufigkeitsnetzes davon ab, ob Bayesianische Informationen oder diagnostische Informationen gezeigt wurden: Der Interaktionseffekt zwischen Richtung und Visualisierung war signifikant. So nahm die Wirkung des Häufigkeitsnetzes erneut signifikant ab, wenn diagnostische statt Bayesianischen Informationen präsentiert wurden. Die Variablen Reihenfolge der Aufgabe und Alter waren in diesem Modell jedoch nicht mehr signifikant. Es konnte (zusätzlich zum Lernen aus einer zuvor bearbeiteten Aufgabe) ein weiterer Lerneffekt beobachtet werden: Das mehrmalige Anschauen eines Videos führte zu einer besseren Leistung (vergleiche die Prozentsätze richtiger Antworten in Abbildung 9 und Abbildung 11), siehe auch das verallgemeinerte lineare gemischte Modell in den

Ergebnisse

inferenzstatistischen Analysen). Der zuvor ersichtliche negative Effekt des Häufigkeitsnetzes bei der Verwendung von Diagnoseinformationen wurde anschaulich reduziert, wenn das Video bereits mehrmalig angesehen wurde oder wenn ähnliche Aufgaben bereits bearbeitet wurden. Das heißt, dass das Häufigkeitsnetz nach einiger Übung keine Barriere mehr darstellte. Interessanterweise war die Häufigkeit der Betrachtung des Videos kein signifikanter Faktor in dem Modell. Dennoch verbesserten sich die Leistungen der Proband*innen, wenn sie ihre Antworten nach mehrmaligem Betrachten des Videos korrigieren durften (vgl. Leistungsraten in Abbildung 9 und Abbildung 11).

*Tabelle 5: Parameterschätzungen des verallgemeinerten linearen gemischten Modells zur Vorhersage der Leistung der Proband*innen, nachdem sie das Video mehrmalig angeschaut haben*

Unabhängige Variablen	Regressionsgewicht	Standardfehler	z	p
y-Achsenabschnitt	0,80	0,27	2,95	0,003
Informationsrichtung	1,69	0,23	7,40	<0,001
Häufigkeitsnetz	0,60	0,16	3,64	<0,001
Informationsrichtung × Häufigkeitsnetz	-0,74	0,17	-4,35	<0,001
Reihenfolge der Aufgabe	0,15	0,16	0,92	0,36
Alter	-0,52	0,27	-1,96	0,05
Geschlecht	0,29	0,26	1,12	0,26
Häufigkeit der Betrachtung des Videos	0,08	0,19	0,42	0,67

Bemerkung: Andere mögliche unabhängige Variablen (wie der Kontext der Aufgabe (über Dummy-Kodierung) oder das höchste Bildungsniveau) waren keine signifikanten Prädiktoren (wie auch das Geschlecht kein signifikanter Prädiktor im Modell war), wenn sie in das Modell implementiert wurden. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

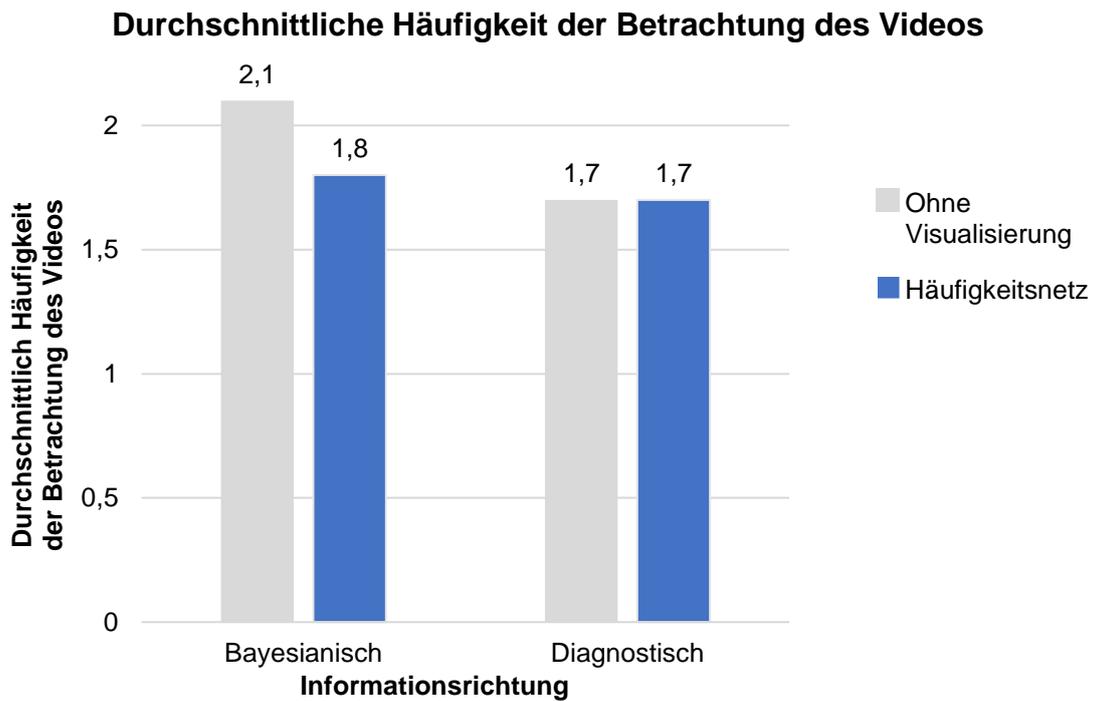


Abbildung 12: Durchschnittliche Häufigkeit der Betrachtung des Videos

Die Verbesserung der 2. abgegebenen Lösungen war durch die erneutes Videoanschauen möglich. Es stellte sich die Frage, wie oft die Proband*innen die Videos angeschaut haben, um zuvor aufgeführte Verbesserung zu erreichen. Dabei durften die Proband*innen die Häufigkeit des Anschauens selbst wählen, bis sie sich des Ergebnisses sicher waren, wurden aber aufgefordert, diese Anzahl anzugeben (siehe Abbildung 12). In die anzugebende Häufigkeit zählte sowohl das erstmalige als auch alle weiteren Male des Anschauens rein. In wenigen Fällen wurde die Häufigkeit des Anschauens nicht angegeben, sodass alle restlichen Daten für die Auswertung der Häufigkeit verwendet wurden. Die Anzahl der Betrachtungen, die notwendig waren, um sich der richtigen Antwort sicher zu sein, unterschieden sich je nach verwendeter Version:

Am seltensten wurden mit durchschnittlichen 1,7-mal die diagnostischen Versionen angeschaut. Die Bayesianischen Fälle mit Häufigkeitsnetz wurden im Schnitt 1,8-mal gesehen. Am häufigsten mussten die Proband*innen die Videos mit Bayesianischen Informationen und ohne Visualisierung ansehen (durchschnittlich 2,1-mal), bis sie sich sicher zu sein schienen. Dieser Fall hatte nach einmaligem Anschauen die niedrigste Lösungsrate gehabt. Vergleicht man

diese aber mit der Lösungsrate nach mehrmaligem Anschauen, scheint sich das mehrmalige Anschauen hier aber besonders gelohnt zu haben.

4.5 Ergebnisgeschwindigkeit

Neben der Ergebnisrichtigkeit wurde in der vorliegenden Studie auch die Ergebnisgeschwindigkeit ausgewertet. Die Zeiten wurden in Casus automatisch erfasst. Dabei handelt es sich um die reine Bearbeitungszeit, welche nach dem Lesen der Aufgabe und dem erstmaligen Betrachten des Videos startete und nach der Beantwortung der Aufgabenfrage endete. Durch dieses andere Setting lassen sich die deutlich kürzeren Zeiten im Vergleich zu ähnlichen Studien, beispielsweise Kunzelmann et al. (2022) und Binder et al. (2021), erklären. In diesen Studien wurde auch die Lesezeit in die ausgewertete Zeit eingerechnet, die offensichtlich einen erheblichen Teil des Lösungsprozesses in Anspruch nimmt, wodurch die Zeiten länger ausfielen. Ein weiterer Unterschied zu den genannten Studien war, dass in dieser Studie die Frage bereits bekannt war und es so einfacher war, sich auf die Informationen zu konzentrieren. Die nachfolgende Tabelle 6 zeigt die Medianzeiten (und 25- und 75-Quartile), wobei zwischen den Proband*innen mit falschen und richtigen Antworten unterschieden wurde.

Ergebnisse

Tabelle 6: Deskriptive Ergebnisse über alle Kontexte und über alle gelösten Aufgaben, nachdem das Video einmal angesehen wurde

		Prozentsatz richtiger Antworten		Benötigte Zeit für eine falsche Antwort				Benötigte Zeit für eine richtige Antwort				Score für die diagnostische Effizienz:
		N	%	N	Q1	Median	Q3	N	Q1	Median	Q3	Mediane Zeit für eine richtige Antwort geteilt durch den Prozentsatz richtiger Antworten
												Zeit in [s]
Bayesianische Informationen	Ohne Visualisierung	109	10	98	12,0	22,5	42,0	11	13,0	106,0	133,0	17:30
	Häufigkeitsnetz	109	37	69	12,0	21,0	32,0	40	9,0	13,5	20,5	0:36
Diagnostische Informationen	Ohne Visualisierung	109	72	31	12,0	16,0	26,0	78	10,0	14,5	19,0	0:20
	Häufigkeitsnetz	109	61	42	13,0	21,5	32,0	67	9,0	14,0	30,0	0:23

Prozentsatz richtiger Antworten, benötigte Zeit für eine falsche oder richtige Antwort mit 1. Quartil (Q1), Median und 3. Quartil (Q3) und Score für die diagnostische Effizienz. Der Wert für die diagnostische Effizienz wurde berechnet, indem die mediane Zeit für eine richtige Antwort durch den Prozentsatz richtiger Antworten geteilt wurde. N gibt die Anzahl der Proband*innen an. Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

4.5.1 Bei falschen Antworten

Zunächst fiel auf, dass falsch gelöste Aufgaben ähnlich viel Zeit zur Beantwortung benötigten. Die Mediane der verschiedenen Versionen lagen zwischen 16,0 und 22,5 Sekunden, wodurch die Zeit bei falsch gelösten Aufgaben offenbar nicht so sehr von der Version abzuhängen schien. Dennoch ließ sich eine leichte Tendenz erkennen, inwiefern die Zeiten mit den Ergebnisrichtigkeiten korrelierten. So wurde im Schnitt für die Version mit der niedrigsten Lösungsrate (Bayesianisch, ohne Visualisierung) am meisten Zeit benötigt, während für die am häufigsten richtige Version (diagnostisch, ohne Visualisierung) Gegenteiliges galt. Die übrigen zwei Versionen gliederten sich dementsprechend mit mittleren medianen Zeiten ein.

4.5.2 Bei richtigen Antworten

Zusätzlich wurden noch differenziert die Geschwindigkeiten jener Proband*innen ausgewertet, welche die Aufgaben richtig gelöst hatten. Diese Zeiten sind weitaus interessanter, da einerseits im klinischen Alltag richtiges Verständnis dringend gewünscht ist, andererseits hierbei auch deutlichere Unterschiede der Bearbeitungszeiten ersichtlich wurden. Vor allem in der Bayesianischen Version ohne Visualisierung zeigte sich ein auffälliges Ergebnis: Um hierbei eine richtige Lösung abzugeben, brauchten die Proband*innen eine mediane Zeit von 106 Sekunden. Demgegenüber wurden die diagnostischen Aufgaben deutlich schneller bearbeitet (Median ohne Visualisierung 14,5 Sekunden bzw. mit Häufigkeitsnetz 14,0 Sekunden). Durch die zusätzliche Verwendung eines Häufigkeitsnetzes (das auch diagnostische Informationen zeigt) konnte die Schwierigkeit Bayesianischer Informationen verringert werden, ersichtlich durch die verkürzte mediane Bearbeitungszeit (Median: 13,5 Sekunden).

Im Rahmen der statistischen Auswertung wurde auch ein lineares gemischtes Modell zur Schätzung der Zeit, die für eine richtige Antwort benötigt wurde, modelliert. Das Modell zeigt, dass die Art der Informationsrichtung ($p < 0,001$) und die Verwendung eines Häufigkeitsnetzes ($p < 0,001$) einen signifikanten Effekt auf die Bearbeitungszeiten für eine richtige Antwort hatten. Darüber hinaus gab es einen signifikanten Interaktionseffekt von Informationsrichtung \times Visualisierung ($p < 0,001$), was bedeutet, dass der Vorteil, der bei der

Verwendung des Häufigkeitsnetzes auf die Bearbeitungszeit gesehen wurde, davon abhängt, ob es sich um Bayesianische oder diagnostische Informationen handelte, die präsentiert wurden. Auch hier zeigte sich, dass die Reihenfolge der Aufgaben (d. h. ob die Aufgabe als erste, zweite, dritte oder vierte Aufgabe präsentiert wurde; $p < 0,001$) einen signifikanten Einfluss auf die Bearbeitungszeit hat. Ebenso führte der Kontext des Cushing-Syndroms dazu, dass die Proband*innen signifikant mehr Zeit benötigten, wobei Frauen signifikant mehr Zeit benötigten als Männer, ohne dass sich die berichteten Auswirkungen der Informationsrichtung oder Visualisierung änderten. Andere Variablen (wie das Alter) hatten keinen signifikanten Einfluss auf die Bearbeitungszeit.

4.5.3 Die diagnostische Effizienz

Sowohl die Richtigkeit als auch die Geschwindigkeit sind in der ärztlichen Kommunikation mit Patient*innen entscheidende Faktoren. Diese lassen sich in einen Score integrieren, der die diagnostische Effizienz angibt. Die diagnostische Effizienz wurde hierbei als Quotient der medianen Zeit für richtige Antworten durch den Anteil richtiger Antworten berechnet und ist in der Tabelle 6 in der rechten Spalte zu finden. Alternativ könnten auch Scores berechnet werden, die die mediane Zeit für eine beliebige Aufgabe durch den Anteil der richtigen Antworten dividieren; vgl. (35). Je niedriger der Score ist, desto effizienter ist die jeweilige Version zur Kommunikation, und umgekehrt. Der beste Wert für die diagnostische Effizienz wurde bei der Verwendung der Version mit diagnostischen Informationen ohne Visualisierung erzielt (20 Sekunden), während die Version mit Bayesianischen Informationen ohne Visualisierung sich als die ineffizienteste Art der Informationsmitteilung zeigte (17 Minuten, 30 Sekunden).

4.6 Subjektive Schwierigkeitsbeurteilung

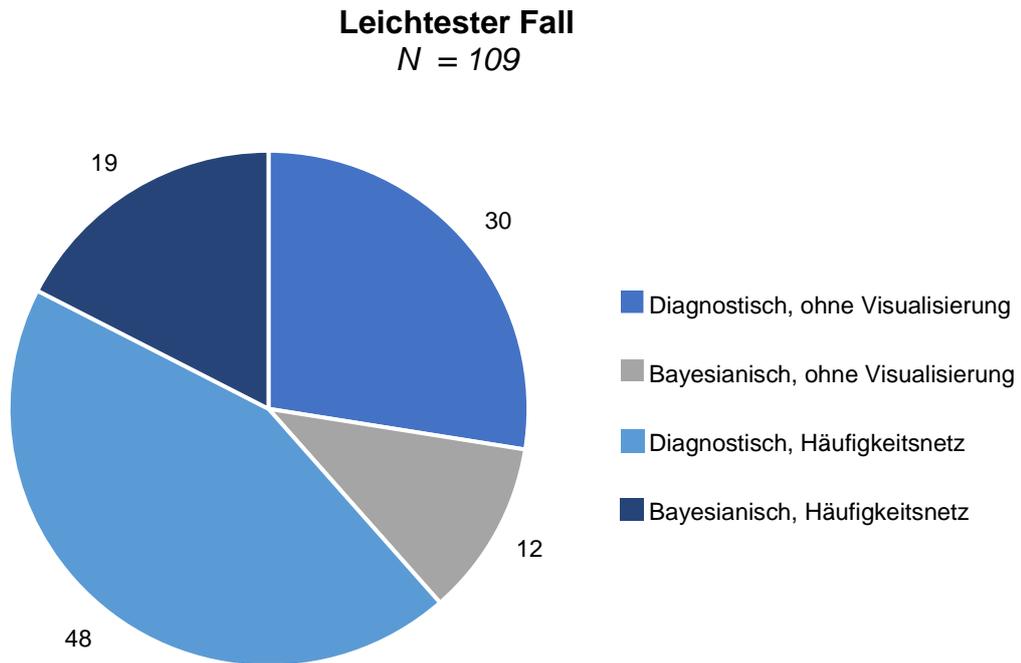


Abbildung 13: Subjektiv leichtester Fall

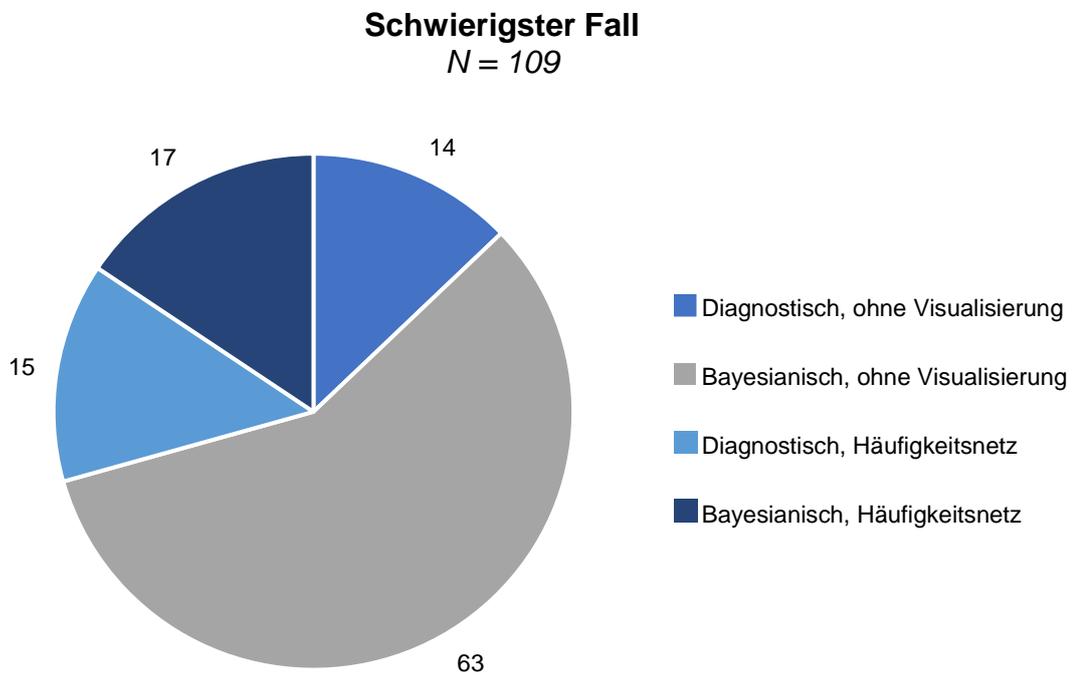


Abbildung 14: Subjektiv schwierigster Fall

Ergebnisse

Nach der Bearbeitung aller Fälle gaben die Proband*innen an, welchen Fall sie am leichtesten (siehe Abbildung 13) und welchen sie am schwierigsten (siehe Abbildung 14) fanden. Beim leichtesten Fall fand die deutliche Mehrheit (48%) den diagnostischen Fall mit Häufigkeitsnetz am leichtesten. Aber auch ohne Visualisierung hielten 30% der Proband*innen den diagnostischen für den leichtesten. Nur wenige Proband*innen wählten die Bayesianischen Fälle ohne Visualisierung bzw. mit Häufigkeitsnetz als den leichtesten aus. Bei der Wahl des schwierigsten Falls fiel das Ergebnis noch deutlicher aus. So waren sich 63% der Proband*innen sicher, dass der Bayesianische Fall ohne Visualisierung der schwierigste war. Die übrigen Proband*innen teilten sich zu fast gleichen Teilen auf die anderen drei Fälle auf (diagnostisch ohne Visualisierung, diagnostisch mit Häufigkeitsnetz, Bayesianisch mit Häufigkeitsnetz).

5 Diskussion

5.1 Zusammenfassung

In dieser Studie stand erstmals das Verständnis von Patient*innen bei Bayesianischen Aufgaben verglichen mit diagnostischen Aufgaben und mit Aufgaben mit visueller Unterstützung bei der Informationsvermittlung im Vordergrund. Insgesamt liegen in dieser Untersuchung die Daten von 109 volljährigen Proband*innen als potenzielle Patient*innen vor. Dabei ließ sich ein deutlich schlechteres Verständnis Bayesianischer Informationen im Gegensatz zu diagnostischen Informationen nachweisen. Dieses schwierige Verständnis Bayesianischer Informationen ließ sich durch eine Visualisierung mittels eines Häufigkeitsnetzes jedoch stark verbessern, da dadurch zusätzlich diagnostische Informationen gezeigt wurden. Bei diagnostischen Informationen hingegen zeigte sich durch die Verwendung des Häufigkeitsnetzes, wodurch Bayesianische Informationen ergänzt wurden, ein anderes Bild: Ein Teil der Proband*innen schnitt hiermit schlechter ab. Mehrmaliges Anschauen des Videos konnte in allen Fällen das Verständnis verbessern. Die Versionen mit niedrigen Ergebnisrichtigkeiten gingen in der Studie mit langen Zeiten zum Lösen der Aufgaben einher und vice versa. Ein richtiges Verständnis auf Seite der Patient*innen ist notwendig, um gute, gemeinsame Entscheidungen von Ärzt*innen und Patient*innen über die weiteren Behandlungsmöglichkeiten fällen zu können.

5.2 Diskussion der Methoden

Im Rahmen der Studie fiel die Entscheidung auf ein $2 \times 2 \times 4$ Design, welches bereits in anderen Studien erprobt und angewendet wurde (35). Neben vier verschiedenen Kontexten, welche jedoch keine Interessensfaktoren der Studie waren, musste sich somit auf zwei verschiedene Parameter beschränkt werden, da eine größere Auswahl die gute und differenzierte Auswertbarkeit der Ergebnisse beeinflusst hätte. Nach ausgiebiger Recherche bereits untersuchter sowie neuer Parameter wurden folgende Parameter für diese Studie herangezogen: die Informationsrichtung und die Informationsvisualisierung. Diese Parameter sind somit nicht die einzig möglichen, welche sich auf die Kommunikation statistischer Informationen auswirken können, sondern jene, die

am relevantesten und aufgrund Ihrer bisherigen spärlichen Studienlage als am interessantesten erschienen.

Der Einfluss einer Bayesianischen vs. diagnostischen Richtung wurde in diesem Modell erstmals untersucht. Bereits lange ist klar, dass Bayesianische Informationen Ärzt*innen Probleme bereiten, obwohl diese häufig mit ihnen im klinischen Alltag konfrontiert werden (84). Es galt zu untersuchen, ob auch Patient*innen Verständnisschwierigkeiten bei Bayesianischen Informationen haben. Um einen Vergleichswert zu haben, wie viele Proband*innen die Aufgaben lösen können, wenn bereits alle gefragten Zahlen gegeben sind, wurde in der Hälfte der Aufgaben eine diagnostische Informationsrichtung eingesetzt. Dadurch lässt sich zeigen, in wie vielen der Fälle vermutlich ein grundlegendes falsches Verständnis von statistischen Informationen zu Grunde liegt und bei welchem Prozentanteil tatsächlich die Bayesianischen Informationsrichtung eine Schwierigkeit darstellt. Da in vorangegangenen Studien Bayesianische Informationen eingesetzt wurden, war dieser Vergleich zu diagnostischen Informationen bisher nicht möglich gewesen. Unabhängig von der Richtung blieb die zu beantwortende Frage, die nach dem positiven prädiktiven Wert, dieselbe. Im medizinischen Arbeitsbereich ist diese vor allem für Patient*innen relevant, häufig aber nicht gegeben. Anzumerken ist, dass in beiden Fällen der Richtungen auch für das Lösen unwichtige Häufigkeiten gegeben wurden, die ein*e Ärzt*in im klinischen Alltag vermutlich ausgelassen hätte. Diese waren im Rahmen der Studie aber notwendig, um herauszufinden, ob die Patient*innen wirklich zwischen den verschiedenen Angaben unterscheiden können oder diese nur raten.

Der Parameter der Informationsvisualisierung war mit weiteren Entscheidungen verbunden, da verschiedene Visualisierungsmöglichkeiten zur Auswahl standen. Neben einfachen reichen diese bis zu komplizierten, welche aber gleichzeitig die beste Darstellung aller Zahlen gleichzeitig erlauben. Dabei sind spezifisch das Häufigkeitsnetz und der Doppelbaum zu nennen, welche zu den besten Visualisierungen für Bayesianische Informationen zählen. Auch letzterer wäre für die Studie denkbar gewesen, unterlag in seiner Darstellung und seinen Möglichkeiten aber dem Häufigkeitsnetz (61), weshalb dieses letztendlich in der Studie Anwendung fand. Ein Vergleich verschiedener Visualisierungen wurde bewusst nicht gewählt, da hierzu bereits mehrere Forschungen betreiben wurden

Diskussion

(59, 61, 63). Deshalb sollte sich auch hier auf einen neuen Aspekt konzentriert werden: der Vergleich einer visuellen Kommunikation zu einer rein auditiven Kommunikation. Denkt man an den Ablauf von ärztlichen Gesprächen in Kliniken und Praxen fällt schnell auf, dass rein auditive Informationsvermittlungen keine Seltenheit sind. Die Gründe hierfür können sich von einer erhofften Zeitersparnis über fehlende vorhandene Visualisierungen bis hin zu einem nicht vermuteten Nutzen erstrecken. Mithilfe des neuen Video-Designs konnte dieser Vergleich gut umgesetzt werden.

Für die Aufgaben wurden echte klinische Fälle verwendet. Neben der Kombination der Krankheiten mit den entsprechenden Tests haben auch die genannten Häufigkeiten jenen aus der Literatur entsprochen. Vor ihrem Einsatz wurden die Fälle zur Qualitätssicherung von Experten aus dem medizinischen und mathematischen Bereich bewertet und geprüft. Die Gütekriterien Objektivität, Reliabilität und Validität konnten für die Studie als erfüllt angenommen werden.

Der Studienablauf auf Casus ermöglichte ein für Proband*innen flexibles Teilnehmen, welches sehr realitätsnah gestaltet war. Nicht nur im Rahmen der medizinischen Lehre sondern auch für die Durchführung von Studie hat sich Casus als geeignet erwiesen (35). Anzumerken ist jedoch, dass es nicht hundertprozentig den klinischen Alltag widerspiegeln kann, da die Proband*innen nur potenzielle Patient*innen waren. Echte Patient*innen hätten sich möglicherweise bereits im Internet über den bei ihnen durchgeführten Test informiert und einen Wissensvorsprung bekommen können. Ebenso kann die Konzentration in der eigenen häuslichen Umgebung der Proband*innen von der der Patient*innen abweichen. Diese befinden sich bei Ärzt*innen in einem neuen Umfeld und haben vermutlich noch weitere Gedanken und Sorgen im Kopf, wodurch die kommunizierten Informationen schlechter als in der Studie aufgenommen werden könnten. Eine Inhomogenität des Untersuchungskollektivs wurde innerhalb der Studie versucht umzusetzen. Bei der Auswertung der soziodemografischen Daten, auf welche nachfolgend eingegangen wird, zeigte sich jedoch, dass die Inhomogenität im klinischen Alltag jedoch häufig anders ausfällt.

5.3 Diskussion der Ergebnisse

5.3.1 Die Proband*innen als potenzielle Patient*innen

In der Studie bestand eine gewisse Diskrepanz zwischen den Proband*innen und echten Patient*innen. Die Proband*innen der Studie waren in der Mehrheit jung und weiblich und hatten einen meist hohen Bildungsabschluss. Ursächlich hierfür könnte sein, dass viele der Studienausschreibungen in Universitäten, Bibliotheken und Kliniken verteilt wurden und besonders diese Altersgruppe hohes Interesse an der Teilnahme von Studien hat. In Praxen und Kliniken finden sich vor allem ältere Patient*innen wieder, da viele Krankheiten im Alter gehäuft auftreten. Ebenso sind von einer gemischteren Verteilung aller Bildungsschichten und einer gleichmäßigeren Verteilung zwischen Männern und Frauen auszugehen. Diese Unterschiede, vor allem der geringe Anteil an Proband*innen höheren Alters und niedrigerer Bildungsschichten, könnte sich auf das Verständnis Bayesianischer Informationen im klinischen Alltag auswirken. Jedoch ist dabei eher von einem schlechteren Verständnis auszugehen, wodurch die in der Studie aufgezeigten Möglichkeiten zur Verständnisverbesserung immer noch die gleichen bleiben und deren Anwendung aber noch mehr an Bedeutung gewinnt.

Da nur volljährige Personen an der Studie teilnehmen durften, lassen die Ergebnisse keine Rückschlüsse das Verständnis von Kindern und Jugendlichen ziehen. Da diese jedoch meist noch keine abgeschlossene Schulbildung haben, wären bei ihnen ebenso wie bei alten Patient*innen schlechtere Ergebnisse zu erwarten. Unter 18-Jährige besuchen aber in der Regel in Begleitung ihrer Erziehungsberechtigten ärztliche Gespräche, sodass es wichtig ist, den Erziehungsberechtigten die Informationen richtig zu vermitteln, sodass sie es an ihre Kinder weiterkommunizieren können.

Das anzugebende medizinische und mathematische Grundwissen wurde stark subjektiv beeinflusst, da die Proband*innen die Note freiwählen sollten ohne konkrete Richtwerte zu haben, was zum Beispiel „sehr gut“ oder „mangelhaft“ heißt. Allgemein war eine Tendenz zur Mitte zu verzeichnen. Bedenkt man, dass der Großteil der Proband*innen mindestens ein Abitur hat, sind möglicherweise die

Noten im Vergleich zu dem Bild der Gesamtbevölkerung hin zu schlechteren Werten verschoben, als dies in der Realität objektiv der Fall ist.

5.3.2 Der Einsatz diagnostischer Informationen ist von Vorteil

Die Studie hatte als Ziel zu klären, wie sich eine diagnostische Informationsrichtung gegenüber einer Bayesianischen auf das Verständnis statistischer Informationen auswirkt. Dabei hat sich gezeigt, dass statistische Informationen für potenzielle Patient*innen große Probleme verursachen: Nur 10% der Proband*innen waren in der Lage, Bayesianische Aufgaben zu lösen, die ihnen rein auditiv vermittelt wurden. Besonders auffällig war in dieser Studie die lange Zeit, die die Proband*innen benötigten, um Bayesianische Aufgaben ohne jegliche Visualisierung richtig zu lösen. Dies stellt ein Problem im klinischen Umfeld dar, wo Zeit eine begrenzte Ressource und ihre Minimierung somit essentiell ist. Durch die Ergebnisse der Studie konnte veranschaulicht werden, dass eine diagnostische Informationsrichtung immer von Vorteil war: Sie hatte deutlich höhere Lösungsraten (72%) mit geringen Zeiten zum Lösen der Aufgaben und geringeren Häufigkeiten des Videoanschauens zu Folge.

Die Gründe für die unterschiedlichen Lösungsraten zwischen den zwei Versionen der Informationsrichtung sind zwar nicht nachgewiesen, liegen aber auf der Hand: In beiden Aufgabenversionen wurde nach dem positiven prädiktiven Wert gefragt. In den diagnostischen Videos wurde der positive prädiktive Wert von einem*einer Ärzt*in aktiv genannt. Somit mussten die Proband*innen lediglich den richtigen Wert zwischen den verschiedenen Häufigkeiten herausfiltern, aber nichts mehr berechnen. Die diagnostischen Aufgaben waren also durch aufmerksames Zuhören und Wiedergeben lösbar. Bei den Bayesianischen Werten war der positive prädiktive Wert gefragt, aber nicht direkt im Video erwähnt. Stattdessen wurde die Sensitivität genannt, welches bei vielen Proband*innen zu kognitiven Überforderungen und Verwechslungen zwischen dem positive prädiktive Wert und der Sensitivität führte. Um den positiven prädiktiven Wert richtig zu ermitteln, mussten zwei genannte Häufigkeiten addiert werden oder, wenn gegeben, aus dem Häufigkeitsnetz abgelesen werden.

Diese algorithmisch komplexere Berechnung kann ebenso als Ursache für die 7-8-fach so langen Zeiten für richtiges Lösen Bayesianischer Aufgaben ohne

Visualisierung, verglichen den Zeiten der anderen Versionen, angenommen werden. Zudem kann dieser Zeitunterschied dadurch mitbeeinflusst worden sein, dass nur wenige Proband*innen diese Aufgabe lösen konnten, weshalb Ausreiser stärker zur Geltung kamen. Zusammenfassend wurde durch die langen Zeiten und die niedrigen Lösungsraten die Schwierigkeit dieser Bayesianischen Version ohne Visualisierung für die Proband*innen deutlich.

Durch die Erhebung der subjektiven Schwierigkeitseinschätzung hat sich gezeigt, dass die Proband*innen diesen Vorteil diagnostischer Informationen anstelle von Bayesianischen selbst auch richtig einschätzen konnten. Der Großteil der Proband*innen gab an, dass diagnostische Aufgaben die leichtesten gewesen seien, während Bayesianische, vorallem ohne Visualisierung, schwierig zu lösen waren.

Die Tatsache der Unterlegenheit der Bayesianischen Informationsrichtung stellt somit keine Überraschung dar, das Ausmaß aber schon. Bedenkt man nun, dass Angaben von Test und Diagnosen in der Medizin regelmäßig Bayesianisch erfolgen (30), lässt sich das Problem schnell erkennen: Kaum Patient*innen können diese Angaben und somit ihre Wahrscheinlichkeit erkrankt zu sein, richtig verstehen. Und wenn sie es doch richtig verstehen, brauchen sie dafür sehr viel Zeit. Mithilfe der Studie konnte gezeigt werden, dass die Verwendung diagnostischer Häufigkeiten ein erster guter Ansatz für eine effizientere Kommunikation ist.

5.3.3 Das Häufigkeitsnetz verringert die Schwierigkeit von Bayesianischen Informationen

Als Alternative zur auditiven Erklärung wurde in der Hälfte der Aufgaben in dieser Studie ein Häufigkeitsnetz zur Visualisierung verwendet. Bei den Bayesianischen Aufgaben führte diese Verwendung zu einem besseren Verständnis, was sich in einer höheren Lösungsrate (37%) widerspiegelte. Dies lässt sich dadurch erklären, dass bei den Aufgaben mit Häufigkeitsnetz beide gefragten Zahlen abgelesen werden konnten, während bei den Aufgaben ohne Häufigkeitsnetz die Proband*innen sich mehrere Zahlen merken und diese verrechnen mussten. Es zeigte sich somit das gleiche Bild wie in früheren Studien: Visualisierungen können Bayesianisches Denken unterstützen (35, 59, 60, 63). Jedoch wurden

Diskussion

dabei meist andere Visualisierungen verwendet und Daten zum Häufigkeitsnetz fehlten fast vollständig. Lediglich von einem Nutzen des Häufigkeitsnetzes im Vergleich zu Texten wird berichtet (61). Nach den Ergebnissen dieser Studie konnte der Nutzen des Häufigkeitsnetzes ebenso im Vergleich zu auditiver Informationsvermittlung belegt werden.

Auch was die Zeiten für richtige Lösungen angeht, scheint das Häufigkeitsnetz Bayesianische Informationen schneller verständlich zu machen. Die zuvor sehr langen Zeiten bei der Bayesianischen Version ohne Visualisierung wurden durch den Einsatz des Häufigkeitsnetzes stark reduziert und ähnelten dadurch den Bearbeitungszeiten der diagnostischen Aufgaben.

Interessanterweise konnten bei diagnostischen Fällen der Nutzen einer Visualisierung nicht nachgewiesen werden. Die Aufgaben ohne Häufigkeitsnetz konnten häufiger richtig gelöst werden, kam das Häufigkeitsnetz hinzu, ging die Lösungsrate sogar leicht auf 61% zurück. Es schien, als ob in diesem Fall die Visualisierung mehr verwirrt als geholfen habe. Dahinter könnte der Mechanismus einer kognitiven Überlastung stecken, dass Proband*innen sich umso schwieriger damit tun, die wichtigen Informationen herauszufiltern, je mehr Informationen sie bekommen (85).

Die Proband*innen überschätzen den Vorteil einer Visualisierung bei diagnostischen Informationen. In der Schwierigkeitsbeurteilung gab der Großteil der Proband*innen an, den diagnostischen Fall mit Häufigkeitsnetz leichter gefunden zu haben als den diagnostischen Fall ohne Visualisierung. Die Auswertung der Ergebnisrichtigkeiten widerlegte dies aber, indem gezeigt wurde, dass der diagnostische Fall ohne Visualisierung am häufigsten richtig gelöst werden konnte. Insofern ist ein gutes Abwägen und gezieltes Einsetzen von Häufigkeitsnetzen durch den*die Ärzt*in erforderlich.

Daraus lässt sich schlussfolgern, dass die Visualisierung durch Ärzt*innen im klinischen Alltag nur dann erfolgen sollte, wenn komplizierte statistische Informationen gegeben sind, die von den Patient*innen anderenfalls im Kopf weiterverarbeiten müssten. Bei leichten Angaben, welche direkt mündlich überbracht werden können, sollte auf eine Visualisierung verzichtet werden.

5.3.4 Mit Hilfe guter Kommunikationen können Patient*innen statistische Informationen angemessen verstehen

Das Verständnis statistischer Informationen hängt stark davon ab, wie sie vermittelt werden. Damit die Patient*innen gute Entscheidungen treffen können, sollten die Ärzt*innen versuchen, mehrere Parameter in ihrer Kommunikation zu optimieren. Nach den Ergebnissen dieser Studie ist die Übermittlung diagnostischer Informationen aus Gründen der Effizienz Bayesianischen Informationen vorzuziehen. Wenn aber nur Bayesianische Informationen zur Verfügung stehen, könnten diese von Ärzt*innen oder medizinischem Personal in diagnostische Informationen umgewandelt werden, bevor ein Gespräch mit den Patient*innen stattfindet. Sollte dies nicht möglich sein, hilft eine Visualisierung zur Erklärung der komplizierteren Bayesianischen Informationen.

Darüber hinaus scheint eine intensivere Auseinandersetzung mit den Aufgaben durch mehrmaliges Ansehen der Videos zu einem Lerneffekt zu führen. In dieser Studie zeigte sich, dass durch die Vermittlung mehrerer verschiedener Diagnosesituationen nacheinander die Lösungsraten deutlich erhöht werden konnten (Vergleich der Prozentsätze richtiger Antworten in Abbildung 9 und Abbildung 10 sowie in Abbildung 9 und Abbildung 11). Im verallgemeinerten linearen gemischten Modell nach mehrmaligem Videoanschauen war die Häufigkeit der Betrachtung des Videos jedoch kein signifikanter Faktor. Vermutlich haben die Proband*innen unserer Studie die Frage, wie oft sie das Video angeschaut haben, nicht ganz ehrlich beantwortet, denn viele von ihnen korrigierten die zuvor falsche Antwort nach mehrmaligem Anschauen des Videos und konnten dann die richtige Antwort geben. Ärzt*innen sollten im klinischen Alltag deshalb öfters Patient*innen die Möglichkeit einräumen, ihnen ihre Ergebnisse nochmals zu erläutern. Dabei könnte dieser Möglichkeit das Problem der mangelnden Zeit im klinischen Alltag gegenüberstehen. Ob ein richtiges Verständnis oder eine zügige Kommunikation wichtiger ist, müsste situationsbedingt abgewogen werden.

Zusammenfassend lässt sich sagen, dass Patient*innen bei guter Kommunikation durchaus in der Lage sind, Bayesianische Situationen zu lösen. Ärzt*innen sollten sich bemühen, dass Patient*innen die Ergebnisse ihrer Tests

richtig verstehen, sodass eine vernünftige, gemeinsame Entscheidung zwischen Ärzt*innen und Patient*innen über das weitere Vorgehen getroffen werden kann.

5.4 Limitationen

Einschränkend ist darauf hinzuweisen, dass als Visualisierungsmöglichkeit in der Studie stets nur das Häufigkeitsnetz verwendet wurde. Somit ist eine Verallgemeinerung der Studienergebnisse auf andere Visualisierungen nur bedingt möglich. Des Weiteren ergibt sich eine Limitation aus dem Proband*innenkollektiv. Wie in Kapitel 5.3.1. ausführlich beschrieben, könnten sich die Studienergebnisse der vorwiegend jungen und weiblichen Proband*innen mit hohem Bildungsabschluss von möglichen Ergebnissen von Patient*innen im klinischen Umfeld, welche häufig älter und verschiedener Bildungsschichten sind, unterscheiden.

5.5 Ausblick

Die in dieser Studie gefundenen Zusammenhänge zwischen der Informationsrichtung bzw. Informationsvisualisierung und der Ergebnisrichtigkeit bzw. Ergebnisgeschwindigkeit geben wichtige Hinweise darauf, wie Kommunikation im klinischen Alltag verbessert werden kann.

Es braucht mehr Studien, die konkrete Vorschläge zur Verbesserung der Risikokommunikation machen (siehe z. B. (86, 87)). Dabei soll die vorliegende Arbeit als Grundstein weiterer Forschung dienen.

An diese Studie anknüpfend wäre eine Replikation der Untersuchung mit anderen, ausgewählten Proband*innen bei gleichbleibendem Studiendesign möglich. Denkbare wäre eine gezielte Rekrutierung von älteren Personen, welche das klinische Patient*innenkollektiv besser widerspiegeln, oder sogar von echten Patient*innen. So könnten neue Erkenntnisse darüber gewonnen werden, ob und wie sich die Ergebnisrichtigkeiten unter noch realeren Bedingungen verändern würden.

Außerdem wären Adaptionen der Studie möglich. Eine Idee wäre beispielsweise, den Proband*innen in der zweiten Runde statt eines wiederholten Anschauens des Videos eine bessere Hilfe anzubieten, um ein richtiges Verständnis zu

Diskussion

erreichen. Bei den rein auditiven Versionen könnte folglich in der zweiten Runde eine Visualisierung mittels Häufigkeitsnetz angeboten werden und bei den Versionen, die bereits ein Häufigkeitsnetz enthielten, eine andere Form der Visualisierung. Alternativ könnten noch weitere Parameter, welche die ärztliche Kommunikation mit Patient*innen effektiver machen können, herausgearbeitet und verglichen werden.

Im nächsten Schritt ist eine Praxisapproximation anzustreben: Medizinstudierende und Ärzt*innen sollen Kommunikationstrainings erhalten, um statistische Informationen patient*innengerecht aufbereiten zu können. Anschließend gilt zu untersuchen, welchen Einfluss dieses Training auf das Verständnis von Patient*innen hat.

6 Literaturverzeichnis

1. Brose SF, Binder K, Fischer MR, Reincke M, Braun LT, Schmidmaier R. Bayesian versus diagnostic information in physician-patient communication: Effects of direction of statistical information and presentation of visualization. *PLOS ONE*. 2023;18(6):e0283947.
2. Wölfel T, Beltermann E, Lottspeich C, Vietz E, Fischer MR, Schmidmaier R. Medical ward round competence in internal medicine - an interview study towards an interprofessional development of an Entrustable Professional Activity (EPA). *BMC Med Educ*. 2016;16:174.
3. Vietz E, März E, Lottspeich C, Wölfel T, Fischer MR, Schmidmaier R. Ward round competences in surgery and psychiatry - a comparative multidisciplinary interview study. *BMC Med Educ*. 2019;19(1):137.
4. Lottspeich C, Braun LT, Fischer MR, Schmidmaier R. Simulated ward round training in the medical curriculum Munich. *GMS J Med Educ*. 2021;38(4):Doc75.
5. Nikendei C, Kraus B, Schrauth M, Briem S, Jünger J. Ward rounds: how prepared are future doctors? *Med Teach*. 2008;30(1):88-91.
6. Schmidmaier R, Holzer M, Angstwurm M, Nouns Z, Reincke M, Fischer MR. Using the Progress Test Medizin (PTM) for evaluation of the Medical Curriculum Munich (MeCuM). *GMS Z Med Ausbild*. 2010;27(5):Doc70.
7. Ackel-Eisnach K, Raes P, Hönikl L, Bauer D, Wagener S, Möltner A, et al. Is German Medical Education Research on the rise? An analysis of publications from the years 2004 to 2013. *GMS Z Med Ausbild*. 2015;32(3):Doc30.
8. Hahn E, Fischer M. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT) [National Competence-Based Learning Objectives for Undergraduate Medical Education (NKLM) in Germany: Cooperation of the Association for Medical Education (GMA) and the Association of Medical Faculties in Germany (MFT)]. 2009;26:Doc35.
9. Fischer MR, Bauer D, Mohn K. Finally finished! National Competence Based Catalogues of Learning Objectives for Undergraduate Medical Education (NKLM) and Dental Education (NKLZ) ready for trial. *GMS Z Med Ausbild*. 2015;32(3):Doc35.
10. Harendza S, Fischer MR, Fabry G. Quo vadis?--Medical education 2020 between politics and science. *GMS J Med Educ*. 2016;33(1):Doc12.
11. Röcker N, Lottspeich C, Braun LT, Lenzer B, Frey J, Fischer MR, et al. Implementation of self-directed learning within clinical clerkships. *GMS J Med Educ*. 2021;38(2):Doc43.

12. Bientzle M, Fissler T, Cress U, Kimmerle J. The impact of physicians' communication styles on evaluation of physicians and information processing: A randomized study with simulated video consultations on contraception with an intrauterine device. *Health Expect.* 2017;20(5):845-51.
13. Dowsett SM, Saul JL, Butow PN, Dunn SM, Boyer MJ, Findlow R, et al. Communication styles in the cancer consultation: preferences for a patient-centred approach. *Psychooncology.* 2000;9(2):147-56.
14. Fissler T, Bientzle M, Cress U, Kimmerle J. The Impact of Advice Seekers' Need Salience and Doctors' Communication Style on Attitude and Decision Making: A Web-Based Mammography Consultation Role Play. *JMIR Cancer.* 2015;1(2):e10.
15. Graugaard PK, Finset A. Trait anxiety and reactions to patient-centered and doctor-centered styles of communication: an experimental study. *Psychosom Med.* 2000;62(1):33-9.
16. Bientzle M, Griewatz J, Kimmerle J, Küppers J, Cress U, Lammerding-Koeppel M. Impact of Scientific Versus Emotional Wording of Patient Questions on Doctor-Patient Communication in an Internet Forum: A Randomized Controlled Experiment with Medical Students. *J Med Internet Res.* 2015;17(11):e268.
17. Gurmankin AD, Baron J, Hershey JC, Ubel PA. The role of physicians' recommendations in medical treatment decisions. *Med Decis Making.* 2002;22(3):262-71.
18. Mendel R, Traut-Mattausch E, Frey D, Bühner M, Berthele A, Kissling W, et al. Do physicians' recommendations pull patients away from their preferred treatment options? *Health Expect.* 2012;15(1):23-31.
19. Meinhardt AL, Eggeling M, Cress U, Kimmerle J, Bientzle M. The impact of a physician's recommendation and gender on informed decision making: A randomized controlled study in a simulated decision situation. *Health Expect.* 2021;24(2):269-81.
20. Eggeling M, Bientzle M, Cress U, Shiozawa T, Kimmerle J. The impact of physicians' recommendations on treatment preference and attitudes: a randomized controlled experiment on shared decision-making. *Psychol Health Med.* 2020;25(3):259-69.
21. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest.* 2007;8(2):53-96.
22. Gigerenzer G, Hoffrage U, Ebert A. AIDS counselling for low-risk clients. *AIDS Care.* 1998;10(2):197-211.
23. Bours MJ. Bayes' rule in diagnosis. *J Clin Epidemiol.* 2021;131:158-60.

24. Wirtz MA. Bayes'sches Theorem und der Basisratenfehler: Zum Vorhersagewert von Risikofaktoren und Screeningverfahren. *Rehabilitation (Stuttg)*. 2018;57(02):108-16.
25. Webb MPK, Sidebotham D. Bayes' formula: a powerful but counterintuitive tool for medical decision-making. *BJA Educ*. 2020;20(6):208-13.
26. Westbury CF. Bayes' rule for clinicians: an introduction. *Front Psychol*. 2010;1:192.
27. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *Science*. 2011;333(6048):1393-400.
28. Ghosh K, Crawford BJ, Pruthi S, Williams CI, Neal L, Sandhu NP, et al. Frequency format diagram and probability chart for breast cancer risk communication: a prospective, randomized trial. *BMC Womens Health*. 2008;8:18.
29. Ellis K, Cokely E, Ghazal S, Garcia-Retamero R. Do People Understand their Home HIV Test Results? Risk Literacy and Information Search. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2014;58:1323-7.
30. Prinz R, Feufel MA, Gigerenzer G, Wegwarth O. What Counselors Tell Low-Risk Clients About HIV Test Performance. *Curr HIV Res*. 2015;13(5):369-80.
31. Ellis KM, Brase GL. Communicating HIV Results to Low-Risk Individuals: Still Hazy After All These Years. *Curr HIV Res*. 2015;13(5):381-90.
32. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. *Medicine*. Communicating statistical information. *Science*. 2000;290(5500):2261-2.
33. Eddy DM. Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Tversky A, Kahneman D, Slovic P, editors. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press; 1982. p. 249-67.
34. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med*. 1998;73(5):538-40.
35. Binder K, Krauss S, Schmidmaier R, Braun LT. Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. *Advances in Health Sciences Education*. 2021;26(3):847-63.
36. Büchter T, Eichler A, Steib N, Binder K, Böcherer-Linder K, Krauss S, et al. How to Train Novices in Bayesian Reasoning. *Mathematics*. 2022;10(9):1558.
37. Böcherer-Linder K, Binder K, Büchter T, Eichler A, Krauss S, Steib N, et al. Communicating conditional probabilities in medical practice. In: Peters S, editor. *Bridging the Gap: Empowering and Educating Today's Learners in Statistics Proceedings of the Eleventh International Conference on*

- Teaching Statistics; Rosario (Argentina): International Association for Statistical Education; 2022.
38. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165(13):1493-9.
 39. Zwaan L, de Bruijne M, Wagner C, Thijs A, Smits M, van der Wal G, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med.* 2010;170(12):1015-21.
 40. Braun LT, Zwaan L, Kiesewetter J, Fischer MR, Schmidmaier R. Diagnostic errors by medical students: results of a prospective qualitative study. *BMC Med Educ.* 2017;17(1):191.
 41. Braun LT, Borrmann KF, Lottspeich C, Heinrich DA, Kiesewetter J, Fischer MR, et al. Guessing right - whether and how medical students give incorrect reasons for their correct diagnoses. *GMS J Med Educ.* 2019;36(6):Doc85.
 42. Braun LT, Lenzer B, Kiesewetter J, Fischer MR, Schmidmaier R. How case representations of medical students change during case processing - Results of a qualitative study. *GMS J Med Educ.* 2018;35(3):Doc41.
 43. Braun LT, Borrmann KF, Lottspeich C, Heinrich DA, Kiesewetter J, Fischer MR, et al. Scaffolding clinical reasoning of medical students with virtual patients: effects on diagnostic accuracy, efficiency, and errors. *Diagnosis (Berl).* 2019;6(2):137-49.
 44. Kiesewetter J, Ebersbach R, Tsalas N, Holzer M, Schmidmaier R, Fischer MR. Knowledge is not enough to solve the problems - The role of diagnostic knowledge in clinical reasoning activities. *BMC Med Educ.* 2016;16(1):303.
 45. Kiesewetter J, Ebersbach R, Görlitz A, Holzer M, Fischer MR, Schmidmaier R. Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions. *PLoS One.* 2013;8(8):e71486.
 46. Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, Norman G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf.* 2017;26(2):104-10.
 47. DA AL, Rotgans JI, Mamede S, Mahzari MM, Al-Ghamdi GA, Schmidt HG. Factors underlying suboptimal diagnostic performance in physicians under time pressure. *Med Educ.* 2018;52(12):1288-98.
 48. Lussier MT, Richard C. Doctor-patient communication. Time to talk. *Can Fam Physician.* 2006;52(11):1401-2.
 49. Committee on Diagnostic Error in Health C, Board on Health Care S, Institute of M, The National Academies of Sciences E, Medicine. Improving Diagnosis in Health Care In: Balogh EP, Miller BT, Ball JR, editors. *Improving Diagnosis in Health Care.* Washington (DC): National Academies Press (US) Copyright 2015 by the National Academy of Sciences. All rights reserved.; 2015.

50. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf.* 2013;22 Suppl 2(Suppl 2):ii21-ii7.
51. Wegwarth O, Gigerenzer G. Less is more: Overdiagnosis and overtreatment: evaluation of what physicians tell their patients about screening harms. *JAMA Intern Med.* 2013;173(22):2086-7.
52. Gigerenzer G, Hoffrage U. How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Judgement and decision making* : Vol 3, 114-153 (2009). 1995;102.
53. 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.
54. McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychol Bull.* 2017;143(12):1273-312.
55. Binder K KS, Bruckmaier G, Marienhagen J. T(H)REE STEPS TO IMPROVE BAYESIAN REASONING. In: Sorto MA WA, & Guyot L, editor. Looking back, looking forward Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan. Voorburg, The Netherlands: International Statistical Institute; 2018. p. 1-6.
56. Bruckmaier G BKK, S. Strategien beim Lösen statistischer Aufgaben – Eine Eyetracking-Studie zur visuellen Durchmusterung von Baumdiagrammen und Vierfeldertafeln. In: Frank A KSBK, editor. Beiträge zum Mathematikunterricht. Münster: WTM-Verlag; 2019. p. 1238.
57. Bruckmaier G, Binder K, Krauss S, Kufner HM. An Eye-Tracking Study of Statistical Reasoning With Tree Diagrams and 2 x 2 Tables. *Front Psychol.* 2019;10:632.
58. Hoffrage U, Krauss S, Martignon L, Gigerenzer G. Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Front Psychol.* 2015;6:1473.
59. Binder K, Krauss S, Bruckmaier G. Effects of visualizing statistical information – an empirical study on tree diagrams and 2 x 2 tables. *Frontiers in Psychology.* 2015;6(1186).
60. Binder K, Braun L. Erhöhung der ärztlichen Diagnoseschnelligkeit durch natürliche Häufigkeiten. In: Frank A, Krauss S, Binder K, editors. Beiträge zum Mathematikunterricht. Münster: WTM-Verlag; 2019. p. 1186.
61. 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).
62. 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 Pract.* 2022;7(1):23814683221086623.

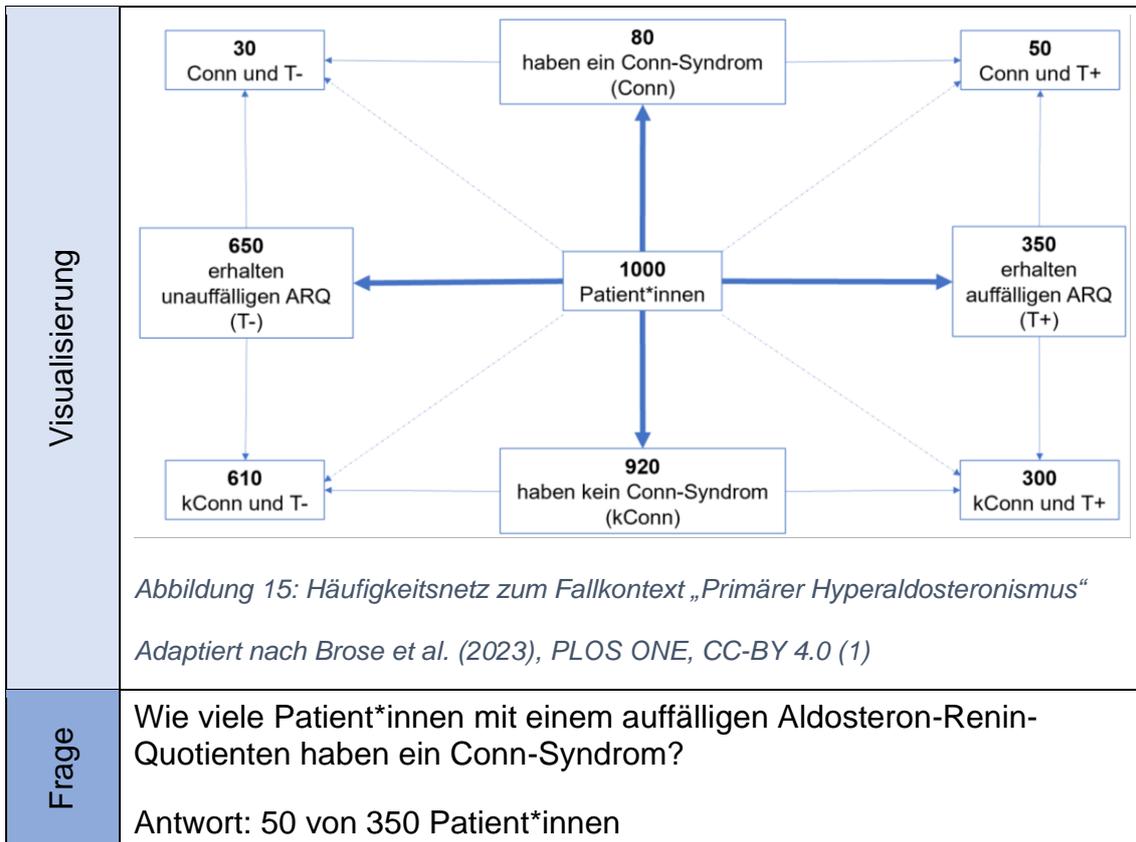
63. Binder K, Krauss S, Bruckmaier G. Welche Visualisierung unterstützt Bayesianisches Denken? In: Caluori F, Linneweber-Lammerskitten H, Streit C, editors. Beiträge zum Mathematikunterricht. Münster: WTM-Verlag; 2015. p. 160-3.
64. Weber P, Binder K, Krauss S. Why Can Only 24% Solve Bayesian Reasoning Problems in Natural Frequencies: Frequency Phobia in Spite of Probability Blindness. *Front Psychol.* 2018;9:1833.
65. 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.
66. Martignon L, Vitouch O, Takezawa M, Forster MR. Naive and Yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees. In: Hardman D, Macchi L, editors. *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*: John Wiley & Sons; 2003. p. 189-211.
67. Woike JK, Hoffrage U, Martignon L. Integrating and testing natural frequencies, naïve Bayes, and fast-and-frugal trees. *Decision.* 2017;4(4):234-60.
68. Böcherer-Linder K, Eichler A. How to Improve Performance in Bayesian Inference Tasks: A Comparison of Five Visualizations. *Front Psychol.* 2019;10:267.
69. Starns JJ, Cohen AL, Bosco C, Hirst J. A visualization technique for Bayesian reasoning. *Applied Cognitive Psychology.* 2019;33(2):234-51.
70. Eichler A, Böcherer-Linder K, Vogel M. Different Visualizations Cause Different Strategies When Dealing With Bayesian Situations. *Frontiers in Psychology.* 2020;11(1897).
71. Böcherer-Linder K, Eichler A. The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Frontiers in Psychology.* 2017;7.
72. Brase G. The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology.* 2013;26:81-97.
73. Brase GL. Pictorial representations in statistical reasoning. *Applied Cognitive Psychology.* 2009;23(3):369-81.
74. 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.
75. Mousavi SY, Low R, Sweller J. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology.* 1995;87(2):319-34.

76. Sweller J, Chandler P. Evidence for Cognitive Load Theory. *Cognition and Instruction*. 1991;8(4):351-62.
77. Schmidt D. [Ultrasound of Thyroid Nodules - Ultrasound Characteristics and Risk of Carcinoma]. *Laryngorhinootologie*. 2019;98(2):79-84.
78. Montori VM, Young WF, Jr. Use of plasma aldosterone concentration-to-plasma renin activity ratio as a screening test for primary aldosteronism. A systematic review of the literature. *Endocrinol Metab Clin North Am*. 2002;31(3):619-32, xi.
79. 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):1372-8.
80. Braun LT, Riester A, Oßwald-Kopp A, Fazel J, Rubinstein G, Bidlingmaier M, et al. Toward a Diagnostic Score in Cushing's Syndrome. *Front Endocrinol (Lausanne)*. 2019;10:766.
81. Fischer MRG, Aulinger B, Baehring T. Computer-based-Training (CBT): Fallorientiertes Lernen am PC mit dem CASUS/ProMediWeb-System. *Dtsch Med Wochenschr*. 1999;124(46):1401-.
82. Team RC. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013.
83. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1 - 48.
84. Gigerenzer G. HIV screening: helping clinicians make sense of test results to patients. *Bmj*. 2013;347:f5151.
85. Jona K, Gerber A. MyOncofertility.org: a web-based patient education resource supporting decision making under severe emotional and cognitive overload. *Cancer Treat Res*. 2010;156:345-61.
86. Brick C, McDowell M, Freeman ALJ. Risk communication in tables versus text: a registered report randomized trial on 'fact boxes'. *R Soc Open Sci*. 2020;7(3):190876.
87. Ellermann C, McDowell M, Schirren CO, Lindemann AK, Koch S, Lohmann M, et al. Identifying content to improve risk assessment communications within the Risk Profile: Literature reviews and focus groups with expert and non-expert stakeholders. *PLoS One*. 2022;17(4):e0266800.
88. Brose SF, Binder K, Fischer MR, Braun LT, Schmidmaier R. Bayesianische vs. diagnostische ärztliche Informationsvermittlung an Patient*innen – Der Einfluss von Richtung der statistischen Information und Visualisierung. *Osteologie*. 2022;31(03):V 2.5.

7 Anhang

Tabelle 7: Die Aufgabenversionen zum Fallkontext "Primärer Hyperaldosteronismus"

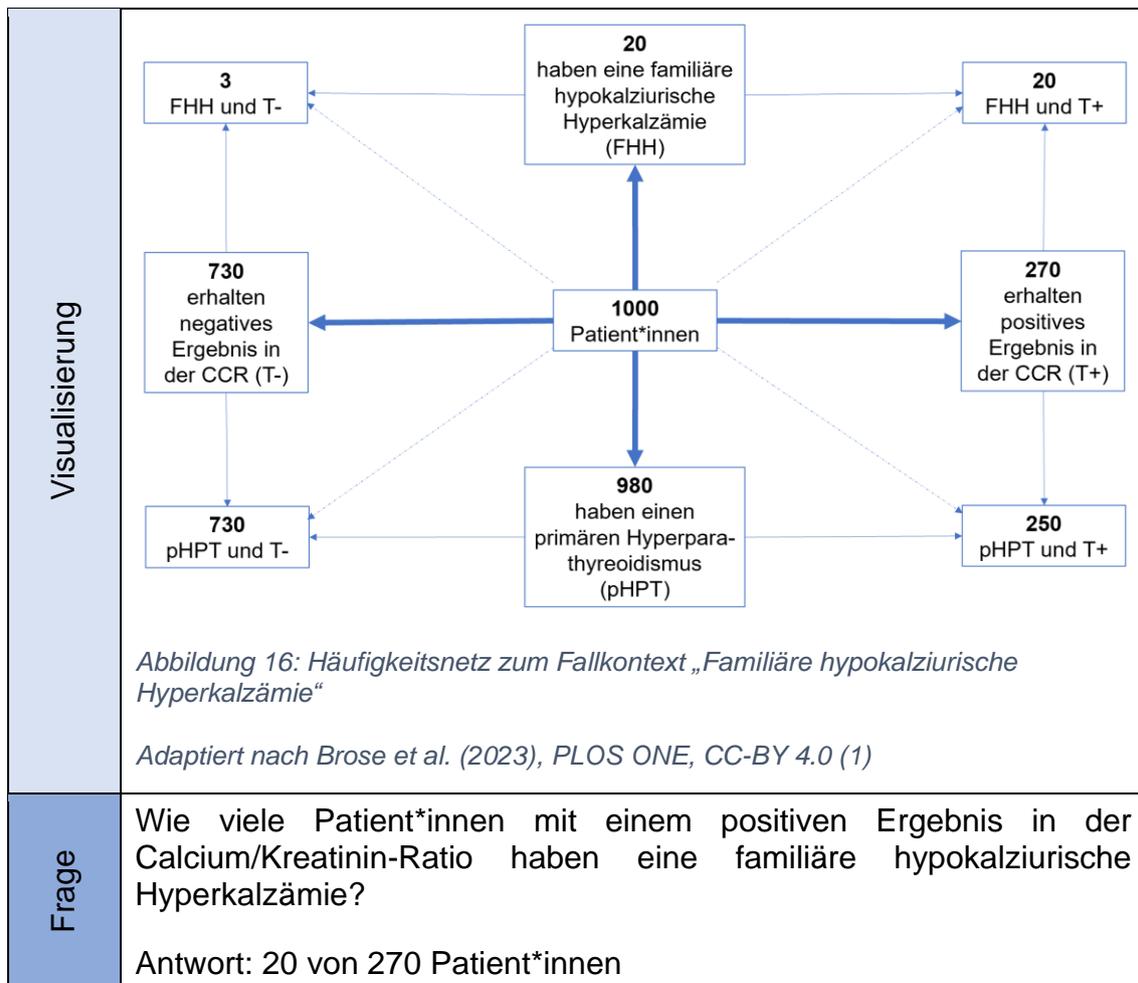
	Version 1	Version 2	Version 3	Version 4
	Diagnostisch		Bayesianisch	
Einleitung	<p>Bei Ihnen wurde ein Bluthochdruck diagnostiziert. Gestern wurde Ihnen Blut abgenommen, um die Ursache Ihres Bluthochdrucks abzuklären. Es wurde der Aldosteron-Renin-Quotient bestimmt, um den Verdacht eines Conn-Syndroms erhärten oder verwerfen zu können. Ein auffälliges Testergebnis würde für ein Conn-Syndrom sprechen. Ich habe Sie heute nochmals eingeladen, um gemeinsam Ihre Untersuchungsergebnisse zu besprechen.</p> <p>Leider muss ich Ihnen mitteilen, dass sich in der Untersuchung bei Ihnen ein auffälliger Aldosteron-Renin-Quotient gezeigt hat. Ich möchte Ihnen nun erklären, was ein auffälliger Untersuchungsbefund genau bedeutet.</p> <p>(In diesem Häufigkeitsnetz wurden 1000 Patient*innen mit Bluthochdruck auf das Vorliegen eines auffälligen Aldosteron-Renin-Quotienten sowie eines Conn-Syndroms untersucht. Die beiden Kriterien eines auffälligen bzw. unauffälligen Aldosteron-Renin-Quotienten und von einem bzw. keinem Conn-Syndrom sind hierbei sowohl einzeln als auch in Kombination visualisiert.)</p>			
Informationsrichtung	<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 350 Patient*innen einen auffälligen Aldosteron-Renin-Quotienten. • Von diesen 350 Patient*innen mit auffälligem Quotienten haben 50 Patient*innen tatsächlich ein Conn-Syndrom. • Andererseits haben von 650 Patient*innen mit einem unauffälligen Ergebnis trotzdem 30 Patient*innen ein Conn-Syndrom. 		<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 80 Patient*innen ein Conn-Syndrom. • Von diesen 80 Patient*innen mit diagnostiziertem Conn-Syndrom haben 50 Patient*innen einen auffälligen Aldosteron-Renin-Quotienten. • Andererseits haben von 920 Patient*innen, die kein Conn-Syndrom haben, trotzdem 300 Patient*innen einen auffälligen Aldosteron-Renin-Quotienten. 	
Informations-visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung



Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Tabelle 8: Die Aufgabenversionen zum Fallkontext "Familiäre hypokalziurische Hyperkalzämie"

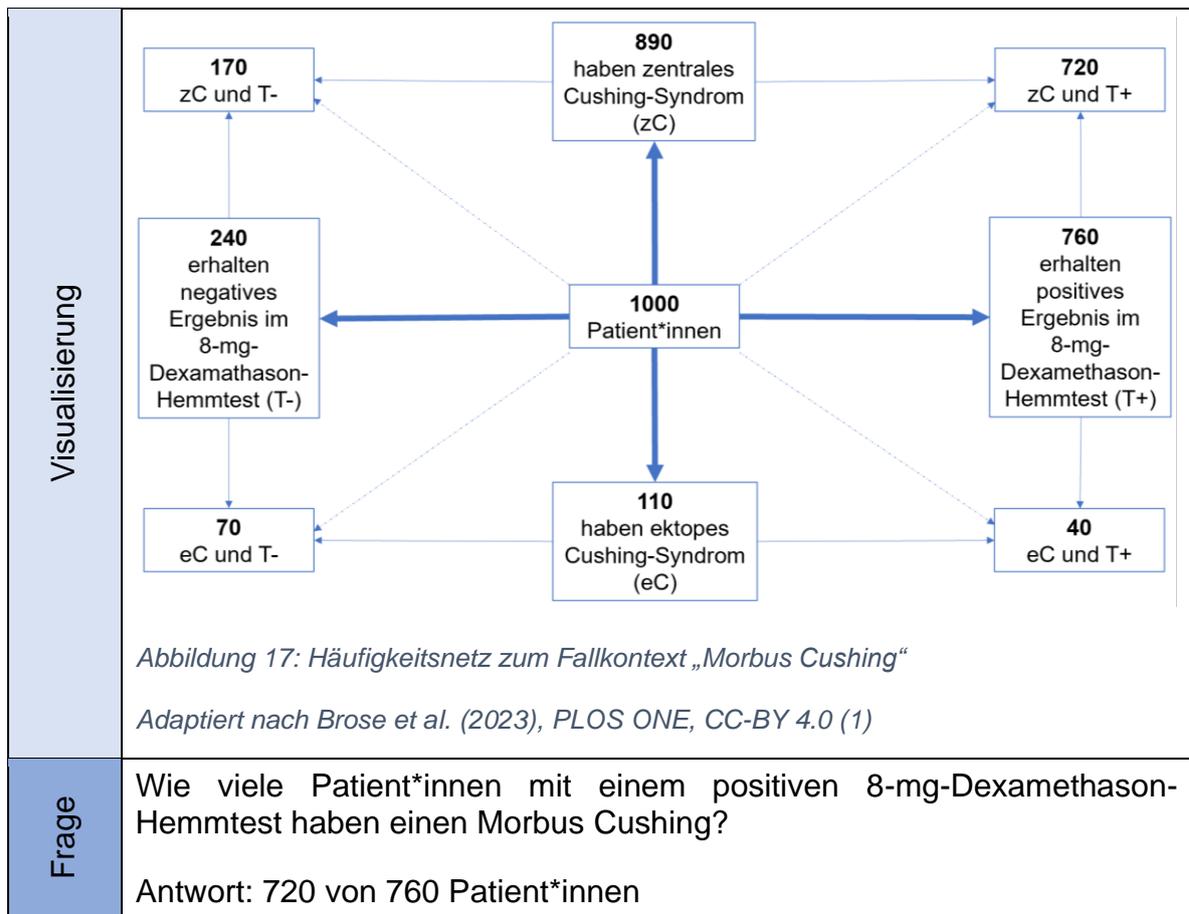
	Version 1	Version 2	Version 3	Version 4
	Diagnostisch		Bayesianisch	
Einleitung	<p>Gestern haben Sie 24h-Sammelurin abgegeben, wodurch man die Calcium/Kreatinin-Ratio im Urin bestimmen konnte. Hiermit sollte getestet werden, ob bei Ihnen eine familiäre hypokalziurische Hyperkalzämie oder ein primärer Hyperparathyreoidismus vorliegt. Ein positives Testergebnis würde für eine familiäre hypokalziurische Hyperkalzämie sprechen. Ich habe Sie heute nochmals eingeladen, um gemeinsam Ihre Untersuchungsergebnisse zu besprechen.</p> <p>In der Untersuchung hat sich ein positives Testergebnis gezeigt. Ich möchte Ihnen nun erklären, was ein positiver Untersuchungsbefund genau bedeutet.</p> <p>(In diesem Häufigkeitsnetz wurden 1000 Patient*innen auf das Vorliegen eines positiven bzw. negativen Ergebnisses in der Calcium/Kreatinin-Ratio sowie einer familiären hypokalziurischen Hyperkalzämie bzw. eines primären Hyperparathyreoidismus untersucht. Die beiden Kriterien des Testergebnisses und der Krankheiten sind hierbei sowohl einzeln als auch in Kombination visualisiert.)</p>			
Informationsrichtung	<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 270 Patient*innen ein positives Ergebnis in der Calcium/Kreatinin-Ratio. • Von diesen 270 Patient*innen mit positivem Ergebnis haben 20 Patient*innen tatsächlich eine familiäre hypokalziurische Hyperkalzämie. • Andererseits haben von 730 Patient*innen mit einem negativen Ergebnis trotzdem 3 Patient*innen eine familiäre hypokalziurische Hyperkalzämie. 		<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 20 Patient*innen eine familiäre hypokalziurische Hyperkalzämie. • Von diesen 20 Patient*innen mit einer diagnostizierten familiären hypokalziurischen Hyperkalzämie haben alle 20 Patient*innen ein positives Ergebnis in der Calcium/Kreatinin-Ratio im Urin. • Andererseits haben von 980 Patient*innen, die einen primären Hyperparathyreoidismus haben, trotzdem 250 Patient*innen ein positives Ergebnis in der Calcium/Kreatinin-Ratio. 	
Informations-visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung



Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Tabelle 9: Die Aufgabenversionen zum Fallkontext "Morbus Cushing"

	Version 1	Version 2	Version 3	Version 4
	Diagnostisch		Bayesianisch	
Einleitung	<p>Kürzlich wurde bei Ihnen der 8-mg-Dexamethason-Hemmtest durchgeführt. Dies ist ein Test, anhand dessen man unterscheiden kann, ob es sich bei Ihrem bereits bestätigten Hyperkortisolismus um einen Morbus Cushing oder um ein ektopes Cushing-Syndrom handelt. Hierfür wurde Ihnen an zwei aufeinander folgenden Morgenden Blut abgenommen, wobei Sie am Abend des ersten Tages eine Tablette einnahmen. Ein positives Testergebnis würde für einen Morbus Cushing sprechen. Heute möchten wir gemeinsam Ihre Untersuchungsergebnisse besprechen.</p> <p>In der Untersuchung hat sich ein positives Testergebnis gezeigt. Ich möchte Ihnen nun erklären, was ein positiver Untersuchungsbefund genau bedeutet.</p> <p>(In diesem Häufigkeitsnetz wurden 1000 Patient*innen mit Hyperkortisolismus auf das Vorliegen eines positiven bzw. negativen 8-mg-Dexamethason-Hemmtests sowie eines Morbus Cushings bzw. ektopen Cushing-Syndroms untersucht. Die beiden Kriterien des Testergebnisses und der Krankheiten sind hierbei sowohl einzeln als auch in Kombination visualisiert.)</p>			
Informationsrichtung	<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 760 Patient*innen ein positives Ergebnis im 8-mg-Dexamethason-Hemmtest. • Von diesen 760 Patient*innen mit positivem Ergebnis haben 720 Patient*innen tatsächlich einen Morbus Cushing. • Andererseits haben von 240 Patient*innen mit einem negativen Ergebnis trotzdem 170 Patient*innen einen Morbus Cushing. 		<ul style="list-style-type: none"> • Von 1000 Patient*innen haben 890 Patient*innen einen Morbus Cushing. • Von diesen 890 Patient*innen mit diagnostiziertem Morbus Cushing haben 720 Patient*innen ein positives Ergebnis im 8-mg-Dexamethason-Hemmtest. • Andererseits haben von 110 Patient*innen, die ein ektopes Cushing-Syndrom haben, trotzdem 40 Patient*innen ein positives Ergebnis im 8-mg-Dexamethason-Hemmtest. 	
Informations-visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung	Häufigkeitsnetz (s. u.)	Ohne Visualisierung



Adaptiert nach Brose et al. (2023), PLOS ONE, CC-BY 4.0 (1)

Datenverfügbarkeit

Alle Rohdaten dieser Studie sind verfügbar unter:

<https://www.doi.org/10.17605/OSF.IO/FZNVWG>

			
<h1>STUDIEN- TEILNEHMER GESUCHT</h1>			
<p>Gesucht werden Personen ab 18 Jahre. Keine Vorkenntnisse erforderlich.</p> <p>Löse Verständnisfragen online auf CASUS!</p> <p>Thema der Studie: Kommunikation statistischer Informationen an Patienten</p>			
<p>Dauer: einmalig 30-60 Minuten</p> <p>Termin: von Zuhause aus jederzeit möglich</p> <p>Aufwandsentschädigung: 10€</p>			
<p>Zur Anmeldung schreibe eine E-Mail mit deinem Namen an: Frederike.Brose@campus.lmu.de</p>			

Abbildung 18: Flyer der Studie



Medizinische Klinik und Poliklinik IV
Lehrstuhl: Endokrinologie und Diabetologie
Direktor: Prof. Dr. med. Martin Reincke

Medizinische Klinik und Poliklinik IV, Direktion
Ziemssenstraße 1, 80336 München

Leitung / Direktor

Prof. Dr. med. Martin Reincke

Ihr Ansprechpartner

Dr. med.
Leah Braun
Assistenzärztin der Klinik
Tel. +49 89 4400-57333
Fax +49 89 4400-57339
Leah.Braun@
med.uni-muenchen.de
www.lmu-klinikum.de

Informationsschrift

Sehr geehrte Probanden und Probandinnen,

Wir möchten Sie fragen, ob Sie an einer wissenschaftlichen Studie teilnehmen möchten. Im klinischen Alltag sind Mediziner oft mit Bayes'schen-Aufgaben konfrontiert. Mit dem Satz von Bayes lassen sich bedingte Wahrscheinlichkeiten berechnen. Im medizinischen Kontext handelt es sich hierbei zum Beispiel um Situationen, in denen ein Arzt erläutern muss, wie wahrscheinlich es ist, dass ein positives Testergebnis auch wirklich bedeutet, dass der Patient die Krankheit hat.

Ziel der vorliegenden Studie ist es herauszufinden, wie statistische Informationen präsentiert werden müssen, damit Mediziner richtig und schnell eine Entscheidung treffen können. Das Ergebnis der Studie soll langfristig Einfluss auf die Präsentationsform von Lehrinhalten haben, bei denen Bayes'sche Entscheidungen gefällt werden müssen.

Die Studie wird an der Medizinischen Klinik und Poliklinik IV des Klinikums der LMU München durchgeführt und von Prof. Dr. med. Ralf Schmidmaier, MME geleitet. Es werden insgesamt ca. 60 Probanden an der Studie teilnehmen.

Jeder Proband soll auf der Online-Lernplattform CASUS 4 verschiedene Fallbeispiele bearbeiten. Die Fallbeispiele werden zunächst anhand eines Videos eingeführt, in welchem ein Arzt oder eine Ärztin die Ergebnisse eines klinischen Tests erläutert. Nach jedem Video soll beantwortet werden, wie wahrscheinlich es ist, an einer bestimmten Erkrankung zu leiden, wenn ein bestimmter Test positiv ausfällt. Die Zeit pro Fall oder die gesamte Bearbeitungszeit der Fälle ist nicht limitiert, wird aber aufgezeichnet. Der Zeitaufwand für die Bearbeitung aller Fallbeispiele liegt insgesamt bei circa einer Stunde.

Im Rahmen der Studie werden ausschließlich Daten in anonymisierter Form erfasst und ausgewertet. Dabei handelt es sich um folgende Daten: soziodemographische Basisdaten, wie Alter, Geschlecht, Studienabschnitt, sowie die Ergebnisse und Bearbeitungszeit der Fallbeispiele. Ein Rückschluss auf die bearbeitende Person ist nicht möglich. Jeder Proband erhält vor Durchführung der CASUS Fallbeispiele einen individuellen Zugangscode per Zufallsprinzip zugeteilt. Es wird keine Liste über die Zuteilung von Person und Zugangscode geführt. Bei der Aushändigung der Zugangscode wird jeder Proband über das Procedere

Vorstand

Ärztlicher Direktor:
Prof. Dr. med. Karl-Walter Jauch
(Vorsitz)

Kaufmännischer Direktor:
Markus Zendler

Pflegedirektor:
Marcus Huppertz

Vertreter der Medizinischen Fakultät:
Prof. Dr. med. dent. Reinhard Hickel
(Dekan)

Institutionskennzeichen:
260 914 050

Umsatzsteuer-ID:
DE813536017

Das Klinikum der Universität
München ist eine Anstalt des
Öffentlichen Rechts

Gemeinsam. Fürsorglich. Wegweisend.

Abbildung 19: Informationsschrift und Einwilligungserklärung (Seite 1/6)

der Anonymisierung aufgeklärt und kann die Einwilligung unterschreiben. Aufklärung und Aushändigung des Zugangscodes erfolgen in einem persönlichen Gespräch mit jedem Probanden.

Die Teilnahme an dieser Studie ist freiwillig. Sie werden nur dann einbezogen, wenn Sie dazu schriftlich Ihre Einwilligung erklären. Sofern Sie nicht an der Studie teilnehmen oder später aus ihr ausscheiden möchten, entstehen Ihnen dadurch keine Nachteile. Sie können jederzeit, auch ohne Angabe von Gründen, Ihre Einwilligung mündlich oder schriftlich widerrufen.

Wenn Sie Ihre Einwilligung widerrufen, werden keine weiteren Daten mehr erhoben. Die bis zum Widerruf erfolgte Datenverarbeitung bleibt jedoch rechtmäßig.

Sollten Sie im Verlauf der Studie Ihre Einwilligung widerrufen, werden Sie gefragt, ob Sie mit der anonymisierten Weiterverwendung Ihrer Daten und ggf. Proben einverstanden sind oder ob eine Löschung der Daten bzw. Vernichtung der Proben erfolgen soll.

Die Studie wurde der zuständigen Ethikkommission vorgelegt. Sie hat keine Einwände erhoben.

Sind mit der Datenverarbeitung Risiken verbunden?

Bei jeder Erhebung, Speicherung, Nutzung und Übermittlung von Daten bestehen Vertraulichkeitsrisiken (z.B. die Möglichkeit, die betreffende Person zu identifizieren). Diese Risiken lassen sich nicht völlig ausschließen und steigen, je mehr Daten miteinander verknüpft werden können. Der Initiator der Studie versichert Ihnen, alles nach dem Stand der Technik Mögliche zum Schutz Ihrer Privatsphäre zu tun und Daten nur an Stellen weiterzugeben, die ein geeignetes Datenschutzkonzept vorweisen können. Medizinische Risiken sind mit der Datenverarbeitung nicht verbunden.

Risiko-Nutzen-Abwägung

Es entsteht kein Eigennutzen für die Studienteilnehmer. Da im Rahmen unserer Studie nur Daten erhoben werden, sind mit der Teilnahme keine medizinischen Risiken verbunden.

Die Ergebnisse dieser Studie können dazu beitragen, die Arzt-Patienten-Kommunikation im Hinblick auf die Kommunikation statistischer Informationen zu verbessern.

Information über Versicherungsschutz

Wir möchten Sie darüber aufklären, dass keine Wegeversicherung abgeschlossen wurde, da die Durchführung der Studie bei Ihnen vor Ort an ihrem Endgerät stattfindet und für Sie somit keine Wege entstehen.

Datenschutz

- Rechtsgrundlage für die Datenverarbeitung ist Ihre freiwillige Einwilligung (Art. 6 Abs. 1 Buchst. c) DSGVO)
- Der Verantwortliche für die Datenverarbeitung ist: **Klinikum der Universität München, Medizinische Klinik und Poliklinik IV, Ziemssenstr. 1, 80336 München**

Die Daten werden zu jeder Zeit vertraulich behandelt. Die Daten werden **in irreversibel anonymisierter Form** an die Initiatorin der Studie, **Dr. med. Leah Braun**, bzw. von ihr beauftragte Stellen zum Zweck der wissenschaftlichen Auswertung und Verbesserung der Patientenversorgung und Lehre weitergeleitet. Zugriff auf die personenbezogenen Daten haben nur die Studienleiter, **Frau Dr. med. Leah Braun** und **Herr Professor Dr. med. Ralf Schmidmaier, MME** sowie vom Studienleiter zur Verschwiegenheit verpflichtete Mitglieder des Studienteams, **Frau Sarah Frederike Brose**.

Die Daten werden **10 Jahre** nach Beendigung oder Abbruch der Studie auf dem Webserver der Medizinischen Klinik und Poliklinik IV aufbewahrt. Sie sind gegen unbefugten Zugriff gesichert. Sie werden gelöscht, wenn sie nicht mehr benötigt werden. Spätestens nach **10 Jahren** werden sie gelöscht.

Die Daten werden zu jeder Zeit (auch im Falle einer Veröffentlichung der Studienergebnisse) vertraulich behandelt.

Gemeinsam. Fürsorglich. Wegweisend.

Sind mit der Datenverarbeitung Risiken verbunden?

Bei jeder Erhebung, Speicherung, Nutzung und Übermittlung von Daten bestehen Vertraulichkeitsrisiken (z.B. die Möglichkeit, die betreffende Person zu identifizieren). Diese Risiken lassen sich nicht völlig ausschließen und steigen, je mehr Daten miteinander verknüpft werden können. Der Initiator der Studie versichert Ihnen, alles nach dem Stand der Technik Mögliche zum Schutz Ihrer Privatsphäre zu tun und Daten nur an Stellen weiterzugeben, die ein geeignetes Datenschutzkonzept vorweisen können. Medizinische Risiken sind mit der Datenverarbeitung nicht verbunden. Es wird ausdrücklich darauf hingewiesen, dass keine Ihrer Angaben aus der Studie an Vorgesetzte o.ä. weitergegeben werden.

Kann ich meine Einwilligung widerrufen?

Sie können Ihre jeweilige Einwilligung jederzeit ohne Angabe von Gründen schriftlich oder mündlich widerrufen, ohne dass Ihnen daraus ein Nachteil entsteht. Wenn Sie Ihre Einwilligung widerrufen, werden keine weiteren Daten mehr erhoben. Die bis zum Widerruf erfolgte Datenverarbeitung bleibt jedoch rechtmäßig.

Welche weiteren Rechte habe ich bezogen auf den Datenschutz?

Das Recht auf Widerruf der Einwilligung und Löschung der erhobenen Daten kann nur bis zur irreversiblen Anonymisierung geltend gemacht werden.

Es besteht ein Beschwerderecht bei der Datenschutz-Aufsichtsbehörde:

Für die LMU und das Klinikum ist zuständig:

Bayerischer Landesbeauftragter für den Datenschutz (BayLfD) Postanschrift: Postfach 22 12 19, 80502

München Hausanschrift: Wagnmüllerstr. 18, 80538 München

Tel.: 089 212672-0

Fax: 089 212672-50

Behördlicher Datenschutzbeauftragter der des Klinikums der Universität München:

Gerhard Meyer

Pettenkoferstraße 8

80336 München

E-Mail: gerhard.meyer@med.uni-muenchen.de

Ansprechpartner für Fragen zur Studie

Wenn Sie Fragen zu dieser Studie haben, wenden Sie sich bitte an:

Sarah Frederike Brose

E-Mail: Frederike.Brose@campus.lmu.de

Gemeinsam. Fürsorglich. Wegweisend.



Medizinische Klinik und Poliklinik IV
Lehrstuhl: Endokrinologie und Diabetologie
Direktor: Prof. Dr. med. Martin Reincke

Seite 4/6

Einwilligungserklärung

Titel der Studie: Kommunikation statistischer Informationen an Patienten

Name des Studienteilnehmers in Druckbuchstaben: _____

- Ich bin von Frau Brose über Wesen, Bedeutung und Tragweite der Studie sowie die sich für mich daraus ergebenden Anforderungen aufgeklärt worden. Ich habe darüber hinaus den Text der Studienteilnehmersaufklärung und dieser Einwilligungserklärung gelesen.
- Ich hatte ausreichend Zeit, Fragen zu stellen und mich zu entscheiden. Aufgetretene Fragen wurden mir von den Studienverantwortlichen beantwortet.
- Ich weiß, dass ich meine freiwillige Mitwirkung jederzeit beenden kann, ohne dass mir daraus Nachteile entstehen.

Ich erkläre mich bereit, an der Studie teilzunehmen.

1. Ich willige ein, dass personenbezogene Daten über mich, insbesondere Alter, Geschlecht und Ergebnisse der CASUS-Fall-Bearbeitung, wie in der Informationsschrift beschrieben erhoben und in Papierform sowie auf elektronischen Datenträgern in der Medizinischen Klinik und Poliklinik IV aufgezeichnet werden.

Soweit erforderlich, dürfen die erhobenen Daten in irreversibel anonymisierter Form weitergegeben werden:

- a) an Prof. Dr. med. Ralf Schmidmaier, MME oder von diesem beauftragte Stellen zum Zweck der wissenschaftlichen Auswertung,
 - b) im Falle unerwünschter Ereignisse: an die jeweils zuständige Ethik-Kommission und zuständige Behörden sowie von dieser an die Europäische Datenbank.
2. Ich bin darüber aufgeklärt worden, dass ich meine Einwilligung jederzeit widerrufen kann. Im Falle des Widerrufs werden keine weiteren Daten mehr erhoben. Ich kann in diesem Fall die Löschung der Daten verlangen.
 3. Ich willige ein, dass die Daten nach Beendigung oder Abbruch der klinischen Prüfung maximal 10 Jahre aufbewahrt werden.

Kontaktdaten der Studienverantwortlichen:

Dr. med. Leah Braun (Studienleitung, Betreuerin)

Assistenzärztin der Klinik

Medizinische Klinik und Poliklinik IV

Klinikum der Universität München

Ziemssenstr. 1

80336 München

E-Mail: Leah.Braun@med.uni-muenchen.de

Telefon: 089-4400-57333

Fax: 089/4400-57339

Gemeinsam. Fürsorglich. Wegweisend.

Abbildung 22: Informationsschrift und Einwilligungserklärung (Seite 4/6)



Medizinische Klinik und Poliklinik IV
Lehrstuhl: Endokrinologie und Diabetologie
Direktor: Prof. Dr. med. Martin Reincke

Seite 5/6

Prof. Dr. med. Ralf Schmidmaier, MME (Studienleitung, Doktorvater)

Stellvertretender Direktor, leitender Oberarzt
Klinikum der Universität München
Ziemssenstr. 1
80336 München
E-Mail: Ralf.schmidmaier@med.uni-muenchen.de
Telefon: 089/4400-53180
Fax: 089/4400-54428

Sarah Frederike Brose (Doktorandin)

E-Mail: Frederike.Brose@campus.lmu.de

Gemeinsam. Fürsorglich. Wegweisend.

Abbildung 23: Informationsschrift und Einwilligungserklärung (Seite 5/6)



Ich willige in die Verarbeitung der genannten Daten ein.

Ein Exemplar der Informationsschrift und der Einwilligungserklärung habe ich erhalten. Ein Exemplar verbleibt im Prüfzentrum.

Unterschrift des Teilnehmers/der Teilnehmerin

(Name und Vorname in Druckschrift)

(Datum)

(Unterschrift)

Erklärung und Unterschrift des aufklärenden Studienverantwortlichen/ der aufklärenden Studienverantwortlichen
Ich habe das Aufklärungsgespräch geführt und die Einwilligung eingeholt.

Sarah Frederike Brose
(Name und Vorname in Druckschrift)

21.08.2021
(Datum)

(Unterschrift)

Gemeinsam. Fürsorglich. Wegweisend.

Abbildung 24: Informationsschrift und Einwilligungserklärung (Seite 6/6)

8 Zusammenfassung

Obwohl Kommunikation in der Medizin eine zentrale und alltägliche Aufgabe ist, ist das richtige Verständnis von statistischen Informationen, vor allem von Bayesianischen, oftmals eine große Herausforderung. Dabei muss zwischen zwei möglichen *Informationsrichtungen* unterschieden werden: *Bayesianische Informationen* (z.B. der Anteil der positiv getesteten Personen unter den erkrankten Personenwagen) und *diagnostische Informationen* (z.B. der Anteil der erkrankten Personen unter den positiv getesteten Personen).

Ziel der Studie war es, den Einfluss der bei einer ärztlichen Kommunikation mit Patient*innen relevanten Parameter *Informationsrichtung* (*Bayesianisch vs. diagnostisch*) und *Informationsvisualisierung* (*mit Häufigkeitsnetz vs. ohne Visualisierung*) auf das Verständnis der statistischen Information zu überprüfen.

Hierzu bearbeiteten 109 Proband*innen als potenzielle Patient*innen jeweils 4 Aufgaben ($2 \times 2 \times 4$ Design) im Videoformat, wobei ein*e Ärzt*in ihnen ihr fiktives Untersuchungsergebnis erklärte. Die Zahlen der endokrinologischen Kontexte wurden in Form von natürlichen Häufigkeiten angegeben. Dabei wurde sowohl die Informationsvisualisierung als auch die Informationsrichtung systematisch variiert. Anschließend sollten die Proband*innen den *positiven prädiktiven Wert* angeben, wobei dessen Richtigkeit und die hierfür benötigte Zeit ausgewertet wurden.

Eine diagnostische Informationsrichtung konnte die Verständnisrichtigkeit erhöhen (72% ohne Visualisierung bzw. 61% mit Häufigkeitsnetz). Auffällig dabei war jedoch, dass eine Visualisierung mittels eines Häufigkeitsnetzes nur bei den Bayesianischen Version von Vorteil war (37% mit Häufigkeitsnetz vs. 10% ohne Visualisierung). Die Zeiten zum Beantworten der Fragen und Häufigkeiten des Videoanschauens verhielten sich dabei konträr zu den Ergebnisrichtigkeiten, sodass die Bearbeitung Bayesianischer Aufgaben ohne Visualisierung am längsten dauerte (Median 106 Sekunden, bei den anderen Versionen 13,5, 14,0 und 14,5 Sekunden).

Im nächsten Schritt steht nun an, die Theorie einer guten ärztlichen Kommunikation mit Patient*innen in den medizinischen Alltags zu integrieren.

Abstract

Although communication is a central and everyday task in medicine, the proper understanding of statistical information, especially Bayesian ones, is often a major challenge. It is important to distinguish between two possible *directions of information*: *Bayesian information* (e.g., the proportion of persons tested positive among those with disease) and *diagnostic information* (e.g., the proportion of persons with disease among those tested positive).

The aim of the study was to examine the influence of the parameters *information direction* (*Bayesian vs. diagnostic*) and *information visualization* (*with frequency net vs. without visualization*), which are relevant in a medical communication with patients, on the understanding of the statistical information.

For this purpose, 109 participants as potential patients each completed 4 tasks ($2 \times 2 \times 4$ design) in video format, with a physician explaining their fictitious examination result. The numbers of endocrinological contexts were given in the form of natural frequencies. Both the information visualization and the information direction were systematically varied. Subsequently, the participants were asked to indicate the *positive predictive value*, and its correctness and the time required for this were evaluated.

A diagnostic information direction could increase the correctness of understanding (72% without visualization and 61% with frequency net). However, it was striking that a visualization by means of a frequency net was only advantageous for the Bayesian version (37% with frequency net vs. 10% without visualization). The times for answering the questions and frequencies of the video visualization behaved contrary to the result correctness, so that the processing of Bayesian tasks without visualization took the longest time (median 106 seconds, with the other versions 13.5, 14.0 and 14.5 seconds).

The next step is to integrate the theory of good medical communication with patients into everyday medical practice.

9 Danksagung

Zuerst möchte mich bei meiner Betreuerin Frau Dr. med. Leah Braun für die großartige Idee zum Promotionsthema bedanken, wodurch diese Doktorarbeit zustande kommen konnte. Ihre ausgezeichnete Betreuung zeichnete sich vor allem dadurch aus, dass ihre Antworten sowie effizienten Korrekturen nie lange auf sich warten ließen und sie bei Problemen sofort hilfreich zur Verfügung stand, wodurch sie maßgeblich zum zügigen Gelingen dieser Arbeit beigetragen hat.

Ein ganz besonderer Dank gilt ebenso Frau Prof. Dr. Karin Binder, die die Studie von der Erstellung bis zur Publikation durchgehend mit großem Enthusiasmus begleitet hat. Ihre wertvolle Unterstützung bei der statistischen Auswertung, ihre konstruktiven Ideen und ihr promptes Beantworten von Fragen haben diese Arbeit außerordentlich positiv beeinflusst.

Mit besonderem Nachdruck möchte ich meinem Doktorvater Herr Prof. Dr. med. Ralf Schmidmaier danken, der mich in wissenschaftlicher Hinsicht stets exzellent unterstützt hat. Darüber hinaus stand er mir auch in inspirierenden Gesprächen mit Ratschlägen für meinen persönlichen Werdegang zur Seite.

Auch Herrn Prof. Dr. med. Martin Reincke und Herrn Univ.-Prof. Dr. med. Martin Fischer gebührt mein Dank für die hervorragende Zusammenarbeit, unter anderem bestehend aus zahlreichen gewinnbringenden Tipps und zielführenden Denkanstößen.

Mein persönlicher Dank gilt außerdem meinem Freund, der jederzeit und uneingeschränkt für mich da war, um mir Halt zu geben, mich zu motivieren und mich aufzumuntern; meinem Papa und meiner Mama, meinen größten Vorbildern, die mich in jeder Hinsicht während des Studiums unterstützt haben und immer ein offenes Ohr für mich hatten; und meinen beiden Schwestern, die stets an mich geglaubt und mit mir mitgefiebert haben. Ebenso möchte ich mich bei dem Rest meiner Familie und all meinen Freund*innen bedanken, die mich während meines Studiums und meiner Promotion helfend begleitet haben.

Zuletzt gilt mein Dank allen Proband*innen, durch deren Teilnahme diese Studie erst ermöglicht wurde.

10 Publikationsliste

Teile dieser Arbeit wurden bereits in der internationalen Fachzeitschrift PLOS ONE veröffentlicht:

- **Brose SF**, Binder K, Fischer MR, Reincke M, Braun LT, Schmidmaier R. Bayesian versus diagnostic information in physician-patient communication: Effects of direction of statistical information and presentation of visualization. PLOS ONE. 2023;18(6):e0283947.
 - DOI: 10.1371/journal.pone.0283947
 - Siehe (1)

Die Ergebnisse dieser Arbeit wurden zudem auf dem Jahreskongress DVO „Osteologie 2022“ in Baden-Baden (18.-20. September 2022) in Form eines Vortrags und eines Posters vorgestellt und als Abstract veröffentlicht:

- **Brose SF**, Binder K, Fischer MR, Braun LT, Schmidmaier R. Bayesianische vs. diagnostische ärztliche Informationsvermittlung an Patient*innen – Der Einfluss von Richtung der statistischen Information und Visualisierung. Osteologie. 2022;31(03):V 2.5.
 - DOI: 10.1055/s-0042-1755852
 - Siehe (88)

11 Affidavit

Eidesstattliche Versicherung

Brose, Sarah Frederike

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

Bayesianische versus diagnostische ärztliche Informationsvermittlung an Patient*innen: Der Einfluss von Richtung der statistischen Information und Visualisierung

selbstständig 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, 17.02.2025

Sarah Frederike Brose

12 Übereinstimmungserklärung

Erklärung zur Übereinstimmung der gebundenen Ausgabe der Dissertation mit der elektronischen Fassung

Ihr Name: Sarah Frederike Brose

Ihre Adresse: Klinikum der Ludwig-Maximilians-Universität München,
Medizinische Klinik und Poliklinik IV, München, Deutschland

Hiermit erkläre ich, dass die elektronische Version der eingereichten Dissertation mit dem Titel

Bayesianische versus diagnostische ärztliche Informationsvermittlung an Patient*innen: Der Einfluss von Richtung der statistischen Information und Visualisierung

in Inhalt und Formatierung mit den gedruckten und gebundenen Exemplaren übereinstimmt.

München, 17.02.2025

Sarah Frederike Brose